

Special Issue Reprint

Cognitive Load Theory

Emerging Trends and Innovations

Edited by
Fred Paas, Andre Tricot, Kim Ouwehand and Florence Lespiau

mdpi.com/journal/education

Cognitive Load Theory: Emerging Trends and Innovations

Cognitive Load Theory: Emerging Trends and Innovations

Guest Editors

Fred Paas

Andre Tricot

Kim Ouwehand

Florence Lespiau



Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Guest Editors

Fred Paas

Psychology, Education and
Child Studies
Erasmus University Rotterdam
Rotterdam
Netherlands

Andre Tricot

Laboratoire de Psychologie
Epsilon
Université de Montpellier
Paul-Valéry
Montpellier
France

Kim Ouwehand

Department of Psychology,
Education and Child Studies
Erasmus University Rotterdam
Oegstgeest
Netherlands

Florence Lespiau

Faculté Psychologie-STAPS
Nîmes University
Nîmes
France

Editorial Office

MDPI AG

Grosspeteranlage 5

4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Education Sciences* (ISSN 2227-7102), freely accessible at: <https://www.mdpi.com/journal/education/special-issues/O7BTYCHS5J>.

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, A.A.; Lastname, B.B. Article Title. <i>Journal Name</i> Year , Volume Number, Page Range.
--

ISBN 978-3-7258-4040-3 (Hbk)

ISBN 978-3-7258-4039-7 (PDF)

<https://doi.org/10.3390/books978-3-7258-4039-7>

© 2025 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Contents

About the Editors	vii
Preface	ix
Kim Ouwehand, Florence Lespiau, André Tricot and Fred Paas	
Cognitive Load Theory: Emerging Trends and Innovations	
Reprinted from: <i>Educ. Sci.</i> 2025 , <i>15</i> , 458, https://doi.org/10.3390/educsci15040458	1
John Guzmán and Jimmy Zambrano R.	
Effects of Split-Attention and Task Complexity on Individual and Collaborative Learning	
Reprinted from: <i>Educ. Sci.</i> 2024 , <i>14</i> , 1035, https://doi.org/10.3390/educsci14091035	7
Björn B. de Koning	
Best of Both Worlds? Combining Physical and Mental Self-Management Strategies to Support Learning from Split-Attention Examples	
Reprinted from: <i>Educ. Sci.</i> 2024 , <i>14</i> , 284, https://doi.org/10.3390/educsci14030284	20
Melanie Trypke, Ferdinand Stebner and Joachim Wirth	
The More, the Better? Exploring the Effects of Modal and Codal Redundancy on Learning and Cognitive Load: An Experimental Study	
Reprinted from: <i>Educ. Sci.</i> 2024 , <i>14</i> , 872, https://doi.org/10.3390/educsci14080872	32
Ouhao Chen, Endah Retnowati, Juan Cristobal Castro-Alonso, Fred Paas and John Sweller	
The Relationship between Interleaving and Variability Effects: A Cognitive Load Theory Perspective	
Reprinted from: <i>Educ. Sci.</i> 2023 , <i>13</i> , 1138, https://doi.org/10.3390/educsci13111138	45
Ghislain Mugisha and Amael Arguel	
Procedural Learning in Mixed Reality: Assessing Cognitive Load and Performance	
Reprinted from: <i>Educ. Sci.</i> 2025 , <i>15</i> , 339, https://doi.org/10.3390/educsci15030339	60
Rebecca B. Brockbank and David F. Feldon	
Cognitive Reappraisal: The Bridge between Cognitive Load and Emotion	
Reprinted from: <i>Educ. Sci.</i> 2024 , <i>14</i> , 870, https://doi.org/10.3390/educsci14080870	82
Madison Graham, Marinela Ilic, Martine Baars, Kim Ouwehand and Fred Paas	
The Effect of Self-Monitoring on Mental Effort and Problem-Solving Performance: A Mixed-Methods Study	
Reprinted from: <i>Educ. Sci.</i> 2024 , <i>14</i> , 1167, https://doi.org/10.3390/educsci14111167	97
Anna Gorbunova, Christopher Lange, Alexander Savelyev, Kseniia Adamovich and Jamie Costley	
The Interplay of Self-Regulated Learning, Cognitive Load, and Performance in Learner-Controlled Environments	
Reprinted from: <i>Educ. Sci.</i> 2024 , <i>14</i> , 860, https://doi.org/10.3390/educsci14080860	125
Lorenzo Consalvi, Kim Ouwehand and Fred Paas	
Effects of Observing Urban and Natural Scenes on Working Memory Depletion and Restoration: An EEG Study	
Reprinted from: <i>Educ. Sci.</i> 2024 , <i>14</i> , 1204, https://doi.org/10.3390/educsci14111204	140

Katrin Schuessler, Vanessa Fischer, Maik Walpuski and Detlev Leutner The Moderating Role of Interest in the Relationship between Perceived Task Difficulty and Invested Mental Effort Reprinted from: <i>Educ. Sci.</i> 2024 , <i>14</i> , 1044, https://doi.org/10.3390/educsci14101044	168
Hee Min Lee and Paul Ayres The Worked-Example Effect and a Mastery Approach Goal Orientation Reprinted from: <i>Educ. Sci.</i> 2024 , <i>14</i> , 597, https://doi.org/10.3390/educsci14060597	183
Kristin Altmeyer, Roland Brünken, Jochen Kuhn and Sarah Malone The Role of Cognitive Learner Prerequisites for Cognitive Load and Learning Outcomes in AR-Supported Lab Work Reprinted from: <i>Educ. Sci.</i> 2024 , <i>14</i> , 1161, https://doi.org/10.3390/educsci14111161	197
Maliha Naushad Mian, Hannah Beder, Nadine Marcus and Paul Ayres Comparing Real and Imitative Practice with No Practice during Observational Learning of Hand Motor Skills from Animations Reprinted from: <i>Educ. Sci.</i> 2024 , <i>14</i> , 949, https://doi.org/10.3390/educsci14090949	220
Conor Desmond Kenneally and Brendan Bentley A Cognitive Load Approach to Molecular Geometries: Augmented Reality Technology and Visuospatial Abilities in Chemistry Reprinted from: <i>Educ. Sci.</i> 2024 , <i>14</i> , 1036, https://doi.org/10.3390/educsci14091036	238
Amedee Marchand Martella, Alyssa P. Lawson and Daniel H. Robinson How Scientific Is Cognitive Load Theory Research Compared to the Rest of Educational Psychology? Reprinted from: <i>Educ. Sci.</i> 2024 , <i>14</i> , 920, https://doi.org/10.3390/educsci14080920	251

About the Editors

Fred Paas

Fred Paas is a Professor of Educational Psychology at the Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Netherlands, and a Professorial Fellow at the School of Education, University of New South Wales, Sydney, Australia. His research career centers on optimizing learning environments through the application of cognitive load theory, with a focus on instructional design, cognitive psychology, and educational psychology. Paas earned a Master's degree in Human Movement Sciences from Maastricht University in 1988 and a Ph.D. in Instructional Technology from the University of Twente in 1993. Since the 1990s, he has utilized cognitive load theory to investigate the instructional control of cognitive load in complex cognitive tasks, contributing over 300 peer-reviewed articles and book chapters, which have garnered more than 70,000 citations. His current projects explore the interplay between physical environments, physical activities (e.g., gestures and exercise), and cognition, as well as evolutionary approaches to cognitive load and learning. His scientific interests include cognitive architecture, working memory, expertise reversal effect, and the design of effective and efficient learning environments, with recent emphasis on the human movement effect and cognitive aging. Paas is a Fellow of the American Educational Research Association, and serves as Editor-in-Chief of *Educational Psychology Review*. He is on the editorial boards of journals such as the *Journal of Educational Psychology*.

Andre Tricot

André Tricot is a Professor of Cognitive Psychology at Université Montpellier Paul Valéry and a member of the Epsilon Laboratory. His research focuses on the interaction between natural memory and artificial memory systems, particularly how the design of documents can aid memory without overloading it. His work has practical implications in instructional design, human–computer interactions, ergonomics, and transport safety, especially aviation. He led the national group that developed France's primary school curricula (Years 1–3) in 2014–2015 and was the Scientific Director of the “Cognition, Behaviors, Language” program at the French National Research Agency (ANR) from 2021 to 2022. He currently co-directs the French National Center for Education Studies (Cnesco, Paris). Professor Tricot collaborates with organizations such as Airbus, Thales, ArianeGroup, Air France, Orange, PSA, Nathan, and Éditions Milan. He is the author of around twenty books and editor of the *Myths and Realities in Education* series published by Retz.

Kim Ouwehand

Kim Ouwehand is an Assistant Professor at the Department of Psychology, Education, and Child Studies, at the Erasmus University Rotterdam, Netherlands. Her research centers on optimizing learning by combining knowledge from fundamental cognitive psychology and educational psychology, with a particular emphasis on cognitive load theory and embodied cognition. Her work has resulted in numerous peer-reviewed publications, including over 20 journal articles and book chapters, with significant contributions to understanding how physical actions and environmental factors influence cognitive processes. She was recently invited to share her knowledge on this topic with an international colloaboration that resulted in her 2025 publication in *Nature Human Behaviour*.

Ouwehand earned her Ph.D. in Psychology from the Erasmus University Rotterdam, focusing on the role of gestures in learning and cognitive load, particularly among older adults. Currently, the main focus of her research is on how physical movement, exercise, and environmental stimuli, such as observing natural versus urban scenes, affect working memory, cognitive restoration, and problem solving. Additionally, she also examines self-regulated learning, the impact of light exposure on cognition, and school-level factors like policy and teacher experiences. Ouwehand has received recognition for her contributions, including invitations to present at international conferences such as the EARLI and the International Convention of Psychological Science. Her research continues to inform instructional design and educational practices, bridging cognitive science with real-world applications.

Florence Lespiau

Florence Lespiau is an Associate Professor in Cognitive Psychology at Nîmes University, France, and a researcher at the APSY-v Laboratory. Since earning her Ph.D. in Cognitive Psychology from the University of Toulouse 2 in 2017, her research has focused on cognitive load theory and the evolutionary approach to learning. Her doctoral dissertation explored how primary knowledge—evolutionarily early knowledge—facilitates logical reasoning and learning compared to secondary knowledge. Building on this foundation, her current work addresses two central questions—“Why are individuals naturally more motivated and efficient at acquiring their native language during early childhood, despite its complexity, while even basic arithmetic concepts (such as $3 + 3 = 6$) require sustained effort?” and “How can this understanding be leveraged to improve educational practices?”. Supported by a grant from the French National Research Agency awarded in 2024, her ongoing projects investigate how cognitive load theory can be integrated with the principle of desirable difficulties to design more effective instructional materials. A key focus is on how primary knowledge reduces cognitive load and increases learner motivation, thereby supporting the acquisition of more complex secondary knowledge. She has presented her work regularly through academic publications and international conferences. Her areas of expertise include cognitive psychology, cognitive ergonomics in education, expertise reversal effects, and the practical application of evolutionary cognitive principles in order to improve teaching and learning.

Preface

This Reprint presents the Special Issue, titled “Cognitive Load Theory: Emerging Trends and Innovations”, published in *Education Sciences* in April 2025. This collection explores the evolving landscape of cognitive load theory (CLT), a foundational framework in educational psychology that bridges cognitive science and instructional design to optimize learning, performance, and training outcomes. Its subject centers on how CLT adapts to modern educational challenges, with a scope encompassing emerging trends such as technology-enhanced learning, embodied cognition, and individual learner differences. The aim of this collection is to showcase innovative research that extends CLT’s applications, while its purpose is to inspire educators, instructional designers, and researchers in order to refine teaching practices in an era of rapid technological and societal changes.

The motivation for this work stems from transformative shifts in learning contexts since CLT’s inception in 1988. Advances in technology and the rise of lifelong learning have reshaped how knowledge is acquired, necessitating fresh perspectives on managing cognitive load. This Special Issue was inspired by the 15th International Cognitive Load Theory Conference hosted by the Epsilon Lab at the University of Montpellier Paul-Valéry, France. It compiles 15 articles that explore these dynamics. It is addressed to a broad audience, i.e., researchers and practitioners in educational psychology, instructional design, and cognitive science, as well as educators seeking evidence-based strategies to enhance learning environments.

We, the editors—Kim Ouwehand, Florence Lespiau, André Tricot, and Fred Paas—are grateful to have collaborated on this endeavor. and express our heartfelt acknowledgment to the authors for their rigorous and innovative work, the reviewers for their thoughtful feedback, and the Epsilon Lab team for their support during the conference that sparked this Special Issue. Special thanks go to the Special Issue editor Mr. Alan Chen, the editorial staff at *Education Sciences*, and MDPI for their assistance in bringing this Reprint to fruition. This collective effort underscores our shared commitment to advancing CLT scholarship, and we hope it serves as a valuable resource for all who engage with it.

Fred Paas, Andre Tricot, Kim Ouwehand, and Florence Lespiau

Guest Editors

Cognitive Load Theory: Emerging Trends and Innovations

Kim Ouwehand ^{1,*}, Florence Lespiau ², André Tricot ³ and Fred Paas ^{1,4}

¹ Department of Psychology, Education, and Child Studies, Erasmus University Rotterdam, 3062 PA Rotterdam, The Netherlands; paas@essb.eur.nl

² Laboratoire APSY-v, Faculté Psychologie et STAPS, Nîmes Université, 30021 Nîmes, France; florence.lespiau@unimes.fr

³ Laboratoire de Psychologie Epsilon, Université de Montpellier Paul-Valéry, 34199 Montpellier, France

⁴ School of Education, University of New South Wales, Sydney, NSW 2308, Australia

* Correspondence: ouwehand@essb.eur.nl

1. Introduction

Cognitive load theory (Sweller et al., 1998, 2019) has been a cornerstone of educational research for nearly four decades, integrating principles from cognitive psychology and instructional design to enhance learning, performance, and training outcomes. Supported by extensive research (for meta-analyses, see, Barbieri et al., 2023; Ginns, 2005, 2006; Schroeder & Cencki, 2018; Sundararajan & Adesope, 2020), cognitive load theory (CLT) provides insights into how the constraints and capacities of the human information processing system can be used to optimize instructional methods and learning environments. Its focus on aligning educational practices with cognitive limitations and strengths has significantly informed approaches to effective teaching and learning.

However, the context of learning has changed significantly since the beginnings of CLT, as introduced by Sweller (1988). Rapid technological developments have transformed learning into a lifelong continuously evolving journey rather than a static school-based concept (Fidalgo & Thormann, 2024). This shift presents both advantages and challenges for learners and instructional designers. On the downside, the increasing flow of stimuli can make it difficult for learners to filter information and focus on what is most relevant. On the upside, understanding human cognitive architecture, including its limitations (such as working memory capacity (Baddeley, 1992, 2020) and strengths (such as its capacity for embodied learning (Barsalou, 2008, 2020) combined with insights into individual learner characteristics, enables instructional designers to create optimized learning environments (Zou et al., 2025). Additionally, exploring the supportive role of modern technology offers promising solutions for enhancing learning efficiency. This Special Issue, based on contributions to the 15th International Cognitive Load Theory Conference hosted by the Epsilon Lab at the University of Montpellier Paul-Valéry, France, brings together 15 articles showcasing emerging trends and innovations in cognitive load theory research.

2. Presenting and Handling Learning Materials

Two articles in this Special Issue address the split-attention effect, a phenomenon that typically results in increased cognitive load and impaired learning when essential information is spread across multiple, separate sources—such as text and images—rather than presented in an integrated format (Chandler & Sweller, 1991; Pouw et al., 2019). In the first contribution to this Special Issue, Guzmán and Zambrano (2024) investigate the split-attention effect by comparing individual versus collaborative learning approaches. They also examine element interactivity, a measure of material difficulty based on the

Received: 18 March 2025

Accepted: 28 March 2025

Published: 7 April 2025

Citation: Ouwehand, K., Lespiau, F., Tricot, A., & Paas, F. (2025). Cognitive Load Theory: Emerging Trends and Innovations. *Education Sciences*, 15(4), 458. <https://doi.org/10.3390/educsci15040458>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

number of informational elements that must be integrated for effective learning. Their results suggest that individual learning is more effective for high element interactivity materials presented in an integrated format, while collaborative learning in dyads is more beneficial when information is presented in a dispersed format. The second contribution, by De Koning (2024), explores strategies to mitigate the split-attention effect by comparing a physical approach (annotation) with a mental approach (imagery and drag-and-drop) in university students. It was found that combining physical and mental strategies significantly reduced the cognitive load, offering practical insights for designing materials with dispersed information by supporting a dual-strategy approach.

While a fragmented presentation of information affects learning due to split attention, redundancy (when the same information is presented in multiple ways or instances) can affect learning too (Kalyuga & Sweller, 2014). In the third contribution, Trypke et al. (2024) investigated modal and codal redundancy and their effect on learning and cognitive load. Modal redundancy referred to stimuli that were presented in two ways in the same modality, in this case the visual modality (images and text). In the codal redundancy condition, the information was presented in two different modalities but used the same linguistic code (spoken and written text). It seemed that modal redundancy facilitated learning and led to lower cognitive load, while codal redundancy elicited a classic (aversive) redundancy effect.

Another key approach in handling learning materials is the distinction between blocked and mixed practice. In the fourth contribution, Chen et al. (2023) studied middle-school students and found that studying multiple domains (e.g., mathematics and language) within a single session was more efficient than studying each domain separately in blocked sessions. However, this effect was only found for materials that looked similar on the surface and not for materials that required completely different approaches. Their findings suggest that alternating between subjects within a study session enhances learning efficiency, supporting a mixed approach for homework across different domains.

In the fifth contribution, Mugisha and Arguel (2025) explored the use of mixed reality in learning knot tying as a technical skill needed for surgical procedures. They tested the principle of spatial contiguity and aimed to measure its impact on cognitive load and procedural learning performance. Their main finding, that an integrated- compared to a non-integrated learning format was specifically beneficial for intrinsic load, has implications for instructional design recommendations for procedural learning tasks.

3. Monitoring

The rise of data-driven and automated education has introduced both opportunities and challenges, with standardized assessments contributing to a high-stakes testing culture that may reduce learners to data points focused on scores, potentially impacting emotional wellbeing (French et al., 2024). Internal factors such as self-regulation (Schunk & Zimmerman, 2008; Yang et al., 2024) offer resilience, as highlighted by the integrative review by Brockbank and Feldon (2024) in the sixth contribution. These researchers propose cognitive reappraisal as a self-monitoring strategy to manage negative emotions in learning. More specifically, they propose that cognitive reappraisal, which involves reframing or reassessing the beliefs underlying an emotional response to better align with goals (Gross, 2015), may reduce the impact of negative emotions on cognitive load.

In the seventh contribution, Graham et al. (2024) investigated self-monitoring via “thinking aloud” in university students solving biology problems to assess its effects on learning and cognitive load. Although thinking aloud did not facilitate learning directly, it led to longer task engagement, with lower monitoring accuracy than the control group.

These findings underscore the importance of careful implementation of monitoring strategies in learning contexts. In the eighth contribution, Gorbunova et al. (2024) investigated graduate law students' prior knowledge and self-regulated learning skills when studying the civil code in an online session. They found that both these factors led to increased germane load and performance. These authors stress the need to carefully consider and select learner-control options in asynchronous online environments.

4. Working Memory Recovery

In today's information-saturated era, stress and mental fatigue due to information overload is a growing concern (Arnold et al., 2023), highlighting the need for interventions to help replenish cognitive energy. Consalvi et al. (2024) explored one such intervention in the ninth contribution by examining the effects of nature exposure on working memory recovery. While previous behavioral research suggests that natural environments can help replenish cognitive resources (Mason et al., 2022; Van Oordt et al., 2023), Consalvi et al. (2024) used psychophysiological measures in addition to behavioral measures to assess whether exposure to nature imagery could restore working memory after a demanding task. Although no significant effects were observed, the study offers a rigorous method for continuous cognitive load assessment, which may inform future interventions. Future research should explore alternative strategies for replenishing cognitive capacity beyond nature exposure, such as structured rest breaks or mindfulness interventions.

5. Individual Learner Characteristics

The trend toward personalized learning underscores the importance of individual factors, such as topic interest and cognitive capacities, in shaping learning outcomes and cognitive load (Bernacki et al., 2021). This was addressed in the tenth contribution by Schuessler et al. (2024), who examined how topic interest moderates the relationship between task complexity and mental effort in chemistry assignments for middle-school students, finding that low-interest tasks required more cognitive effort for simple assignments but not for more complex ones. In the eleventh contribution, Lee and Ayres (2024) demonstrated that using worked examples in mathematics improved retention and reduced cognitive load compared to problem-solving strategies, particularly for students with a strong mastery approach orientation. In the twelfth contribution, Altmeyer et al. (2024) investigated the effects of augmented reality (AR) on learning about electricity, focusing on individual differences such as spatial abilities and verbal working memory. Results suggest that tablet-based AR benefitted students with low spatial ability, while AR glasses better supported those with low verbal working memory—though these effects were not observed in primary school children, likely due to developmental factors.

6. Embodied Learning

Cognitive load theory also considers the affordances of human cognitive architecture, including the role of embodied learning. Grounded in embodied cognition (Barsalou, 2008, 2020) and evolutionary educational psychology (Geary, 2012, 2024), embodied learning emphasizes that humans understand the world through bodily interactions. In the thirteenth contribution, Mian et al. (2024) compared real and imitative practice in learning piano pieces and paper folding. The findings suggest that imitation enhanced learning efficiency in piano, while real practice was more effective for paper folding, in the female participants specifically indicating that the level of embodiment may need to be tailored to specific tasks and gender. Embodiment in learning was studied in the domain of geometry by Kenneally and Bentley (2024) in the fourteenth contribution to investigate the use of AR in teaching molecular geometry and determined that allowing students to explore 3D

molecular structures through physical interaction reduced the cognitive load and improved understanding compared to mental rotation alone.

7. Reflections on CLT in Modern Research

Given rapid technological advancements, it is crucial to continually evaluate whether the guiding theories in educational research, including CLT, remain relevant. In the fifteenth contribution, Martella et al. (2024) reviewed recent CLT research (2020–2023), noting a decrease in experimental studies and a shift toward intervention research, which may impact study replicability. They emphasized the need for ongoing experimental validation to support robust and reliable CLT research outcomes.

8. Conclusions

Overall, this Special Issue reaffirms the ongoing relevance of cognitive load theory in modern learning contexts. The studies presented underscore the importance of leveraging technology and instructional design to address cognitive limitations in learners, offering promising interventions to manage cognitive load in our rapidly evolving society. Together, these findings suggest that CLT will continue to be a valuable framework for future educational research and practice, emphasizing the need for adaptable and experimentally validated approaches in instructional design.

Author Contributions: Conceptualization and methodology—K.O.; writing—original draft preparation—K.O.; writing, review, and editing, K.O., F.L., A.T. and F.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We are grateful to the authors who contributed their work to this Special Issue. The authors who contributed to this issue were invited based on their presentations during the International Cognitive Load Conference, at the Epsilon Laboratoire de Psychologie in Montpellier, France, 2023.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Altmeyer, K., Brünken, R., Kuhn, J., & Malone, S. (2024). The role of cognitive learner prerequisites for cognitive load and learning outcomes in AR-supported lab work. *Educational Sciences*, 14(11), 1161. [CrossRef]
- Arnold, M., Goldschmitt, M., & Rigotti, T. (2023). Dealing with information overload: A comprehensive review. *Frontiers in Psychology*, 14, 1122200. [CrossRef]
- Baddeley, A. (1992). Working memory. *Science*, 255(5044), 556–559. [CrossRef]
- Baddeley, A. (2020). Working memory. In A. Baddeley, M. W. Eysenck, & M. C. Anderson (Eds.), *Memory* (3rd ed., pp. 71–111). Routledge. [CrossRef]
- Barbieri, C. A., Miller-Cotto, D., Clerjuste, S. N., & Chawla, K. (2023). A meta-analysis of the worked examples effect on mathematics performance. *Educational Psychology Review*, 35(1), 11. [CrossRef]
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59(1), 617–645. [CrossRef] [PubMed]
- Barsalou, L. W. (2020). Challenges and opportunities for grounding cognition. *Journal of Cognition*, 3(1), 31. [CrossRef]
- Bernacki, M. L., Greene, M. J., & Lobczowski, N. G. (2021). A systematic review of research on personalized learning: Personalized by whom, to what, how, and for what purpose(s)? *Educational Psychology Review*, 33, 1675–1715. [CrossRef]
- Brockbank, R. B., & Feldon, D. F. (2024). Cognitive reappraisal: The bridge between cognitive load and emotion. *Educational Sciences*, 14(8), 870. [CrossRef]
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8(4), 293–332. Available online: <http://www.jstor.org/stable/3233596> (accessed on 14 March 2025).
- Chen, O., Retnowati, E., Castro-Alonso, J. C., Paas, F., & Sweller, J. (2023). The relationship between interleaving and variability effects: A cognitive load theory perspective. *Educational Sciences*, 13(11), 1138. [CrossRef]

- Consalvi, L., Ouwehand, K., & Paas, F. (2024). Effects of observing urban and natural scenes on working memory depletion and restoration: An EEG study. *Educational Sciences*, 14(11), 1204. [CrossRef]
- De Koning, B. B. (2024). Best of both worlds? Combining physical and mental self-management strategies to support learning from split-attention examples. *Educational Sciences*, 14(3), 284. [CrossRef]
- Fidalgo, P., & Thormann, J. (2024). The future of lifelong learning: The role of artificial intelligence and distance education. In F. G. Paloma (Ed.), *Lifelong learning: Education for the future world* (1st ed., pp. 1–17). IntechOpen. [CrossRef]
- French, S., Dickerson, A., & Mulder, R. A. (2024). A review of the benefits and drawbacks of high-stakes final examinations in higher education. *Higher Education*, 88, 893–918. [CrossRef]
- Geary, D. C. (2012). Evolutionary educational psychology. In K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra, & J. Sweller (Eds.), *APA educational psychology handbook, Vol. 1. Theories, constructs, and critical issues* (pp. 597–621). American Psychological Association. [CrossRef]
- Geary, D. C. (2024). *The evolved mind and modern education: Status of evolutionary educational psychology*. Cambridge University Press.
- Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction*, 15(4), 313–331. [CrossRef]
- Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, 16(6), 511–525. [CrossRef]
- Gorbunova, A., Lange, C., Savelyev, A., Adamovich, K., & Costley, J. (2024). The interplay of self-regulated learning, cognitive load, and performance in learner-controlled environments. *Educational Sciences*, 14(8), 860. [CrossRef]
- Graham, M., Ilic, M., Baars, M., Ouwehand, K., & Paas, F. (2024). The effect of self-monitoring on mental effort and problem-solving performance: A mixed-methods study. *Educational Sciences*, 14(11), 1167. [CrossRef]
- Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, 26(1), 1–26. [CrossRef]
- Guzmán, J., & Zambrano, R. (2024). Effects of split-attention and task complexity on individual and collaborative learning. *Educational Sciences*, 14(9), 1035. [CrossRef]
- Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 247–262). Cambridge University Press. [CrossRef]
- Kenneally, C. D., & Bentley, B. (2024). A cognitive load approach to molecular geometries: Augmented reality technology and visuospatial abilities in chemistry. *Education Sciences*, 14(9), 1036. Available online: <https://www.proquest.com/scholarly-journals/cognitive-load-approach-molecular-geometries/docview/3110442754/se-2> (accessed on 14 March 2025).
- Lee, H. M., & Ayres, P. (2024). The worked-example effect and a mastery approach goal orientation. *Educational Sciences*, 14(6), 597. [CrossRef]
- Martella, A. M., Lawson, A. P., & Robinson, D. H. (2024). How scientific is cognitive load theory research compared to the rest of educational psychology? *Education Sciences*, 14(8), 920. [CrossRef]
- Mason, L., Ronconi, A., Scrimin, S., & Pazzaglia, F. (2022). Short-term exposure to nature and benefits for students' cognitive performance: A review. *Educational Psychology Review*, 34, 609–647. [CrossRef]
- Mian, M. N., Beder, H., Marcus, N., & Ayres, P. (2024). Comparing real and imitative practice with no practice during observational learning of hand motor skills from animations. *Educational Sciences*, 14(9), 949. [CrossRef]
- Mugisha, G., & Arguel, A. (2025). Procedural learning in mixed reality: Assessing cognitive load and performance. *Education Sciences*, 15(3), 339. [CrossRef]
- Pouw, W., Rop, G., de Koning, B., & Paas, F. (2019). The cognitive basis for the split-attention effect. *Journal of Experimental Psychology: General*, 148(11), 2058–2075. [CrossRef]
- Schroeder, N. L., & Cenkci, A. T. (2018). Spatial contiguity and spatial split-attention effects in multimedia learning environments: A meta-analysis. *Educational Psychology Review*, 30, 679–701. [CrossRef]
- Schuessler, K., Fischer, V., Walpuski, M., & Leutner, D. (2024). The moderating role of interest in the relationship between perceived task difficulty and invested mental effort. *Educational Sciences*, 14(10), 1044. [CrossRef]
- Schunk, D. H., & Zimmerman, B. J. (Eds.). (2008). *Motivation and self-regulated learning: Theory, research, and applications*. Routledge.
- Sundararajan, N., & Adesope, O. (2020). Keep it coherent: A meta-analysis of the seductive details effect. *Educational Psychology Review*, 32(3), 707–734. [CrossRef]
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. [CrossRef]
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251–296. [CrossRef]
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31, 261–292. [CrossRef]
- Trypke, M., Stebner, F., & Wirth, J. (2024). The more, the better? Exploring the effects of modal and codal redundancy on learning and cognitive load: An experimental study. *Educational Sciences*, 14(8), 872. [CrossRef]
- Van Oordt, M., Ouwehand, K., & Paas, F. (2023). Restorative effects of observing natural and urban scenery after working memory depletion. *International Journal of Environmental Research and Public Health*, 20(1), 188. [CrossRef]

- Yang, Y., Zheng, Z., Wang, J., & Sun, D. (2024). The relationships of motivation and self-regulation to students' cognitive presence and empowerment in online learning environment. *Educational Information Technologies*, 29, 23391–23414. [CrossRef]
- Zou, L., Zhang, Z., Mavilidi, M., Chen, Y., Herold, F., Ouwehand, K., & Paas, F. (2025). The synergy of embodied cognition and cognitive load theory for optimized learning. *Nature Human Behaviour*, 1–9. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Effects of Split-Attention and Task Complexity on Individual and Collaborative Learning

John Guzmán ^{1,2,*} and Jimmy Zambrano R. ^{2,*}¹ Ministerio de Educación, Quito 170515, Ecuador² Facultad de Ciencias de la Educación y Ciencias Sociales, Universidad Del Pacífico, Guayaquil 090603, Ecuador

* Correspondence: john.guzman@educacion.gob.ec (J.G.); jimmy.zambrano@upacifico.edu.ec (J.Z.R.)

Abstract: School tasks often include individual and collaborative activities supported by a wide variety of learning materials. These materials can elicit varied levels of attention and learning depending on the complexity (i.e., element interactivity level) and physical separation of the information elements in the study material. The aim of this study was to explore the potential effects of the element interactivity level (i.e., high vs. low) and split attention (i.e., integrated vs. separated information) on individual and collaborative learning. An experimental design was implemented with 192 high school learners, with 64 working individually and 128 in dyads. The results revealed that in tasks with high element interactivity and integrated information, individual students learned more than groups. However, separated information benefited groups more than individual learners. It is concluded that the benefits of individual and group learning are mediated by task element interactivity and the physical separation of information sources in the study material, and recommendations for education professionals are presented.

Keywords: split attention; element interactivity; individual learning; collaborative learning; cognitive load

Citation: Guzmán, J.; Zambrano R., J. Effects of Split-Attention and Task Complexity on Individual and Collaborative Learning. *Educ. Sci.* **2024**, *14*, 1035. <https://doi.org/10.3390/educsci14091035>

Academic Editors: Fred Paas, Andre Tricot, Kim Ouwehand and Florence Lespiau

Received: 13 May 2024

Revised: 28 June 2024

Accepted: 4 July 2024

Published: 23 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Learning and problem-solving in groups have come to be understood as key skills of the 21st century, both for education and for the global economy [1]. In an educational context, learning to solve problems may include both individual and collaborative tasks supported by a wide variety of materials and domains [2,3]. Consequently, more research is needed to guide decision-making regarding which instructional combination is most effective.

From the perspective of cognitive load theory [4], the effectiveness of individual or group learning depends on the levels of cognitive load based on the number of information elements and their interrelation [5,6]. Likewise, learning materials can elicit varied levels of cognitive load depending on the degree of physical separation between information segments, i.e., split attention [7]. However, little is known about the potential effects of social learning conditions and split attention on problem-solving learning. This article reports on a field experimental study that examined the level of element interactivity (i.e., high versus low) and split attention (i.e., integrated versus separated information) and discusses the findings.

1.1. Cognitive Load and Element Interactivity

Cognitive load theory is an instructional approach based on the characteristics of human cognitive architecture [4,8]. The fundamental components of this architecture are working memory, long-term memory, and the interaction between the two types of memory. When learning new information, working memory is severely limited because it can only

process a few pieces of information simultaneously, while long-term memory has unlimited storage capacity [9,10]. As the number of new information elements increases, working memory experiences a higher processing load [11,12]. The limitations of working memory disappear when specific information elements are stored in long-term memory [13,14].

Cognitive load refers to the intensity of cognitive activities and information elements that need to be processed in working memory over a period of time (i.e., about 20 s) [4,15–17]. The load varies depending on the number of interacting information elements of the learning task. An information element can be a concept or procedure that needs to be processed and learned [11,12]. The element interactivity level (i.e., level of complexity) depends on the number of element tasks and their interconnection, as well as the learner's level of expertise. [11,14]. For example, a novice learning chemical elements must process pairs of elements, their letters, and their respective names (e.g., H for hydrogen). Learning 118 pairs of elements is very difficult and time-consuming, but it imposes low element interactivity because each element can be processed individually. However, learning algebraic tasks, such as $2(3 + x) = 0$, *solve for x*, may feel overwhelming because the learner must process many interacting elements (e.g., applying basic mathematical operations, breaking parentheses, and rearranging coefficients and constants, among others). Solving this problem requires processing 16 symbols (i.e., $6 + 2x = 0$, $6 + 2x = 0$, $x = -3$) by applying mathematical operations (i.e., distributing the 2 and multiplying it, subtracting 6, dividing by 2) while considering that changing one element affects other elements of the task. In contrast, an expert learner who has accumulated a lot of knowledge about solving problems like these would see this problem as a single element (i.e., low element interactivity level) that requires one or two solution steps (e.g., multiply both terms by 2 and change the sign of 3).

There are two categories of cognitive load that are related to the effectiveness of teaching [12]. Intrinsic load refers to the processing of the information elements that must be learned. This load can be low or high (e.g., $2x = 6$ or $2(3 + x) = 0$, respectively). Extraneous load refers to the processing of information elements unrelated to the task, such as learning to solve a problem through generic cognitive strategies like means–end analysis or trial and error or with poorly designed instructional materials and environments, i.e., split attention [7]. This load should be minimized by manipulating the presentation of information, for example, by using worked or partially worked examples, so that learners focus their working memory resources on the information elements to be learned. Instruction may be more effective when the interacting information elements of the task are kept within the limits of working memory and elements that impose an extraneous cognitive load are minimized. If the level of interactive information elements in the task is low, reducing elements that induce extraneous cognitive load may not be necessary if it does not overload working memory.

1.2. Collaborative Cognitive Load

Collaborative learning is an instructional strategy that involves creating small groups of learners with the purpose of learning to solve academic problems together [2,6,18]. Collaboration has been studied through a variety of approaches [3,19,20], and cognitive and social mechanisms that facilitate or inhibit collaboration have been identified [21]. An emerging approach is to consider groups as information processing systems [22,23]. These systems consist of limited individual working memories that create an extended workspace to process the information elements of the learning task through transactional activities [5]. From this perspective, transactional activities, defined as the elements of communication and interindividual communication, are an information processing mechanism whose cognitive load may either promote or hinder learning [8,22].

Recent research on collaborative learning from the perspective of cognitive load suggests that groups are more efficient and effective when the interactivity of information elements is high [5,6,24–27]. In other words, group members may acquire more knowledge because the cognitive load imposed by the interactive elements of task information and

transactional activities (i.e., communication and interindividual coordination) is shared among group members. However, this extended capacity and the collaborative cognitive load associated with transactional activities would be redundant when the level of element interactivity is low. In these conditions, it is better to learn individually [25,28].

1.3. Split-Attention Effect

A common source of extraneous cognitive load is split attention, which occurs when learners must integrate multiple relevant sources of information that are spatially separated [7,29,30]. Shifting attention from one information element to another requires holding previously learned information in working memory while searching for and processing information from another source [7]. Searching for and integrating multiple sources of physically separated information consume working memory resources that may be used to understand and store the information in long-term memory.

Accordingly, for learners to comprehend and learn information, it should be presented in an integrated, spatial, and temporal manner [7,29]. However, it is worth noting that the split-attention effect has been mostly investigated under individual learning conditions. Nevertheless, the effect of the collective capacity of working memory [5,31] may allow groups to optimally manage separate sources of information. In other words, group members may distribute different information elements among themselves and integrate them through appropriate transactional activities [18,32].

1.4. The Present Study

This study examined how the level of element interactivity (i.e., low (partially worked examples) versus high (conventional problems)) and the type of attention induced by the material (i.e., integrated or separated information) influence individual and group learning. It was expected that the extraneous cognitive load associated with split attention in low-element-interactivity tasks would not be relevant because the cognitive activities of integrating sources of information do not overload the working memory capacity [11,12]. However, low element interactivity is unfavorable for groups because the benefits of the extended capacity of collective working memory are not greater than the cognitive costs induced by transactional activities [5,26,28]. Consequently, in tasks with low information element interactivity, we expected that individual learners would outperform collaborative groups using materials with integrated (h1.2) or separated information (h1.2).

Unlike low-element-interactivity tasks, tasks with high element interactivity demand more cognitive resources, which imposes a high intrinsic cognitive load due to the limited capacity of working memory [11,12]. In these conditions, integrated information material may alleviate the cognitive load of individual learners. On the other hand, groups may leverage their collective working memory capacity to address the extraneous processing demand associated with split attention and the intrinsic processing demand related to the high element interactivity of the task [5,8]. Consequently, it was expected that materials with integrated information would benefit individual learners more than groups (h2.1). In contrast, materials inducing split attention would not affect collaborative groups due to their large working memory capacity. Consequently, we expected that groups would outperform individuals in tasks with a high level of task element interactivity and with material containing separated information (h2.2).

2. Materials and Methods

2.1. Participants

The sample size was calculated using G*Power with a medium effect size of $\eta_p^2 = 0.06$, error = 0.05, and power = 0.80, resulting in 126 samples. An additional 50% of learners were added due to the collaborative learning condition in pairs, resulting in 192 participants. Students were Ecuadorian from a public school in Guayaquil, and the tasks were part of their mathematics class. Public schools tend to have large classes, and mathematics allows for the manipulation of element interactivity. The average age of the learners, of

whom 97 were male (51%) and 95 were female (49%), was 15.50 years ($SD = 0.79$). Learners were randomly assigned, and institution authorities were informed of and approved their participation. All learners were novices in the learning topic.

2.2. Design and Procedure

A factorial design with 2 factors was used: element interactivity (i.e., low vs. high), material type (i.e., separated information vs. integrated information), and social condition (i.e., individual vs. group). The dependent variable was academic performance in mathematics. The study was conducted in four stages: pretest of knowledge, learning stage, immediate test stage, and delayed test stage. Each stage was structured with four 40 min sessions, equivalent to four regular class periods. A teacher provided participants with guidance on how to use study methods and treatment fidelity to ensure proper application. They were instructed to use a digital clock application to display the number of minutes assigned to each activity. Additionally, they were informed that they would receive 10 academic compensation points for participating.

The four involved stages were executed as follows: In the first stage, there was a preliminary test, where all learners in the class were individually assessed on their knowledge of complex numbers. In the second stage, learners were assigned using a randomization scheme. Four groups were formed, each with 16 learners, in the following way: (1) individual learning with low-element-interactivity tasks and material with integrated information; (2) individual learning with low-element-interactivity tasks and material with separated information; (3) individual learning with high-element-interactivity tasks and material with integrated information; and (4) individual learning with high-element-interactivity tasks and material with separated information.

Another four groups were formed, each with 32 learners, who worked in pairs as described below: (1) collaborative learning in pairs with low-element-interactivity tasks and material with integrated information; (2) collaborative learning in pairs with low-element-interactivity tasks and material with separated information; (3) collaborative learning in pairs high-element-interactivity tasks and material with integrated information; (4) collaborative learning with high-element-interactivity tasks with separated information. Instructional guidance was provided according to the type of learning for each treatment. The third stage involved an immediate test after the learning stage, which included retention and transfer tests. Finally, the fourth stage was an 8-day-delayed test. Additionally, learners were informed that they could ask the teacher if they had any doubts.

2.3. Materials

The material was based on operations with complex numbers (i.e., real and imaginary), which are part of the content of mathematics in the national curriculum. Learners received instructional guidance and techniques for adding, subtracting, multiplying, and dividing complex numbers, as well as solved examples and practical problems to solve. Materials were designed considering three key sources of information: a detailed introduction to each operation (i.e., addition, subtraction, multiplication, and division), a worked example for each operation (i.e., detailed and numbered steps accompanied by its solution), and a practice problem for each operation. The level of interactivity of each element of the tasks was manipulated in the practice problem. A task with low element interactivity consisted of completing a partially worked example. For instance, Step 1 of the task below requires regrouping two pairs of information elements. But, the partially solved Step 1 only requires completing each group. Step 2 requires identifying, calculating, and representing imaginary numbers. However, the partially solved Step 2 indicates the imaginary number to be calculated and represented. These instructional procedures decreased the element interactivity level and its respective cognitive load.

Task 1: Complete the processes and solve the following sum with complex numbers.

$$\left(\frac{2}{5} \sqrt[3]{7} + \sqrt{-25}\right) + \left(\frac{3}{2} \sqrt[3]{7} + \sqrt{-36}\right) =$$

Step 1: Group the real part and the imaginary part.

$$\left(\frac{2}{5}\sqrt[3]{7} + \sqrt{-25}\right) + \left(\frac{3}{2}\sqrt[3]{7} + \sqrt{-36}\right) =$$

Step 2: Represent the roots of the imaginary numbers using the imaginary unit.

$$\begin{aligned}\sqrt{-25} &= \\ \sqrt{-36} &= \end{aligned}$$

Step 3: Add the real grouping and the imaginary grouping.

$$\left(\frac{2}{5}\sqrt[3]{7} + \frac{3}{2}\sqrt[3]{7}\right) + (6i) =$$

A task with high element interactivity consisted of solving a conventional problem. Conventional problems only included the step description, requiring learners to solve each step. An example of a conventional problem is as follows:

Task 1: Complete the processes and solve the following sum with complex numbers.

$$\left(\frac{2}{5}\sqrt[3]{7} + \sqrt{-25}\right) + \left(\frac{3}{2}\sqrt[3]{7} + \sqrt{-36}\right) = \quad (1)$$

Step 1: Group the real part and the imaginary part.

Step 2: Represent the roots of the imaginary numbers using the imaginary unit.

Step 3: Add the real grouping and the imaginary grouping.

The integrated information material for the tasks with low element interactivity consisted of numbered step descriptions, followed by the respective steps' solutions, and the worked example was followed by the respective practice problem for each operation (i.e., the practice problem for multiplication after the worked example for multiplication).

Separated information for tasks with low element interactivity comprised two sections. The first section presented a worked example for each operation, and each example had two blocks: one block contained the numbered step descriptions, and the second block contained the corresponding numbered solved steps. Learners needed to pair each step description with its corresponding solution. The second section consisted of four practice problems, one per operation. Learners had to complete or solve each practice problem by integrating it with the respective worked example provided in the first section. Separating blocks and sections of information imposed a higher extraneous load compared to the integrated information condition.

For tasks with high element interactivity, the solution steps were removed. The material used for both individual and group settings was identical but varied in element interactivity and information integration according to the condition. However, groups were instructed to interact to study and solve the problems.

2.4. Performance Measure

During the pretest stage, learners were assessed with a written mathematics test comprising four tasks related to the operations of addition, subtraction, multiplication, and division of complex numbers, a new topic not yet taught, which they had to solve. In the learning stage, learners received instructional guidance containing the topic of complex numbers with four worked examples and the respective procedure related to the operations of addition, subtraction, multiplication, and division, as well as four practice problems to solve according to their condition. The academic performance of learners who worked with high element interactivity was graded out of 13 points. Participants had to solve four tasks with a total of 13 steps. Participants were awarded 1 point for each step answered correctly, 0 points if the step was not completed or the answer was incorrect, and a proportional score if the step was partially resolved. For students who worked with low element interactivity (i.e., they had to solve partially worked examples), they could earn 6.5 points: 0.5 points for each step solved correctly, 0 points if the step was not completed or the answer was

incorrect, and a proportion if the step was partially resolved. In the immediate assessment, participants received a booklet with four problems to solve one day after the learning stage. In the delayed test, participants received another booklet with four problems eight days after the learning stage.

In the pretest, immediate posttest, and delayed posttest, participants received a document with learning activities containing the procedures for the operations of addition, subtraction, multiplication, and division of complex numbers with the same level of complexity as the learning tasks, along with their respective assessment criteria. These tasks were performed independently by the participants. They were also instructed to complete all practice tasks provided to them in the learning stage.

3. Results

Analyses were conducted using SPSS version 26 and a 2 (element interactivity: high vs. low) $\times$ 2 (social condition: individual vs. group) $\times$ 2 (format: separated information vs. integrated information) analysis of variance (ANOVA). The dependent variable was academic performance, measured during the learning, immediate test, and delayed test stages. A significance level of 0.05 was used for all analyses. Partial eta squared (η_p^2) was used as the effect size measure, with values of 0.01, 0.06, and 0.14 corresponding to small, medium, and large effects, respectively [33]. The levels of element interactivity in the learning stage were analyzed separately.

3.1. Learning

Descriptive results for the learning stage are shown in Table 1. Regarding tasks with low element interactivity, the ANOVA showed that the main effect of material was statistically significant in favor of material with integrated information ($M = 6.01$, $SD = 0.47$, $F(1, 92) = 13.55$, $MSE = 0.50$, $p < 0.01$, $\eta_p^2 = 0.13$) compared to separated information ($M = 5.45$, $SD = 0.90$). The main effect of the social condition was also significant ($F(1, 92) = 5.05$, $MSE = 0.50$, $p = 0.03$, $\eta_p^2 = 0.05$), suggesting that group performance ($M = 5.84$, $SD = 0.79$) was better than individual performance ($M = 5.50$, $SD = 0.67$). However, the interaction between main effects was not significant ($p = 1.00$).

Table 1. Means and standard deviations for performance in the learning stage.

	Integrated Information		Separated Information	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Low element interactivity				
Individual	5.78	0.55	5.22	0.68
Group	6.12	0.38	5.56	0.98
High element interactivity				
Individual	12.00	1.21	9.63	2.80
Group	9.44	1.72	11.38	1.64

Regarding performance on tasks with high element interactivity, the ANOVA showed that the main effects of the material and social condition were not significant ($p = 0.59$ and $p = 0.31$, respectively). However, the interaction between these effects was significant ($F(1, 92) = 28.96$, $MSE = 3.43$, $p < 0.01$, $\eta_p^2 = 0.24$). Bonferroni post hoc tests suggest that material with integrated information is more beneficial for the individual condition than for the group condition ($p < 0.01$, $\eta_p^2 = 0.18$) and that separated information favors the group condition more than the individual condition ($p < 0.01$, $\eta_p^2 = 0.09$).

3.2. Immediate Test

The ANOVA (Table 2) revealed that only the main effect of element interactivity was significant, suggesting that low element interactivity ($M = 9.92$, $SD = 0.31$) resulted in

superior performance compared to high element interactivity ($M = 7.5$, $SD = 0.31$). The remaining main effects were not statistically significant.

Table 2. Two-way ANOVA for immediate test.

Source	MS	F (1184)	p	η_p^2
Element interactivity (E)	250.26	29.77	<0.01	0.14
Material (M)	1.50	0.18	0.67	0.01
Social condition (S)	30.38	3.61	0.06	0.02
E $\times$ M	0.01	0.01	0.97	<0.01
E $\times$ S	495.04	58.88	<0.01	0.24
M $\times$ S	106.26	12.64	<0.01	0.06
E $\times$ M $\times$ S	117.04	13.92	<0.01	0.07
Error	8.41			

Regarding the significant interactions between main effects (Table 2), Bonferroni post hoc tests for the interaction between element interactivity and social condition showed that for tasks with low element interactivity, learning individually ($p < 0.01$, $\eta_p^2 = 0.20$, $M = 12.05$, $SD = 0.51$) was more effective than learning in groups ($M = 7.80$, $SD = 0.36$). However, for tasks with high element interactivity, learning in groups ($p < 0.01$, $\eta_p^2 = 0.08$, $M = 8.78$, $SD = 0.36$) was better than learning individually ($M = 6.22$, $SD = 0.51$).

Bonferroni tests for the interaction between the material and social condition indicated that material with integrated information benefits individual learners more ($p < 0.01$, $\eta_p^2 = 0.08$, $M = 9.83$, $SD = 0.51$) than groups ($M = 7.41$, $SD = 0.36$). However, material with separated information produced similar results for both individual and group learning ($p = 2.4$).

The interaction between element interactivity, social condition, and material was significant. This suggests that tasks with low element interactivity, whether with integrated information ($p < 0.01$, $\eta_p^2 = 0.11$) or separated information ($p < 0.01$, $\eta_p^2 = 0.11$), benefit individual learners more than groups (Table 3). Tasks with high element interactivity and integrated information benefited both individual and group learners equally ($p = 0.45$). However, separated information proved to be better for group learners ($p < 0.01$, $\eta_p^2 = 0.19$).

Table 3. Means and standard deviations for immediate test performance.

	Integrated Information		Separated Information	
	M	SD	M	SD
Low element interactivity				
Individual	11.91	1.13	12.19	1.57
Group	7.73	2.55	7.86	3.93
High element interactivity				
Individual	7.75	4.41	4.69	3.98
Group	7.08	2.64	10.48	1.43

3.3. Delayed Test

The ANOVA (Table 4) showed that the main effect of the social condition was significant and suggests that individual learning ($M = 7.87$, $SD = 0.41$) resulted in better performance than group learning ($M = 6.06$, $SD = 0.29$). The remaining main effects were not statistically significant.

Regarding statistically significant main-effect interactions (Table 4), Bonferroni tests for the interaction between element interactivity and material indicated that for low element interactivity, integrated information ($M = 7.91$, $SD = 0.51$) is more effective than separate information ($p < 0.01$, $\eta_p^2 = 0.05$, $M = 5.70$, $SD = 0.51$). However, for high element interactivity, both types of material are equally effective ($p = 0.52$).

Bonferroni tests on the interaction between the material and social condition indicated that integrated information material benefits individual learners more ($p < 0.01$, $\eta_p^2 = 0.13$, $M = 9.27$, $SD = 0.58$) than those learning in groups ($M = 5.53$, $SD = 0.41$). However, both learning conditions equally achieved similar results with separated information material ($p = 0.87$).

Table 4. Two-way ANOVA for delayed test.

Source	MS	F (1184)	p	η_p^2
Element interactivity (E)	4.27	0.39	0.53	<0.01
Material (M)	32.38	2.97	0.09	0.02
Social condition (S)	139.56	12.78	<0.01	0.07
E $\times$ M	75.70	6.94	0.01	0.04
E $\times$ S	9.85	0.90	0.34	<0.01
M $\times$ S	158.24	14.50	<0.01	0.07
E $\times$ M $\times$ S	0.71	0.07	0.80	<0.01
Error	10.92			

4. Discussion

Developing problem-solving skills is gaining popularity among those advocating for the existence of key knowledge for contemporary society [34,35]. Meanwhile, in classrooms, instructional decisions for problem-solving learning often combine collaborative and individual learning activities with the support of a wide variety of instructional materials. In this context, the decision between individual and group learning to promote the acquisition of problem-solving skills will be more accurate and effective if the characteristics of human cognition, the task, and the learning material are considered [36]. Consequently, the purpose of this study was to examine how the level of interactivity of task elements (i.e., low (partially worked examples) versus high (problem-solving)) and split attention (i.e., material with integrated and separated information) interact in individual and collaborative learning.

Our hypotheses regarding tasks with low element interactivity, such as partially worked examples, assumed that these induce low cognitive load, leaving space available to store task information in long-term memory and cope with other sources of cognitive load. Therefore, we expected that individual learners would exhibit superior performance compared to groups, whether they studied materials that induced split attention (h1.1) or those that did not (h1.2). Immediate and delayed tests provide evidence supporting both hypotheses. These results are consistent with previous studies on the split-attention effect [37–39] and low-element-interactivity tasks (i.e., worked examples) [26,28]. This finding suggests that the attentional resources invested during the individual learning stage may have been sufficient to cope with the intrinsic cognitive demand of the learning task and the extraneous load associated with separated information [7]. The low element interactivity induced by partially solved tasks probably induced lower activity in working memory or cognitive load. This perhaps allowed individual learners to integrate the separated information from the material without exceeding the total capacity of working memory.

On the other hand, collaborative groups had to invest additional working memory resources in transactional activities [5,32]. Activities such as sharing problem information, performing shared calculations, or discussing the result of a task step [32,40] likely further increased working memory activity. Prior research about resource interdependence suggests that learners can combine their resources to achieve their goals [41–43]. Group members may have effectively shared their working memory resources to meet the cognitive demand of both the interactions and mental operations involved in integrating separate sources of information. Regarding the latter, it is plausible that group members may have coped with the cognitive costs of split attention by sharing elements of the material that were physically separated. In other words, by sharing task information,

the negative effect of split attention led to increased collaborative interactions. However, transactional activities may have consumed a significant amount of attentional resources to the point that they interfered with the storage of information in long-term memory [25]. Apparently, as previous research suggests [5,24,26,28], the advantages of being able to distribute the cognitive load of a task among group members and having a larger reservoir of collective memory resources are not justified in tasks with low element interactivity. Transactional activities in undemanding learning tasks impose cognitive costs that seem to inhibit information acquisition [32].

The lack of evidence for these hypotheses during the learning phase might be explained by factors associated with the low element interactivity. Although individual learners performed better than groups in the subsequent tests, their similar performance to groups during the learning tasks may be due to their better information acquisition [6] being equivalent to the group performance associated with the extended working memory capacity [5]. It appears that the groups leveraged their interactions to complete the tasks, although this did not translate into better retention. The greater knowledge acquired by individual learners and the higher cognitive capacity of the groups may explain why the split-attention material did not affect performance during the learning phase, as we predicted in our hypothesis.

Regarding our hypotheses about the effects of tasks with a high level of element interactivity (i.e., problem-solving), we expected them to consume substantial mental resources to the point of interfering with the storage of task information in long-term memory and the handling of extraneous cognitive load. Therefore, it was hypothesized that tasks with high element interactivity with materials containing integrated information would benefit individual learners more than groups because groups have extra-high cognitive costs associated with transactional activities (h2.1). Only the results of the learning stage provided evidence for this hypothesis. Apparently, during the learning process, materials with integrated information allowed individual learners to invest more working memory resources in acquiring information compared to groups. Problem-solving required a series of steps following a worked example. The worked example with integrated information may have reduced the number of working memory activities, leaving more mental resources to solve the learning tasks [44,45]. However, the quality of the acquired mental representations was not enough to solve problems on subsequent tests. It is plausible that, although the worked example with integrated information may have reduced the cognitive load during learning, solving each of the steps following the example was still cognitively demanding. As a result, working memory resources were focused on problem-solving, leaving few attentional resources available to create lasting learning [46].

Unlike individual learners, collaborative groups had to invest more working memory resources in the learning stage due to transactional activities. These activities may include distributing the elements of the task or discussing how to complete partial solutions, imposing additional information processing on working memory [5,32]. Although groups may have been able to handle the collaborative load due to their extended working memory capacity, previous research [32] provides some evidence suggesting that some interactions may have interfered with information retention processes (e.g., task-unrelated conversations), leading to a lower performance. Nonetheless, group work resulted in learning equivalent to that of individual learners in the subsequent test due to operational transactive discussions, such as sharing information elements and calculations or explaining solutions [47,48]. These interactions may have resulted in enduring task elaborations similar to those of individual learners.

Regarding tasks with high element interactivity with materials that induce split attention, according to the collective working memory effect [5,31], it was estimated that collaborative groups would leverage their large memory resources to meet the demands of task information, as well as transactional activities and distributed information. Consequently, we expected that groups would perform better than individuals (h2.2). The findings from both the learning stage and the immediate test support this hypothesis.

Apparently, groups more effectively managed the attentional load associated with the task, interactions, and material through a distributive advantage and collective working memory [23]. Because of this, the cognitive load associated with split attention was due to the physical separation not only between the step descriptions but also between the worked example and the practice problem. It is plausible that group members might have decided to process the high cognitive demand of separated information by productive coordination interactions (i.e., mutual cognitive interdependence principle) [8,32,49]. Previous studies suggest that group members implicitly or explicitly agree to process separated sources of information and integrate them with productive interactions such as verbally shared calculations [32,40,42]. Furthermore, it is reasonable to suggest that groups agree to process separated information by having one member review the information from the material at hand while another integrates that information with what they have in another section of the material. In this way, the larger working memory capacity, along with more productive transactional activities, likely induced the creation of durable mental representations in long-term memory.

5. Conclusions

In the educational context, learning to solve problems can encompass both individual and collaborative tasks across a wide variety of materials and domains. This study provides a contribution to the current evidence indicating that the outcomes of individual and group learning are mediated by the characteristics of the task and the learning material. Tasks with a low level of element interactivity (e.g., partially worked examples), presented either integrated or spatially separated (inducing or not inducing split attention, respectively), result in similar performance, whether learned individually or in groups. However, tasks with a high level of element interactivity (i.e., conventional problems), whose information is integrated, seem to benefit individual learning processes, but the test results are similar to those of collaborative groups. Nevertheless, if the information in the material is physically separated, inducing split attention, groups learn more due to their greater collective working memory capacity. These findings were observed only in immediate tests; in tests delayed by 8 days, the performance was similar.

This study has limitations and implications for future research. One limitation was that we did not conduct an analysis to explore the transactional activities that may explain the results of the subsequent tests. Future studies should confirm our findings and include analyses, for example, of the types of interactions deployed to process material with separated information. Split attention is likely to evoke more regulatory management interactions [32,50], as is the case with individual learning [51].

Our study was conducted with pairs of learners in order to keep interactions focused on the task [52]. However, in classrooms, there may be groups with varying numbers of members. A future line of research may estimate the limits of the distributive advantage and collective working memory of groups in high-element-interactivity tasks considering the number of members per group [8]. An increase in the number of members would increase both the group's processing capacity and interactions, but this may reduce responsibility and effort in learning the task [53]. Further, it is surprising that the evidence from the 8-day test did not support our hypotheses. Nevertheless, this opens up an opportunity for future research to explore the cognitive load factors associated with group interactions that can enhance long-term retention, such as collaborative retrieval practice [54] and prior knowledge [6]. Additionally, further studies will be valuable in confirming our hypotheses.

Finally, the results of this study have implications for education professionals. Collaborative learning is a strategy that can be used to foster the acquisition of problem-solving skills in several curriculum areas, such as mathematics, computer science, or writing. However, grouping students to work together and solve problems requires a significant amount of time, effort, and resources and does not always result in learning gains compared to individual learning. This study suggests that acquiring domain-specific problem-solving skills should carefully consider the complexity of the task and the spatial distribution of

the information in the study material. Individual or collaborative learning can be effective when the task information is not very complex, regardless of whether it induces split attention or not. However, it is important to consider the spatial distribution of the information when tasks are highly complex. If the material requires integrating many sources of physically separated information, it is more effective to promote problem-solving learning with collaborative groups.

Author Contributions: Conceptualization, J.Z.R.; methodology, J.Z.R. and J.G.; formal analysis, J.Z.R.; writing—original draft preparation, J.Z.R. and J.G.; writing—review and editing, J.Z.R. and J.G.; supervision, J.Z.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Universidad Del Pacífico (form code GP-FR-T-01 and date of approval 11 March 2023).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are available on request from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Graesser, A.C.; Fiore, S.M.; Greiff, S.; Andrews-Todd, J.; Foltz, P.W.; Hesse, F.W. Advancing the science of collaborative problem solving. *Psychol. Sci. Public Interest* **2018**, *19*, 59–92. [CrossRef] [PubMed]
2. Slavin, R.E. Cooperative learning in elementary schools. *Education* **2014**, *43*, 5–14. [CrossRef]
3. Hmelo-Silver, C.; Chinn, C.A. Collaborative learning. In *Handbook of Educational Psychology*; Corno, L., Anderman, E.M., Eds.; Routledge: New York, NY, USA, 2015; pp. 349–363.
4. Sweller, J.; Ayres, P.; Kalyuga, S. *Cognitive Load Theory*; Springer: New York, NY, USA, 2011. [CrossRef]
5. Kirschner, F.; Paas, F.; Kirschner, P.A. Task complexity as a driver for collaborative learning efficiency: The collective working-memory effect. *Appl. Cogn. Psychol.* **2011**, *25*, 615–624. [CrossRef]
6. Zambrano, J.; Kirschner, F.; Sweller, J.; Kirschner, P.A. Effects of prior knowledge on collaborative and individual learning. *Learn. Instr.* **2019**, *63*, 101214. [CrossRef]
7. Ayres, P.; Sweller, J. The split-attention principle in multimedia learning. In *The Cambridge Handbook of Multimedia Learning*; Mayer, R.E., Fiorella, L., Eds.; Cambridge University Press: New York, NY, USA, 2022; pp. 199–211. [CrossRef]
8. Kirschner, P.A.; Sweller, J.; Kirschner, F.; Zambrano, J. From cognitive load theory to collaborative cognitive load theory. *Int. J. Comput.-Support. Collab. Learn* **2018**, *13*, 213–233. [CrossRef]
9. Baddeley, A.D. Working memory: Theories, models, and controversies. *Annu. Rev. Psychol.* **2011**, *63*, 1–29. [CrossRef]
10. Gilchrist, A.L.; Cowan, N. Can the focus of attention accommodate multiple, separate items? *J. Exp. Psychol. Learn. Mem. Cogn.* **2011**, *37*, 1484. [CrossRef]
11. Chen, O.; Paas, F.; Sweller, J. A cognitive load theory approach to defining and measuring task complexity through element interactivity. *Educ. Psychol. Rev.* **2023**, *35*, 63. [CrossRef]
12. Sweller, J. Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educ. Psychol. Rev.* **2010**, *22*, 123–138. [CrossRef]
13. Chen, O.; Kalyuga, S.; Sweller, J. The expertise reversal effect is a variant of the more general element interactivity effect. *Educ. Psychol. Rev.* **2016**, *29*, 393–405. [CrossRef]
14. Tricot, A.; Sweller, J. Domain-specific knowledge and why teaching generic skills does not work. *Educ. Psychol. Rev.* **2014**, *26*, 265–283. [CrossRef]
15. Kalyuga, S. Knowledge elaboration: A cognitive load perspective. *Learn. Instr.* **2009**, *19*, 402–410. [CrossRef]
16. Sweller, J.; Van Merriënboer, J.J.G.; Paas, F. Cognitive architecture and instructional design: 20 years later. *Educ. Psychol. Rev.* **2019**, *31*, 261–292. [CrossRef]
17. Peterson, L.R.; Peterson, M.J. Short-term retention of individual verbal items. *J. Exp. Psychol.* **1959**, *58*, 193–198. [CrossRef] [PubMed]
18. Webb, N.M. Task-related verbal interaction and mathematics learning in small groups. *J. Res. Math. Educ.* **1991**, *22*, 366–389. [CrossRef]
19. Webb, N.M.; Burnheimer, E. Cooperative and collaborative learning. In *International Encyclopedia of Education*; Elsevier: Amsterdam, The Netherlands, 2023; pp. 593–599. [CrossRef]
20. Crook, C. CSsCL: The performance of collaborative learning. *Int. J. Comput.-Support. Collab. Learn* **2022**, *17*, 169–183. [CrossRef]
21. Nokes-Malach, T.; Richey, J.E.; Gadgil, S. When is it better to learn together? Insights from research on collaborative learning. *Educ. Psychol. Rev.* **2015**, *27*, 645–656. [CrossRef]

22. Tindale, R.S.; Kameda, T. 'Social sharedness' as a unifying theme for information processing in groups. *Group Process. Intergroup Relat.* **2000**, *3*, 123–140. [CrossRef]
23. Hinsz, V.B.; Tindale, R.S.; Vollrath, D.A. The emerging conceptualization of groups as information processors. *Psychol. Bull.* **1997**, *121*, 43–64. [CrossRef]
24. Zambrano, J.; Kirschner, F.; Sweller, J.; Kirschner, P.A. Effects of group experience and information distribution on collaborative learning. *Instr. Sci.* **2019**, *47*, 531–550. [CrossRef]
25. Kirschner, F.; Paas, F.; Kirschner, P.A.; Janssen, J. Differential effects of problem-solving demands on individual and collaborative learning outcomes. *Learn. Instr.* **2011**, *21*, 587–599. [CrossRef]
26. Retnowati, E.; Ayres, P.; Sweller, J. Can collaborative learning improve the effectiveness of worked examples in learning mathematics? *J. Educ. Psychol.* **2016**, *109*, 666–679. [CrossRef]
27. Zambrano, J.; Kirschner, F.; Kirschner, P.A. The effect of the prior collaborative experience on the effectiveness and efficiency of collaborative learning. In *Rethinking Learning in the Digital Age: Making the Learning Sciences Count, Proceedings of the 13th International Conference of the Learning Sciences (ICLS), London, UK, 23–27 June 2018*; Kay, J., Luckin, R., Eds.; International Society of the Learning Sciences: London, UK, 2018; pp. 112–119.
28. Retnowati, E.; Ayres, P.; Sweller, J. Worked example effects in individual and group work settings. *Educ. Psychol.* **2010**, *30*, 349–367. [CrossRef]
29. Sweller, J.; Ayres, P.; Kalyuga, S. The split-attention effect. In *Cognitive Load Theory*; Sweller, J., Ayres, P., Kalyuga, S., Eds.; Springer: London, UK, 2011; pp. 111–128. [CrossRef]
30. De Koning, B.B.; Rop, G.; Paas, F. Effects of spatial distance on the effectiveness of mental and physical integration strategies in learning from split-attention examples. *Comput. Hum. Behav.* **2020**, *110*, 106379. [CrossRef]
31. Kirschner, F.; Paas, F.; Kirschner, P.A. A cognitive load approach to collaborative learning: United brains for complex tasks. *Educ. Psychol. Rev.* **2009**, *21*, 31–42. [CrossRef]
32. Zambrano, J.; Kirschner, F.; Sweller, J.; Kirschner, P.A. Effect of task-based group experience on collaborative learning: Exploring the transaction activities. *Br. J. Educ. Psychol.* **2023**, *93*, 879–902. [CrossRef]
33. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; L. Erlbaum Associates: Hillsdale, NJ, USA, 1988; p. xxi, 567p.
34. Han, A.; Krieger, F.; Greiff, S. Assessment of collaborative problem-solving: Past achievements and current challenges. In *International Encyclopedia of Education*; Elsevier: Amsterdam, The Netherlands, 2023; pp. 234–244. [CrossRef]
35. Karaca-Atik, A.; Meeuwisse, M.; Gorgievski, M.; Smeets, G. Uncovering important 21st-century skills for sustainable career development of social sciences graduates: A systematic review. *Educ. Res. Rev.* **2023**, *39*, 100528. [CrossRef]
36. Zhang, L.; Kirschner, P.A.; Cobern, W.W.; Sweller, J. There Is an Evidence Crisis in Science Educational Policy. *Educ. Psychol. Rev.* **2022**, *34*, 1157–1176. [CrossRef]
37. Schroeder, N.L.; Ceneci, A.T. Spatial contiguity and spatial split-attention effects in multimedia learning environments: A meta-analysis. *Educ. Psychol. Rev.* **2018**, *30*, 679–701. [CrossRef]
38. Noetel, M.; Griffith, S.; Delaney, O.; Harris, N.R.; Sanders, T.; Parker, P.; del Pozo Cruz, B.; Lonsdale, C. Multimedia design for learning: An overview of reviews with meta-meta-analysis. *Rev. Educ. Res.* **2021**, *90*, 6–23. [CrossRef]
39. Chandler, P.; Sweller, J. Cognitive load while learning to use a computer program. *Appl. Cogn. Psychol.* **1996**, *10*, 151–170. [CrossRef]
40. Webb, N.M.; Troper, J.D.; Fall, R. Constructive activity and learning in collaborative small groups. *J. Educ. Psychol.* **1995**, *87*, 406. [CrossRef]
41. Johnson, D.W. Social Interdependence: Interrelationships among Theory, Research, and Practice. *Am. Psychol.* **2003**, *58*, 934–945. [CrossRef] [PubMed]
42. Wong, R.M.; Adesope, O.O. Promoting collaboration through resource interdependence to enhance collaborative concept mapping. *Soc. Psychol. Educ.* **2023**, *27*, 89–113. [CrossRef]
43. Butera, F.; Buchs, C. Social interdependence and the promotion of cooperative learning. In *Social Psychology in Action*; Sassenberg, K., Vliek, M.L.W., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 111–127. [CrossRef]
44. Renkl, A.; Atkinson, R.K.; Plass, J.L.; Moreno, R.; Brunken, R. Learning from worked-out examples and problem solving. In *Cognitive Load Theory*; Cambridge University Press: Cambridge, UK, 2010; pp. 91–108. [CrossRef]
45. Chen, O.; Retnowati, E.; Chan, B.K.Y.; Kalyuga, S. The effect of worked examples on learning solution steps and knowledge transfer. *Educ. Psychol.* **2023**, *43*, 914–928. [CrossRef]
46. Soderstrom, N.C.; Bjork, R.A. Learning versus performance: An integrative review. *Perspect. Psychol. Sci.* **2015**, *10*, 176–199. [CrossRef]
47. Webb, N.M. Student interaction and learning in small groups. In *Learning to Cooperate, Cooperating to Learn*; Slavin, R.E., Sharan, S., Kagan, S., Hertz-Lazarowitz, R., Webb, C., Schmuck, R., Eds.; Plenum Press: New York, NY, USA, 1985; pp. 147–172. [CrossRef]
48. Jurkowski, S.; Mundelse, L.; Hänze, M. Strengthening collaborative learning in secondary school: Development and evaluation of a lesson-integrated training approach for transactive communication. *Learn. Instr.* **2024**, *92*, 101934. [CrossRef]
49. Janssen, J.; Kirschner, P.A. Applying collaborative cognitive load theory to computer-supported collaborative learning: Towards a research agenda. *Educ. Technol. Res. Dev.* **2020**, *68*, 783–805. [CrossRef]
50. Malmberg, J.; Järvelä, S.; Järvenoja, H. Capturing temporal and sequential patterns of self-, co-, and socially shared regulation in the context of collaborative learning. *Contemp. Educ. Psychol.* **2017**, *49*, 160–174. [CrossRef]

51. Eitel, A.; Endres, T.; Renkl, A. Self-management as a bridge between cognitive load and self-regulated learning: The illustrative case of seductive details. *Educ. Psychol. Rev.* **2020**, *32*, 1073–1087. [CrossRef]
52. Zhan, Z.; He, G.; Li, T.; He, L.; Xiang, S. Effect of groups size on students' learning achievement, motivation, cognitive load, collaborative problem-solving quality, and in-class interaction in an introductory AI course. *J. Comput. Assist. Learn.* **2022**, *38*, 1807–1818. [CrossRef]
53. Latané, B.; Williams, K.; Harkins, S. Many hands make light the work: The causes and consequences of social loafing. *J. Personal. Soc. Psychol.* **1979**, *37*, 822–832. [CrossRef]
54. Wissman, K.T.; Rawson, K.A. Why does collaborative retrieval improve memory? Enhanced relational and item-specific processing. *J. Mem. Lang.* **2015**, *84*, 75–87. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Best of Both Worlds? Combining Physical and Mental Self-Management Strategies to Support Learning from Split-Attention Examples

Björn B. de Koning

Department of Psychology, Education and Child Studies, Erasmus School of Social and Behavioural Sciences, Erasmus University Rotterdam, 3062 PA Rotterdam, The Netherlands; b.b.dekoning@essb.eur.nl

Abstract: The self-management principle holds that higher learning performance is obtained when learners actively use instructional strategies to manage the working memory load imposed by a learning task. Self-management studies with spatially separated but mutually referring text and pictures (split-attention examples) demonstrate the learning benefits of physical (e.g., annotation) and mental (imagined drag-and-drop) strategies. We investigated whether combining physical and mental strategies supports learning beyond a single strategy. Eighty-four participants studied a split-attention example with or without using a physical strategy and/or a mental strategy. Participants completed retention, comprehension, and transfer tests, and rated their cognitive load. Results showed that the combined use of physical and mental strategies resulted in lower cognitive load during learning than using the physical strategy and was more instructionally efficient compared to all other conditions. There were no significant differences regarding learning outcomes. Together, this suggests that combining physical and mental strategies is most supportive for studying split-attention examples.

Keywords: self-management of cognitive load; self-management strategies; split-attention examples; cognitive load theory; comprehension of text and graphics

Citation: de Koning, B.B. Best of Both Worlds? Combining Physical and Mental Self-Management Strategies to Support Learning from Split-Attention Examples. *Educ. Sci.* **2024**, *14*, 284. <https://doi.org/10.3390/educsci14030284>

Academic Editor: Myint Swe Khine

Received: 13 February 2024

Revised: 4 March 2024

Accepted: 6 March 2024

Published: 7 March 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The presentation of text and pictures for explaining concepts or procedures, such as the functioning of the heart or assembly of musical instruments, is by now quite common in formal (e.g., study books) and informal (e.g., websites) instructional resources. While multimedia learning research has yielded numerous guidelines for how to present text and pictures in a way that learner processing is optimized (for an overview, see [1]), many available instructional resources are still not presented in the most effective format. For example, mutually referring text and pictures are frequently presented in a spatially separated format (i.e., split-attention format), whereas research has shown that presenting text and pictures close to each other in space (i.e., integrated format) improves learning and reduces cognitive demands on the learner; this has become known as the split-attention effect or spatial contiguity effect [2,3]. Therefore, there is increasing attention on equipping learners with strategies to help them learn more effectively from such sub-optimally designed instructional materials [4]. The present study builds on this emerging line of research and extends it by investigating the relative effectiveness of such strategies and whether the combined use of two strategies is more effective for learning than using a single strategy.

2. The Self-Management Principle

The centrality of the learner to manage their cognitive load when studying text–picture materials is at the core of the self-management principle. The self-management principle holds that higher learning performance is obtained when learners actively use instructional strategies to manage the working memory load imposed by a learning task [5,6]. According to cognitive load theory (CLT; [1]), an instructional theory to optimize learning activities

and materials while taking into account the limited capacity of working memory, the provision of specific guidance to learners enables learners to reduce the high and unnecessary cognitive load that instructional materials impose and help them more effectively learn from these materials [4,6]. The emphasis on the learner as a manager of their own cognitive load contrasts with the decades-long practiced “instructor-managed approach” wherein teachers/designers made decisions—based on CLT—on how to present learning materials to optimize learning [5]. The “self-management approach” is a relatively unexplored direction in CLT research, but an increasing number of studies have investigated what self-management strategies can be used to effectively support learning.

2.1. Self-Management with Physical and Mental Strategies

Prior studies investigating the self-management of cognitive load demonstrate that learning from mutually referring but spatially separated text and pictures (i.e., split-attention examples) is enhanced when learners use a physical or a mental self-management strategy compared to when they do not use these strategies (for an overview see [6]). Specifically, physical strategies that require learners to identify the key information in the text and connect this with the corresponding element in the picture using (1) annotation, such as drawing connecting arrows, numbering, and underlining [5,7]; (2) dragging and dropping text into the picture [8]; or (3) finger pointing (e.g., [9]) have been shown to support learning outcomes such as retention, recognition, and transfer performance and/or to reduce cognitive load. However, there is also research showing no benefits for learning and cognitive load when using a physical self-management strategy, and this appears to be especially the case when instructing learners to only physically move text boxes to the corresponding element in the accompanying picture [10–12].

Research on the mental self-management strategy shows clear and unequivocal learning benefits, consistent with the benefits of using an imagination strategy in, for example, reading comprehension tasks [13]. In studies by De Koning et al. [11,12], learners using the strategy of imagining dragging and dropping text boxes to the corresponding element in the picture had higher retention and comprehension scores than learners using a physical drag-and-drop strategy and learners not using any strategy. No differences were found in the cognitive load scores. The authors explained the differences in learning by arguing that the mental self-management strategy is possibly more generative in nature, encouraging learners to engage in active and meaningful cognitive processing (germane cognitive load), whereas the act of physically moving text to a specific location in the picture might be performed without any cognitive processing. This raises the question of whether the same findings would be obtained if the mental self-management strategy were compared to a physical self-management strategy that has proven effective and might elicit more generative processing in learners. One such physical self-management strategy that was successfully used in prior self-management studies to support learning from text and pictures in a split-attention format [5,7] is making annotations. Using an annotation strategy has been shown to be an effective strategy in other fields of study, as well such as learning from text (e.g., [14]). While there exist obvious differences between the mental drag-and-drop strategy and the annotation strategy (e.g., in terms of physical involvement), with both, the focus is on reducing the search-and-match processes for corresponding text-picture information (extraneous cognitive load) and encouraging learners to actively make sense of the presented information in order to construct an accurate mental representation of the content (germane cognitive load).

One aim of the present study was, therefore, to directly compare the learning effectiveness of the mental self-management strategy and that of the annotation self-management strategy. So far, prior self-management research has focused primarily on investigating whether self-management strategies are effective for learning, and much less attention has been devoted to investigating the relative effectiveness of different strategies. Having this focus in the present study contributes to further understanding of the effective use of self-management strategies in split-attention examples.

2.2. Combining Physical and Mental Strategies

Another aim of the present study was to investigate whether combining physical and mental self-management strategies would further enhance the effectiveness of self-managing the cognitive load during learning from split-attention examples. The study by Gordon and colleagues [15] represents the only reported attempt to investigate the effects of combining two self-management strategies when studying split-attention examples on learning and cognitive load. For the combined strategy condition, they instructed participants to use the physical drag-and-drop strategy together with the annotation strategy (highlighting, drawing arrows). They compared this to a control condition not using any strategy to study and a condition where the text and picture information was spatially integrated. The results showed that there were no significant differences between the self-management condition and control condition on recognition or transfer performance, nor on cognitive load. While, as argued by the Gordon et al. [15], the lack of an effect might be due to the low number of participants, there is another explanation that is more directly connected to the self-management strategies themselves. In the self-management condition, learners were encouraged to engage in two physical strategies, and it is possible that this was not supportive for learning. For example, using one physical strategy and then trying to engage in an additional strategy in the same modality (physical movement) might cause interference between strategies in such a way that engaging in one physical strategy prevents one from engaging in the other physical strategy. Additionally, coordination costs associated with alternating the two physical strategies might unnecessarily increase the cognitive load on the learner, which could hinder learning [16]. Moreover, the requirement to use two physical strategies has possibly led learners to believe that behavioral involvement was the main focus of their task, which might have occurred at the expense of any more substantial cognitive involvement.

In the present study, we therefore investigated whether the combination of a physical and a mental self-management strategy might be a more effective combination to support learning from split-attention examples. Engaging in mental and physical strategies eliminates the competition between two physical strategies, and, moreover, offers the opportunity to execute the two strategies in tandem or simultaneously. That is, while using the physical annotation strategy to highlight relevant information in the text and then the corresponding element in the picture, learners could use the mental drag-and-drop strategy to mentally integrate the text into the picture, as if creating a personal spatially integrated format in their mind that they can continuously revise and update with new incoming information.

Evidence outside the context of self-management strategies and split-attention examples suggests that the combination of a mental strategy and a physical strategy might indeed be an effective combination to improve performance. For example, combining mental and physical training is an effective means to increase motor skill learning (e.g., [17]), enhance executive functioning (e.g., [18]), and support comprehension of text [13]. Also, in learning from mutually referring graphical and textual information, there is preliminary evidence that combining physical and mental strategies supports learning [19]. In the study by Wang et al. [19], it was investigated whether performance on a mathematics test was impacted by having studied mathematics examples by tracing out the study examples with hand movements, tracing out the examples with hand movements followed by imagining tracing out the examples with the eyes closed, and without any physical or imagined tracing. The results showed that participants in the combined tracing and imagination condition solved more mathematics examples correctly, and accomplished this more quickly and with less cognitive load than those in the control condition and physical-tracing-only condition. Translating these findings to the present study, it can be assumed that the combination of mental and physical self-management strategies might be particularly effective in split-attention examples, at least more effective than a condition not engaging in any self-management strategies and a condition that only uses a physical strategy. By examining these comparisons, the current study extends earlier self-management

research in the sense that now, insight is also given into the relative effectiveness of using a combination of two strategies versus using either of these strategies alone, instead of only comparing it to a no-strategy condition (as was the case in Gordon et al. [15]).

3. Hypotheses

To achieve the aims of the study outlined above, learners were presented with a learning task containing a mutually referring text and picture (i.e., split-attention example). They studied this split-attention example while engaging in a physical self-management strategy (annotation), a mental self-management strategy (imagined drag-and-drop), both physical and mental self-management strategies, or no physical or mental self-management strategies. After learning, the learners completed retention comprehension and transfer tests, and indicated their cognitive load after learning and after each test. Based on theoretical considerations from CLT and prior research, the following hypotheses were formulated. First, we made a prediction regarding the involvement in self-management strategies versus not engaging in such strategies. It was expected that, compared to not using any self-management strategy, using either the physical annotation self-management strategy or the mental self-management strategy in isolation, or the two strategies in combination, resulted in higher learning outcomes (Hypothesis 1a) and lower cognitive load (Hypothesis 1b). Second, we made a prediction regarding the relative effectiveness of combined vs. single self-management strategies. It was expected that the combination of the physical annotation strategy and the mental self-management strategy would result in higher learning outcomes (Hypothesis 2a) and lower cognitive load (Hypothesis 2b) than using either of these self-management strategies in isolation. Third, we were interested in the relative effectiveness of using a single self-management strategy. The annotation self-management strategy and the mental self-management strategy both outperformed a control condition without engagement in a self-management strategy, and there has been no direct comparison between self-management strategies in prior studies that can inform a directional hypothesis regarding the annotation strategy and the mental self-management strategy. The physical strategy used in the studies of De Koning et al. [11,12] did not support learning, yielding an unfair comparison to the mental strategy to which it was compared in that study. Therefore, we did not formulate a directional hypothesis, but were interested in exploring the relative effectiveness of the annotation self-management strategy and the mental self-management strategy.

4. Methods

4.1. Design

The study had a 2×2 between-subjects design, with the factors physical strategy (yes, no) and mental strategy (yes, no). Participants were randomly assigned to one of four conditions ($n = 21$ per condition): physical condition, mental condition, physical and mental condition, or control condition (without physical or mental strategies). The comparability of conditions regarding the participants' ages was tested using an univariate ANOVA, with condition as the independent variable and participant age as the dependent variable, which showed no significant differences between conditions: $F(3.83) = 0.49$, $p = 0.691$, $\eta_p^2 = 0.018$. The gender distribution across conditions was examined using the chi-square test, for which the chi-square goodness of fit indicated no significant differences: $\chi^2(6, N = 84) = 4.09$, $p = 0.665$. Together, the conditions were comparable regarding age and gender.

4.2. Participants

The participants comprised 84 students from various higher education institutes in the Netherlands. This number was justified by power calculations in G*Power (version 3.1.9.6; [20]), indicating that, using an alpha of 0.05 and 80% power, at least 80 participants were required in order to find a medium-sized effect (see [9]). Participants identified as female ($n = 61$),

male ($n = 22$), or non-binary ($n = 1$) and had a mean age of 23.77 years ($SD = 3.34$). They participated voluntarily and gave active consent before the study.

4.3. Materials

This study used the materials from De Koning et al. [12]. For all conditions, the tests were administered via Qualtrics, a digital survey tool, but the learning task was presented on paper. The digital test administration had several advantages, including automatic recording of participants' answers. The learning task was paper-based to enable participants in the conditions involving the physical strategy to make physical annotations. This avoided any potential unwanted effects for participants of having to learn to use a digital annotation tool for this study.

4.3.1. Prior Knowledge Test

Participants' prior knowledge about electrical circuits was assessed with a one-item self-rating asking them to indicate their knowledge of electrical circuits on a five-point scale (1—very little; 5—very much) and six statements about electrical circuits that required a yes/no answer (e.g., I know what a starter is). Each "yes" answer yielded one point, and no points were given to 'no' answers. Per participant, the scores on the self-rating item (0–5) and the statements (0–6) were added to one overall prior knowledge score (0–11). There were no significant differences in prior knowledge across the four conditions: $F(3.83) = 2.52$, $p = 0.064$, $\eta_p^2 = 0.086$.

4.3.2. Tests

Retention Test

The retention test assessed participants' memory of the studied electrical circuit's components and their position in the system. Five retention questions asked participants to name the symbol from the circuit that was shown in a picture. A correct answer yielded one point, while an incorrect answer was not awarded points. One question asked participants to draw the studied system from memory. One point was awarded to each component that was drawn in the right location and for each component that was correctly linked to another component. For this question, a score between 0 and 28 could be obtained. Per participant, the overall retention score was calculated by summing the scores on all retention questions (0–33).

Comprehension Test

The comprehension test contained 11 open-ended questions asking about the functioning of the studied electrical circuit. An example question is: "Which switches are pressed when the light is operating?" For each question, there was one correct answer that was given one point (incorrect answers were not awarded points). Per participant, all points were summed to obtain an overall comprehension score (0–11). Consistent with prior research using these materials [11,12,21], participants were given the picture depicting the on/off light-switching circuit (without text) on paper while answering the comprehension questions (this also was the case for the transfer questions).

Transfer Test

The transfer test contained six open-ended questions asking participants to reason regarding the studied electrical circuit. An example of a transfer question is: "After the stop button is released, the bell and the light start working again. What is the cause of this problem?". For each question, one point was obtained if the correct answer was given (no points for incorrect answers). Per participants, all points were summed to obtain an overall transfer score (0–6).

Cognitive Load

Participants' cognitive load during the learning task and tests was assessed with a 9-point rating scale (1—very, very little; 9—very, very much) originally developed by Paas [22].

4.4. Procedure

Participants were tested individually in a quiet room, and sat at a desk with a 24-inch computer screen (Eizo FlexScan, (Eizo, Hakusan, Japan)), keyboard, and mouse in front of them. The experimenter was present for the entire duration of the experiment and sat to the side of the participant. After a short welcome, participants completed the prior knowledge test. Then, they received instructions commensurate with their condition regarding studying the on/off-light switching circuit. They engaged in the practice task, and subsequently were given the paper for the learning task.

The learning task focused on the operation of an on/off light-switching circuit that was explained in a picture with accompanying text. In all conditions, the picture and text were presented in exactly the same way, in a split-attention format on A4-sized paper (portrait orientation). The picture and text were presented as far apart from each other as possible within the dimensions of the paper to increase the need for integration of the mutually referring picture and text. The paper was placed on a desk in front of the participant. In all conditions, participants were instructed to try to understand how the electrical system depicted in the picture and text worked. In the physical strategy condition, participants were required to support their learning by making annotations (i.e., circle and/or underline relevant pictorial and textual information; draw arrows and/or lines to connect textual and pictorial information) on the paper to help them connect the corresponding information in the text with that in the picture (cf. [5,7]). Participants were given a pen, pencil, and highlighter for annotating. In the mental strategy condition, participants had to support their learning by imagining that they moved a text segment (one at the time) to the corresponding part in the picture (cf. [12]). No annotation tools were available in this condition. In the physical and mental condition, participants engaged in both physical (annotation tool available) and mental strategies to support their learning. There was no specific guidance regarding whether or how to sequence the two strategies. In the control condition, participants studied the learning task without physical or mental strategies, and no annotation tools were provided. To familiarize participants with the strategies they had to use during the learning task, they engaged in a short practice task in which the to-be-used strategy/strategies was/were explained and practiced depending on the condition to which a participant belonged. To check for compliance with the instructions, we asked participants afterwards whether they had used the strategy/strategies they had to use to support their learning. All except three participants (one in the physical and mental condition, two in the mental strategy condition), indicated that they had used the physical and/or mental strategy/strategies. The response patterns of these three participants were comparable to those of the other participants, and including their data in the analyses yielded comparable results, so we retained all scores in the analyses so as not to give in terms of the power of the study. All participants were given four minutes to study the learning task.

After having completed the learning task, the experimenter took away the paper from which participants had studied, and participants rated their cognitive load. Thereafter, participants completed the retention test, comprehension test, and transfer test, and after each test, they rated their cognitive load. The entire experiment lasted about 40 min.

5. Results

Consistent with our 2×2 between-subjects design, separate univariate analyses of variance (ANOVAs), with physical strategy (yes, no) and mental strategy (yes, no) as between-subjects factors, were conducted in SPSS on the learning outcome measures (retention, comprehension, transfer) and cognitive load measures. Given the small sample and

mild skewness of the data, we also report (after the regular ANOVA statistics) bootstrapped ANOVAs, and we calculated the bias-corrected and accelerated confidence interval (BCa CI) estimates for the main effects and the interaction effects (including the simple effects analysis following a significant interaction). If a 95% confidence interval includes the null value, there was no statistically significant difference between conditions. For effect sizes, partial eta-squared (η_p^2) is reported where values of 0.01, 0.06, and 0.14 represent small, medium, and large effect sizes, respectively [23].

5.1. Learning Outcomes

Table 1 displays the means and standard deviations of the scores on the retention, comprehension, and transfer tests per condition.

Table 1. Means (standard deviations) per condition of the learning outcome measures.

	Physical	Mental	Physical and Mental	Control
Retention (0–33)	16.05 (5.38)	16.38 (5.96)	17.95 (5.84)	16.95 (5.52)
Comprehension (0–11)	5.38 (1.47)	5.24 (2.14)	5.71 (2.10)	5.95 (2.18)
Transfer (0–6)	2.81 (1.21)	2.38 (1.40)	2.95 (1.20)	2.57 (1.33)

On the retention test, there were no significant main effects of physical strategy ($F(1.80) = 0.07$, $p = 0.789$, $\eta_p^2 = 0.001$; $B = 0.168$, $p = 0.768$, BCa CI = $[-0.989, 1.411]$) or mental strategy ($F(1.80) = 0.29$, $p = 0.592$, $\eta_p^2 = 0.004$; $B = 0.335$, $p = 0.584$, BCa CI = $[-0.956, 1.478]$), and there was no significant physical $\times$ mental strategy interaction ($F(1.80) = 0.99$, $p = 0.321$, $\eta_p^2 = 0.012$; $B = 0.627$, $p = 0.315$, BCa CI = $[-0.508, 1.747]$).

On the comprehension test, there were no significant main effects of physical strategy ($F(1.80) = 0.01$, $p = 0.913$, $\eta_p^2 < 0.001$; $B = -0.024$, $p = 0.914$, BCa CI = $[-0.417, 0.371]$) or mental strategy ($F(1.80) = 0.19$, $p = 0.663$, $\eta_p^2 = 0.002$; $B = -0.096$, $p = 0.670$, BCa CI = $[-0.504, 0.278]$), and there was no significant physical $\times$ mental strategy interaction ($F(1.80) = 1.45$, $p = 0.232$, $\eta_p^2 = 0.018$; $B = 0.265$, $p = 0.231$, BCa CI = $[-0.130, 0.674]$).

On the transfer test, there were no significant main effects of physical strategy ($F(1.80) = 2.08$, $p = 0.153$, $\eta_p^2 = 0.025$; $B = 0.204$, $p = 0.153$, BCa CI = $[-0.067, 0.467]$) or mental strategy ($F(1.80) = 0.01$, $p = 0.933$, $\eta_p^2 < 0.001$; $B = -0.012$, $p = 0.943$, BCa CI = $[-0.274, 0.276]$), and there was no significant physical $\times$ mental strategy interaction ($F(1.80) = 0.35$, $p = 0.554$, $\eta_p^2 = 0.004$; $B = 0.084$, $p = 0.543$, BCa CI = $[-0.165, 0.309]$).

5.2. Cognitive Load

Table 2 displays the means and standard deviations for the cognitive load ratings reported after learning and after each of the learning outcome tests per condition.

Table 2. Means (standard deviations) per condition for the cognitive load ratings.

	Physical	Mental	Physical and Mental	Control
Cognitive load—learning	7.05 (1.50)	6.67 (1.43)	5.71 (1.93)	6.38 (1.72)
Cognitive load—retention	6.24 (2.17)	6.29 (1.95)	5.48 (2.23)	5.90 (1.97)
Cognitive load—comprehension	7.48 (1.25)	6.91 (1.48)	6.95 (1.47)	6.29 (1.62)
Cognitive load—transfer	7.42 (1.60)	6.95 (1.50)	6.95 (1.36)	6.48 (2.04)

For the cognitive load reported after learning, there was a significant physical $\times$ mental strategy interaction ($F(1.80) = 5.03$, $p = 0.028$, $\eta_p^2 = 0.059$; $B = -0.410$, $p = 0.019$, BCa CI = $[-0.749, -0.099]$). Simple effects analyses indicated that lower cognitive load ratings were given when physical and mental strategies were combined compared to when the physical strategy was not combined with the mental strategy ($p = 0.01$; BCa CI = $[-2.364, -0.295]$). There was no significant difference in cognitive load ratings between the conditions without the physical strategy, nor the mental strategy versus no strategy conditions

($p = 0.577$; BCa CI = $[-1.347, 0.779]$). There were no significant main effects for physical strategy ($F(1.80) = 0.16$, $p = 0.693$, $\eta_p^2 = 0.002$; $B = -0.072$, $p = 0.688$, BCa CI = $[-0.422, 0.270]$) or mental strategy ($F(1.80) = 2.11$, $p = 0.151$, $\eta_p^2 = 0.026$; $B = -0.263$, $p = 0.148$, BCa CI = $[-0.614, 0.084]$).

For the cognitive load reported after the tests, there were no significant main effects of physical strategy (retention: $F(1.80) = 0.27$, $p = 0.602$, $\eta_p^2 = 0.003$; $B = -0.120$, $p = 0.607$, BCa CI = $[-0.565, 0.314]$; comprehension: $F(1.80) = 3.78$, $p = 0.055$, $\eta_p^2 = 0.045$; $B = 0.311$, $p = 0.059$, BCa CI = $[-0.025, 0.611]$; transfer: $F(1.80) = 2.54$, $p = 0.115$, $\eta_p^2 = 0.031$; $B = 0.287$, $p = 0.106$, BCa CI = $[-0.038, 0.596]$) or mental strategy (retention: $F(1.80) = 0.18$, $p = 0.676$, $\eta_p^2 = 0.002$; $B = -0.096$, $p = 0.655$, BCa CI = $[-0.571, 0.336]$; comprehension: $F(1.80) = 0.02$, $p = 0.881$, $\eta_p^2 < 0.001$; $B = 0.024$, $p = 0.883$, BCa CI = $[-0.288, 0.344]$; transfer: $F(1.80) = 0.07$, $p = 0.791$, $\eta_p^2 = 0.001$; $B = -0.048$, $p = 0.785$, BCa CI = $[-0.373, 0.303]$). There were also no significant physical $\times$ mental strategy interactions for cognitive load after any of the tests (retention: $F(1.80) = 1.58$, $p = 0.212$, $\eta_p^2 = 0.019$; $B = -0.289$, $p = 0.240$, BCa CI = $[-0.752, 0.164]$; comprehension: $F(1.80) = 3.22$, $p = 0.076$, $\eta_p^2 = 0.039$; $B = -0.289$, $p = 0.090$, BCa CI = $[-0.609, 0.041]$; transfer: $F(1.80) = 2.54$, $p = 0.115$, $\eta_p^2 = 0.031$; $B = -0.289$, $p = 0.139$, BCa CI = $[-0.664, 0.062]$).

5.3. Efficiency

The mean learning outcome scores and cognitive load ratings together suggest that there might have been differences between conditions in the efficiency of the instructional strategies. An instructional strategy is more efficient when the same test performance is achieved with fewer cognitive resources [24]. Therefore, we calculated instructional efficiency based on the test performance (separately for retention, comprehension, and transfer) and the cognitive load reported after learning [25], resulting in three efficiency scores. Table 3 displays the means and standard deviations for the efficiency scores per condition.

Table 3. Means (standard deviations) per condition for instructional efficiency.

	Physical	Mental	Physical and Mental	Control
Efficiency (retention)	−0.35 (1.05)	−0.15 (0.99)	0.45 (0.94)	0.05 (0.87)
Efficiency (comprehension)	−0.32 (0.79)	−0.21 (0.93)	0.36 (1.35)	0.17 (0.71)
Efficiency (transfer)	−0.18 (0.71)	−0.25 (0.88)	0.46 (1.18)	−0.03 (0.82)

The results showed a significant physical $\times$ mental strategy interaction for the efficiency scores based on retention performance ($F(1.80) = 5.50$, $p = 0.022$, $\eta_p^2 = 0.064$; $B = 0.250$, $p = 0.021$, BCa CI = $[0.041, 0.449]$), comprehension performance ($F(1.80) = 6.06$, $p = 0.016$, $\eta_p^2 = 0.070$; $B = 0.266$, $p = 0.021$, BCa CI = $[0.075, 0.446]$), and transfer performance ($F(1.80) = 4.64$, $p = 0.035$, $\eta_p^2 = 0.055$; $B = 0.217$, $p = 0.035$, BCa CI = $[0.018, 0.417]$). Simple effects analyses indicated that there was higher efficiency if physical and mental strategies were combined than if the physical strategy was not combined with the mental strategy for efficiency based on retention performance ($p = 0.009$; BCa CI = $[0.203, 1.390]$), comprehension performance ($p = 0.028$; BCa CI = $[-0.032, 1.353]$), and transfer performance ($p = 0.027$; BCa CI = $[0.040, 1.263]$). There were no significant differences in efficiency scores between the conditions without the physical strategy, nor the mental strategy versus no strategy conditions; this was the case for efficiency based on retention performance ($p = 0.523$; BCa CI = $[-0.774, 0.399]$), comprehension performance ($p = 0.218$; BCa CI = $[-0.883, 0.140]$), and transfer performance ($p = 0.429$; BCa CI = $[-0.252, 0.681]$). There were no significant main effects of physical strategy or mental strategy on any of the three efficiency scores (all $ps > 0.10$).

6. Discussion

This study investigated the individual and combined effects of physical (annotation) and mental (imagined drag-and-drop) self-management strategies in learning from a split-

attention example. Overall, the results provide initial evidence that the combination of physically annotating the split-attention example and imagining dragging and dropping text segments to the corresponding element in the picture resulted in a lower cognitive load during learning compared to only using the annotation strategy. This was not accompanied by higher learning outcomes on retention, comprehension, or transfer tests, meaning that the same learning performance was obtained with less cognitive effort (Hypothesis 2a). The efficiency scores confirm this and further demonstrate that the combination of strategies appeared to be more instructionally efficient than using either strategy in isolation or not using any strategy. In support of Hypothesis 2b, this shows that the combined use of both physical and mental self-management strategies is a less cognitively demanding way to learn from split-attention examples. These findings add to prior self-management research by showing that combining self-management strategies can be beneficial for learners, particularly when physical and mental strategies are combined. This contrasts with the findings of Gordon et al. [15], who did not observe any learning benefits or changes in cognitive load when learners were required to use two physical strategies. Together, this suggests that a key element for the effectiveness of the combined use of two self-management strategies lies in combining strategies that not only require physical involvement, but include a component that explicitly encourages more active mental processing of the content. This interpretation needs to be tested in future research, for example, by directly comparing physical–physical strategies versus physical–mental strategies.

The results do not provide evidence for higher learning outcomes (Hypotheses 1a) or reduced cognitive load (Hypothesis 1b) when either using the annotation strategy or the mental drag-and-drop strategy over not using any of these strategies. This contrasts with earlier research reporting the learning benefits of using the annotation strategy (e.g., [5]) and the mental strategy [11,12] when learning from split-attention examples. Particularly the lack of an effect of the mental self-management strategy was unexpected, given that we used exactly the same materials and procedures as the studies by De Koning et al., which reported medium to large effects for the mental strategy compared to a no-strategy control condition and a physical self-management condition. One key difference, however, is that in their studies, the learning task was presented on a computer screen, while in the present study the learning task was presented on a sheet of paper lying on a desk in front of the learner. Both presentation types have different affordances, which might contribute to an explanation of these diverging findings. That is, paper-based learning materials more strongly invite physical interaction (e.g., placing the finger on a relevant element) than computer-based learning materials do. A non-touchscreen-based computer screen elicits much less physical interaction, because learners know that the screen cannot be controlled with direct physical interaction. Instead, learners presumably might have relied more on mental processing when confronted with the computer-based learning task that was used in the studies by De Koning et al. [11,12]. Using the paper-based materials in the present study might possibly have directed learners more to physical interactions than encouraged mental integration processes; therefore, the potential benefits of the mental strategy might not have been fully realized. This interpretation is in line with other research showing that there are differences in learning outcomes when using paper-based or computer-based materials in learning from split-attention examples (e.g., [3]), and more broadly, such as in reading comprehension (e.g., [26]). Future research could investigate to what extent the medium and associated affordances impact the effectiveness of physical and mental self-management strategies.

Regarding the physical annotation strategy, the differences between the current findings and those of prior research could be attributed to the available study duration. In the present study, there was a 4 min learning phase, which might have been too short of a time to use the annotation strategy to make the relevant connections between spatially separated text and picture elements and also use this as a basis to cognitively process this information into an accurate and coherent mental representation. In earlier studies using the physical annotation strategy, the learning phase was considerably longer. For

example, the Roodenrys study [5] had a 12 min learning phase. It is conceivable that, in earlier studies, benefits of using the physical annotation strategy could have materialized because there was sufficient time to engage in both the physical and mental activities, thus supporting learning. Future research is warranted to investigate whether learning time is actually a determining factor in the effectiveness of the physical annotation strategy and other physical or mental self-management strategies.

7. Limitations

The results and interpretations are subject to some limitations. First, as is consistent with other studies investigating the self-management effect, and to enable direct comparisons of our results to those in earlier studies, we used an overall measure of the cognitive load. One-item measures of the cognitive load can give reliable and valid indications of the experienced cognitive load, but are insufficient to detect changes in different types of cognitive load. Future studies could consider using other measures, such as multi-item self-report scales, which can give more specific insight into how different types of cognitive load are impacted by using self-management strategies (e.g., [27,28]). Second, given that our aim was to investigate the relative effectiveness of self-management strategies on learning and cognitive load, the study does not provide insight into how the two self-management strategies were used. Especially for the combination of strategies, this would be valuable to investigate in future studies, as it might provide more information regarding the way in which self-management strategies are used. It would, for example, be insightful to investigate whether and when learners alternate between strategies by analyzing video recordings of participants employing self-management strategies during learning. Third, the physical strategy and mental strategies investigated in this study did not produce a learning benefit over the control condition (which might have been due to the quite high performance of this control condition). It was, therefore, not possible to examine the relative effectiveness of each of these strategies when used in isolation. Except for the study by De Koning et al. [11,12], no other self-management studies have attempted to directly compare the effectiveness of two different self-management strategies, and this would thus be a useful future direction to explore further. In pursuing this direction of research, comparison of a variety of self-management strategies and/or use of different learning tasks could provide further insights into the effectiveness of self-management strategies.

8. Conclusions

The present study attempted to gain more insight into the whether it is helpful for learners studying split-attention examples to engage in physical and/or mental self-management strategies to support their learning. The results indicate that, within the context of this study, using a physical annotation strategy together with a mental drag-and-drop strategy is less cognitively demanding (compared to the physical strategy) and more instructionally efficient than using either a single self-management strategy or no self-management strategy. The present study also offers new research directions for future studies that could provide more insight into the effective use of physical and mental self-management studies.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and was conducted in the context of a research project approved by the Ethics Committee of Erasmus University Rotterdam (18-026, 16 October 2018).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data is available on request by the author.

Conflicts of Interest: The author declares no conflict of interest.

References

- Sweller, J.; van Merriënboer, J.J.G.; Paas, F. Cognitive Architecture and Instructional Design: 20 Years Later. *Educ. Psychol. Rev.* **2019**, *31*, 261–292. [CrossRef]
- Ginns, P. Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learn. Instr.* **2006**, *16*, 511–525. [CrossRef]
- Schroeder, N.L.; Cenkci, A.T. Spatial Contiguity and Spatial Split-Attention Effects in Multimedia Learning Environments: A Meta-Analysis. *Educ. Psychol. Rev.* **2018**, *30*, 679–701. [CrossRef]
- Castro-Alonso, J.C.; de Koning, B.B.; Fiorella, L.; Paas, F. Five Strategies for Optimizing Instructional Materials: Instructor- and Learner-Managed Cognitive Load. *Educ. Psychol. Rev.* **2021**, *33*, 1379–1407. [CrossRef]
- Roodenrys, K.; Agostinho, S.; Roodenrys, S.; Chandler, P. Managing One's Own Cognitive Load when Evidence of Split Attention is Present. *Appl. Cogn. Psychol.* **2012**, *26*, 878–886. [CrossRef]
- Zhang, S.; De Koning, B.B.; Agostinho, S.; Tindall-Ford, S.; Chandler, P.; Paas, F. The cognitive load self-management principle. In *The Cambridge Handbook of Multimedia Learning*, 3rd ed.; Mayer, R.E., Fiorella, L., Eds.; Cambridge University Press: New York, NY, USA, 2021; pp. 430–436.
- Sithole, S.T.M.; Chandler, P.; Abeysekera, I.; Paas, F. Benefits of guided self-management of attention on learning accounting. *J. Educ. Psychol.* **2017**, *109*, 220–232. [CrossRef]
- Tindall-Ford, S.; Agostinho, S.; Bokosmaty, S.; Paas, F.; Chandler, P. Computer-based learning of geometry from integrated and split-attention worked examples: The power of self-management. *Educ. Technol. Soc.* **2015**, *18*, 89–99.
- Zhang, S.; de Koning, B.B.; Paas, F. Finger pointing to self-manage cognitive load in learning from split-attention examples. *Appl. Cogn. Psychol.* **2022**, *36*, 767–779. [CrossRef]
- Agostinho, S.; Tindall-Ford, S.; Roodenrys, K. Adaptive diagrams: Handing control over to the learner to manage split-attention online. *Comput. Educ.* **2013**, *64*, 52–62. [CrossRef]
- de Koning, B.B.; Rop, G.; Paas, F. Learning from split-attention materials: Effects of teaching physical and mental learning strategies. *Contemp. Educ. Psychol.* **2020**, *61*, 101873. [CrossRef]
- de Koning, B.B.; Rop, G.; Paas, F. Effects of spatial distance on the effectiveness of mental and physical integration strategies in learning from split-attention examples. *Comput. Hum. Behav.* **2020**, *110*, 106379. [CrossRef]
- Glenberg, A.M.; Gutierrez, T.; Levin, J.R.; Japuntich, S.; Kaschak, M.P. Activity and Imagined Activity Can Enhance Young Children's Reading Comprehension. *J. Educ. Psychol.* **2004**, *96*, 424–436. [CrossRef]
- List, A.; Lin, C.-J. Content and quantity of highlights and annotations predict learning from multiple digital texts. *Comput. Educ.* **2023**, *199*, 104791. [CrossRef]
- Gordon, C.; Tindall-Ford, S.; Agostinho, S.; Paas, F. Learning from Instructor-managed and Self-managed Split-attention Materials. *Appl. Cogn. Psychol.* **2016**, *30*, 1–9. [CrossRef]
- Skulmowski, A.; Xu, K.M. Understanding Cognitive Load in Digital and Online Learning: A New Perspective on Extraneous Cognitive Load. *Educ. Psychol. Rev.* **2022**, *34*, 171–196. [CrossRef]
- Gaggioli, A.; Morganti, L.; Mondoni, M.; Antonietti, A. Benefits of Combined Mental and Physical Training in Learning a Complex Motor Skill in Basketball. *Psychology* **2013**, *4*, 1–6. [CrossRef]
- Schmidt, M.; Mavilidi, M.-F.; Singh, A.; Englert, C. Combining physical and cognitive training to improve kindergarten children's executive functions: A cluster randomized controlled trial. *Contemp. Educ. Psychol.* **2020**, *63*, 101908. [CrossRef]
- Wang, B.; Ginns, P.; Mockler, N. Sequencing Tracing with Imagination. *Educ. Psychol. Rev.* **2022**, *34*, 421–449. [CrossRef]
- Faul, F.; Erdfelder, E.; Lang, A.-G.; Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **2007**, *39*, 175–191. [CrossRef]
- Kalyuga, S.; Chandler, P.; Sweller, J. Levels of Expertise and Instructional Design. *Hum. Factors J. Hum. Factors Ergon. Soc.* **1998**, *40*, 1–17. [CrossRef]
- Paas, F.G.W.C. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *J. Educ. Psychol.* **1992**, *84*, 429–434. [CrossRef]
- Richardson, J.T.E. Eta squared and partial eta squared as measures of effect size in educational research. *Educ. Res. Rev.* **2011**, *6*, 135–147. [CrossRef]
- van Gog, T.; Paas, F. Instructional Efficiency: Revisiting the Original Construct in Educational Research. *Educ. Psychol.* **2008**, *43*, 16–26. [CrossRef]
- Paas, F.G.W.C.; Van Merriënboer, J.J.G. The Efficiency of Instructional Conditions: An Approach to Combine Mental Effort and Performance Measures. *Hum. Factors J. Hum. Factors Ergon. Soc.* **1993**, *35*, 737–743. [CrossRef]
- Delgado, P.; Salmerón, L. The inattentive on-screen reading: Reading medium affects attention and reading comprehension under time pressure. *Learn. Instr.* **2021**, *71*, 101396. [CrossRef] [PubMed]

27. Klepsch, M.; Seufert, T. Understanding instructional design effects by differentiated measurement of intrinsic, extraneous, and germane cognitive load. *Instr. Sci.* **2020**, *48*, 45–77. [CrossRef]
28. Leppink, J.; Paas, F.; Van Der Vleuten, C.P.M.; Van Gog, T.; Van Merriënboer, J.J.G. Development of an instrument for measuring different types of cognitive load. *Behav. Res. Methods* **2013**, *45*, 1058–1072. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

The More, the Better? Exploring the Effects of Modal and Codal Redundancy on Learning and Cognitive Load: An Experimental Study

Melanie Trypke ^{1,*}, Ferdinand Stebner ¹ and Joachim Wirth ²

¹ Institute of Educational Science, University of Osnabrück, Heger-Tor-Wall 9, 49074 Osnabrück, Lower Saxony, Germany; ferdinand.stebner@uni-osnabrueck.de

² Institute of Educational Research, Ruhr University Bochum, Universitätsstraße 150, 44801 Bochum, North Rhine Westphalia, Germany; lehrlernforschung@rub.de

* Correspondence: melanie.trypke@uni-osnabrueck.de

Abstract: This study explores how receiving identical information from different sources affects learning and cognitive load, focusing on two types of redundancy: modal redundancy, where redundant information comes from two visual sources (images and written text), and codal redundancy, where redundant information comes from two sources of different modalities which utilize the same symbol system (spoken and written text). Using a 2×2 between-subjects design involving modal (yes/no) and codal (yes/no) redundancy, 158 participants completed twenty learning tasks, consisting of ten construction and ten recall tasks. Additionally, they rated their cognitive load by indicating their perceived task difficulty and mental load. Overall, results indicate positive main effects of modal redundancy and negative effects of codal redundancy on learning and cognitive load. Furthermore, significant interaction effects suggest that modal redundancy may counterbalance the negative effects of codal redundancy, implying a compensatory mechanism in cognitive processing for construction tasks. These results highlight the importance of considering both modal and codal redundancy and their interaction in instructional design.

Citation: Trypke, M.; Stebner, F.; Wirth, J. The More, the Better? Exploring the Effects of Modal and Codal Redundancy on Learning and Cognitive Load: An Experimental Study. *Educ. Sci.* **2024**, *14*, 872. <https://doi.org/10.3390/educsci14080872>

Academic Editors: Fred Paas, Andre Tricot, Kim Ouwehand and Florence Lespiau

Received: 1 July 2024
Revised: 25 July 2024
Accepted: 26 July 2024
Published: 9 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: redundancy effect; redundancy principle; multimedia instruction; cognitive processing; cognitive load; mental load; perceived task difficulty

1. Introduction

Multimedia learning materials play an essential role in daily teaching practices, including the integration of text, images, and visualizations (e.g., images or animations) to build learning environments. This integration aligns with the principles of multimedia learning, which advocate for the incorporation of both verbal (written or spoken text) and visual components (graphics, animations, and images) within instructional materials [1]. Empirical studies have consistently underscored the advantages of multimedia learning environments, demonstrating that combining written text with visualizations enhances learning compared to text-only presentations, a phenomenon known as the multimedia principle [2]. Similarly, the modality effect [3] suggests that presenting text in spoken rather than written form, in addition to visualizations, can further enhance learning. However, while research on multimedia learning offers numerous advantages, it also introduces challenges related to cognitive load, particularly regarding the redundancy effect [4]. The redundancy effect refers to the negative impact when two or more sources provide redundant information simultaneously, potentially overloading the cognitive capacity of working memory [5]. However, empirical evidence demonstrates a diverse range of redundancy effects [6], from enhancements in learning [7] to drawbacks [8] or no effects [9]. This variability arises from the multiple facets of redundancy. One aspect of redundancy, as proposed by Cognitive Load Theory (CLT) [5], involves the contentual overlap within learning materials,

exemplified by instances where written text restates information depicted in diagrams [10]. Results indicate that presenting identical information in different formats can increase cognitive load and lead to poorer learning performance [11]. Moreover, the Cognitive Theory of Multimedia Learning (CTML) [1] highlights another facet of redundancy related to separate processing channels for verbal and visual information. According to CTML, an ineffective combination of verbal and visual information can overwhelm working memory capacity, thereby hindering learning due to the effort required to process and compare multiple streams of information. For instance, if individuals learn more effectively from visualizations paired with spoken text rather than from visualizations, spoken text, and redundant written text. In such cases, the visual channel can become overloaded as learners must switch between visualizations and written text and expend additional mental effort trying to reconcile spoken and written text. [2].

While prior research has primarily focused on redundancy concerning the overlap between spoken and written text, recent studies have introduced distinctions between different types of redundancy that account for the presentation of redundant information across various working memory channels [12]. This study aims to expand on these findings by exploring the concept of working memory channel redundancy, as referred to by Trypke et al. [6]. An experimental study was conducted to analyze the impact of two types of redundancy on learning and cognitive load.

2. Why Should We Distinguish Different Types of Redundancy?

Distinguishing between different types of redundancy is essential due to its multifaceted nature, which impacts learning and cognitive load in various ways. Researchers such as Albers et al. [12], Bohec and Jamet [13], Mayer [2], and Trypke et al. [6] have highlighted the importance of understanding these distinctions.

First, considering the nature of redundant information is crucial. Learners may encounter unnecessary or irrelevant information (unrelated to the learning content) within instructional materials, leading to a higher extraneous cognitive load and impeding learning. This perspective aligns with CLT, suggesting that irrelevant content increases cognitive load and detracts from the learning process [5]. This observation also corresponds with the findings from the CTML, particularly the coherence effect [14] and the seductive details effect [15], where irrelevant information diverts attention from the essential learning content.

Additionally, the presence of identical information (the focus of this study) can have varying impacts on learning. For example, adding written text that restates information depicted in diagrams can lead to negative effects, as observed by many researchers [16–18]. However, some studies indicate the positive effects of providing identical information through multiple sources, as it can guide learners' attention to relevant aspects of the material [19–21]. This principle is known as the signaling or cueing principle [22]. Nevertheless, these studies often do not account for the degree of contentual overlap, but recent empirical evidence [12] illustrates that even presenting identical information (100% overlap) can enhance learning outcomes and reduce cognitive load.

The second crucial aspect of redundancy is the presentation mode of information, which refers to the use of multiple modalities. Based on Albers et al. [12] and Trypke et al. [6], we distinguish between modal and codal redundancy.

2.1. Modal Redundancy

Modal redundancy involves scenarios where images and written text are presented and contain identical information. In this scenario, cognitive processing in working memory channels is influenced as images and written text “compete for limited cognitive resources in the visual channel” [23] (p. 124). This may lead to an overload of the visual channel as learners may struggle to differentiate between the images and accompanying written text or process them simultaneously.

2.2. Codal Redundancy

Codal redundancy occurs when learners receive narration (spoken text) and identical written text. In this scenario, the narration competes with the written text for cognitive processing resources, particularly impacting the verbal processing channel in working memory. This phenomenon aligns with Mayer's concept of verbal redundancy [2,24].

Distinguishing between the nature and presentation mode of redundant information is crucial for several reasons. First, this distinction allows for a more precise understanding of how different types of redundancy affect learning and cognitive load. Prior research has often treated redundancy as a uniform concept, failing to account for the nuanced ways in which different forms of redundant information can interact with working memory channels. By considering both content and the modality, instructional designers can create more effective learning materials that optimize cognitive processing and enhance learning efficiency. Moreover, understanding these distinctions can lead to the development of instructional strategies that mitigate the negative effects of redundancy while leveraging its potential benefits. For instance, educators can design multimedia presentations that strategically use redundancy to highlight critical information without overloading working memory, as demonstrated by the verbal redundancy effect [24].

2.3. Purpose of the Study

The specific research gap this study aims to fill involves the nuanced understanding of how different types of redundancy (modal and codal) affect learning and cognitive load when the provided content is identical. While previous research has examined the general effects of redundancy on learning, there has been limited exploration into how these effects differ based on the type of redundancy and the specific cognitive processes involved. This study seeks to clarify these distinctions by systematically investigating the interplay between modal and codal redundancy and their respective impacts on learning and cognitive load while the provided content is identical among all groups. Therefore, we formulated the following research question: How do modal and codal redundancy affect learning and cognitive load when the provided content is identical?

3. Hypotheses

This experimental study examines whether different types of redundancy affect learning and cognitive load differently. To investigate the research question, we formulated the following hypotheses:

Hypothesis 1. *Presenting identical information through images accompanied by narration (no modal or codal redundancy) results in superior learning, along with reduced cognitive load compared to narration and written text (codal redundancy) or images and written text (modal redundancy). This superiority is linked to the distribution of information across different modalities, alleviating cognitive processing systems in line with the modality effect [1,3].*

Hypothesis 2. *Presenting identical information through narration and written text (codal redundancy) leads to lower learning and higher cognitive load compared to the other experimental groups. This simultaneous engagement of both sensory channels requires working memory to coordinate information from both sources, leading to potential inferences between the narration and written text, as illustrated by Diao and Sweller [25].*

Hypothesis 3. *Presenting identical information through images, narration, and written text (modal and codal redundancy) will enhance learning and decrease cognitive load compared to groups with only one type of redundancy (group 2 codal redundancy and group 3 modal redundancy). In this scenario, we anticipate that redundancy can enhance learning, as the transient nature of narrations may be complemented by additional images and written text, affording learners' multiple retrieval cues and facilitating information processing [2,22].*

4. Method

4.1. Participants and Design

The online data collection was administered via the digital survey tool Unipark, with participants recruited through the subject pool Prolific (Prolific is an online platform that connects researchers with participants for survey and study recruitment. <https://www.prolific.com> (accessed on 11 March 2024)). To qualify for participation, participants had to be fluent in German and had access to either a tablet, laptop, or desktop computer with an audio device (headphones). A total of 158 participants (68.4% male; 31.73 years (Min. 18; Max. 66), $SD = 11.63$) ($n = 1$ (0.6%) divers; $n = 2$ (1.3%) not mentioned) participated in the experiment. The study utilizes a 2×2 factorial between-subjects design, with modal redundancy (yes vs. no) and codal redundancy (yes vs. no). As outlined in Table 1, participants were randomly assigned to one of four groups resulting from the factorial design. Utilizing G*Power 3.1.9.7 [26], we determined the necessary sample size, which indicated that a minimum of 128 participants was required, assuming a medium effect size ($f = 0.25$), a Type I error of 0.05, and a Type II error of 0.20.

Table 1. Assignment of participants to groups.

Redundancy		Presentation	N = 158
Modal	Codal		
No	No	Images and narration (IN)	$n = 43$ (27.2%)
No	Yes	Narration and written text (NT)	$n = 35$ (22.2%)
Yes	No	Images and written text (IT)	$n = 41$ (25.9%)
Yes	Yes	Images, narration, and written text (INT)	$n = 39$ (24.7%)

4.2. Learning Material (Rhombus Logic Tasks)

Since the expertise reversal effect [27] demonstrated that additional redundant information (e.g., additional written text to a diagram) initially benefits novice learners, it becomes detrimental as they become more familiar with the task; we aimed to exclude the effects of prior knowledge and used an adapted form of system-paced rhombus logic tasks from Albers et al. [12]. Participants were presented with rhombus shapes containing numbers and positions. Each task consisted of a row of three rhombuses with a position (top, down, left, right) and a number changing according to a logical pattern.

For the *construction tasks*, we tested learners' ability to understand and apply logical patterns and use inductive reasoning to construct the fourth rhombus. For the corresponding *recall tasks*, learners had to reconstruct one randomly chosen rhombus from the original three. Therefore, the recall tasks assessed memory retrieval, particularly the ability to remember and reconstruct information. All groups received the same tasks, but they were presented differently depending on the types of redundancy (Figure 1). For modal redundancy, identical information was presented via images and written text, while for codal redundancy, it was presented via narration and written text. The group with modal and codal redundancy received identical information via images, written text, and narration, whereas the group with neither modal nor codal redundancy via images and narration. Overall, participants had to complete the construction and corresponding recall tasks in a randomized order.

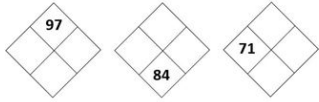
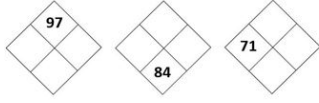
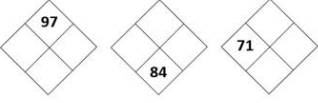
modal redundancy \ codal redundancy	no		yes	
no	1. IN		2. NT	
	I		I	
	N	Ninety-seven top Eighty-four down Seventy-one left	N	Ninety-seven top Eighty-four down Seventy-one left
yes	3. IT		4. INT	
	I		I	
	N	Ninety-seven top Eighty-four down Seventy-one left	N	Ninety-seven top Eighty-four down Seventy-one left
	T	Ninety-seven top Eighty-four down Seventy-one left	T	Ninety-seven top Eighty-four down Seventy-one left

Figure 1. Example task illustrating the different presentations for each of the four groups. I = Image; N = Narration; T = Written text.

4.3. Tests and Scoring

4.3.1. Learning Outcomes

Learning outcomes are operationally defined as the participants' performance on the rhombus logic tasks. The rhombus logic tasks represent a formal learning activity, as they are designed with a clear structure, requiring participants to follow specific instructions to complete them. The tasks are standardized, measuring participants' performance through accuracy, and aim to test specific cognitive abilities such as logical reasoning and memory retrieval. Participants could attain up to two points for each of the ten construction tasks and two points for each of the ten recall tasks, based on the correctness of the number and position (for each task, participants had to write the number in a designated field and select one of the four positions (top, down, left, right) from a drop-down menu). Consequently, each task had a potential maximum score of four points. Therefore, the total possible score for all ten construction tasks was 20 points, and similarly, the total possible score for all ten recall tasks was also 20 points. The overall learning score was calculated by adding the total scores from the construction and the recall tasks, leading to a maximum possible overall score of 40 points.

4.3.2. Perceived Task Difficulty and Mental Load

To measure cognitive load after each construction task and after each recall task, participants were asked to rate their perceived task difficulty and their experienced mental load. Both are related concepts but represent different aspects of cognitive processing. Perceived task difficulty reflects learners' subjective evaluation of how challenging they perceived the task to be. In contrast, the task-related experienced mental load indicates the amount of cognitive resources required to perform a task [28,29].

To measure participants' perceived task difficulty, they had to rate how easy or difficult it was to solve the task. The items were oriented on Bratfisch et al. [30] and Kalyuga et al. [17]. For the construction tasks, the item (in German) was "How easy or difficult was it for you to construct the fourth rhombus?". For the recall tasks, the item (in German) was "How easy or difficult was it for you to recall the [first, second, or third] rhombus?". Ratings were made using a nine-point Likert scale ranging from 1 = very, very easy to

9 = very, very difficult. To measure the mental load, we implemented a facet of the NASA Task load index (TLX) from Hart and Staveland [31]. To investigate the experienced mental load, we asked the questions (in German) “How mentally demanding was it to construct the fourth rhombus?” (for construction tasks) and “How mentally demanding was it to recall the [first, second, or third] rhombus?” (for recall tasks). Ratings were made using a nine-point Likert scale ranging from 1 = very, very little to 9 = very, very much.

The perceived task difficulty score and the mental load score were each computed using the mean score of the ratings for each participant. Finally, three scores were calculated: one construction, one recall, and one overall mean score of both construction and recall scores. All three scores could range from 1 to 9.

4.4. Procedure

After accessing the study via the subject pool, Prolific participants gave their informed consent to participate. They then completed an audio device check and answered questions on their demographics (gender, age, hearing impairment, mother tongue ($n = 140$ (88.6%) German as mother tongue; $n = 18$ (11.4%) other language). All participants indicated to be fluent in German, and analyses revealed no significant impact of mother tongue on the dependent variables), and the existence of left–right confusion). Afterward, they had to pass an attention check (type the color “green” into the text field below). At this point, they were randomly assigned to one of four groups according to the experimental design of the study (Table 1). Before participants entered the tasks, they watched an instructional video (2 min) explaining the procedure and providing an example task. After that, participants had to solve twenty tasks following a standardized procedure: After participants pressed the “Start” button, they viewed a video displaying each rhombus for four seconds. Afterward, they engaged in a distraction task, during which they were shown five animals or countries in different modalities (narration, written text, or images), followed by a question like “Which animal is smallest?”. The purpose of the distraction tasks was to occupy participants’ working memory. They had to type the answer into a text field. In the following, participants received instructions to construct the fourth rhombus (“What should the 4th rhombus in the logical series look like?”) and rated their perceived task difficulty and mental load on a nine-point Likert scale. For the recall tasks, participants were asked “What did the [first, second, or third] rhombus in the logical series look like?” and rated their perceived task difficulty and mental load. After that, they received the prompt that they had to click “next” to go to the next task. Once participants had completed the tasks, they had to pass an additional attention check. At the end of the survey, they were asked questions regarding cheating and technical issues and should indicate if they tried to ignore some information in the learning material. The survey duration was approximately 25–30 min, and participants received their monetary reward afterward (6 euros).

5. Results

5.1. Experimental Factors (Modal and Codal Redundancy)

The results were computed using SPSS 28. The internal consistency was satisfying, with Cronbach’s alpha >0.80 for construction and recall tasks. To ensure group comparability, we examined the distributions of gender and age. A chi-square test revealed no significant differences in gender distribution among the four groups, $\chi^2(9, N = 158) = 13.705$, $p = 0.133$. Similarly, ANOVA results indicated no significant differences in age distribution across the groups, $F(3, 154) = 0.494$, $p > 0.05$.

A 2×2 MANOVA with modal and codal redundancy as between-subject variables was performed on the construction and recall tasks and on perceived task difficulty and mental load for construction and recall tasks. For effect sizes, η^2 is reported where values of 0.01, 0.06, and 0.14 represent small, medium, and large effect sizes [32]. The results from Levene’s test for homogeneity of variance indicated that homogeneity of variance was met for all dependent variables except the recall tasks (to account for the violation of homogeneity of variances, we additionally conducted a Welch–ANOVA for the recall tasks).

Assumption of normality (measured with a Kolmogorov–Smirnov test) was applied for perceived task difficulty and mental load but was violated for the test scores (construction and recall). However, since MANOVAs are considered reasonably robust against violations of normality [33], we refrained from presenting non-parametrical methods in this case.

5.2. Learning

To investigate how modal and codal redundancy affect learning, we examined the learning scores for construction and recall tasks (Table 2 and Figure 2). For construction tasks, modal redundancy had a positive effect, $F(1,154) = 5.592$, $p < 0.05$, $\eta p^2 = 0.035$, enhancing learning outcomes, whereas codal redundancy had a negative effect, $F(1,154) = 5.432$, $p < 0.05$, $\eta p^2 = 0.034$. Additionally, a significant interaction between modal and codal redundancy was observed, $F(1,154) = 14.943$, $p < 0.001$, $\eta p^2 = 0.088$, indicating that modal redundancy mitigates the negative effects of codal redundancy. No significant effects were found for recall tasks (this non-significant finding was confirmed by a Welch–ANOVA, which was conducted to account for the violation of homogeneity of variances, $F(3, 83.696) = 1.211$, $p = 0.311$).

Table 2. Learning Scores for all groups ($N = 158$).

Groups	Construction Score		Recall Score		Overall Learning Score *	
	M	SD	M	SD	M	SD
IN ($n = 43$)	12.98	4.13	11.21	5.26	24.18	8.57
NT ($n = 35$)	8.74	5.05	8.80	6.09	17.54	10.66
IT ($n = 41$)	11.95	4.43	10.61	4.79	22.56	8.24
INT ($n = 39$)	13.00	3.45	10.79	4.51	23.79	6.88

* Minimum Score is 0; Maximum Score is 20 for Construction and Recall Score, 40 for Overall Learning score. Redundancy (modal/codal): IN (no/no), NT (no/yes), IT (yes/no), INT (yes/yes).

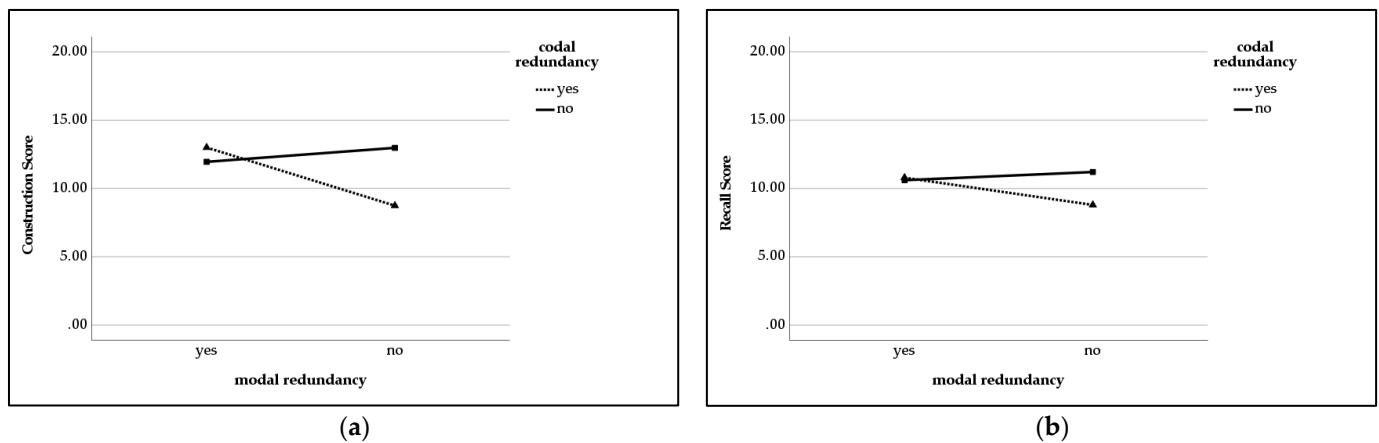


Figure 2. (a) Learning Scores for modal and codal redundancy (construction tasks); (b) Learning Scores for modal and codal redundancy (recall tasks).

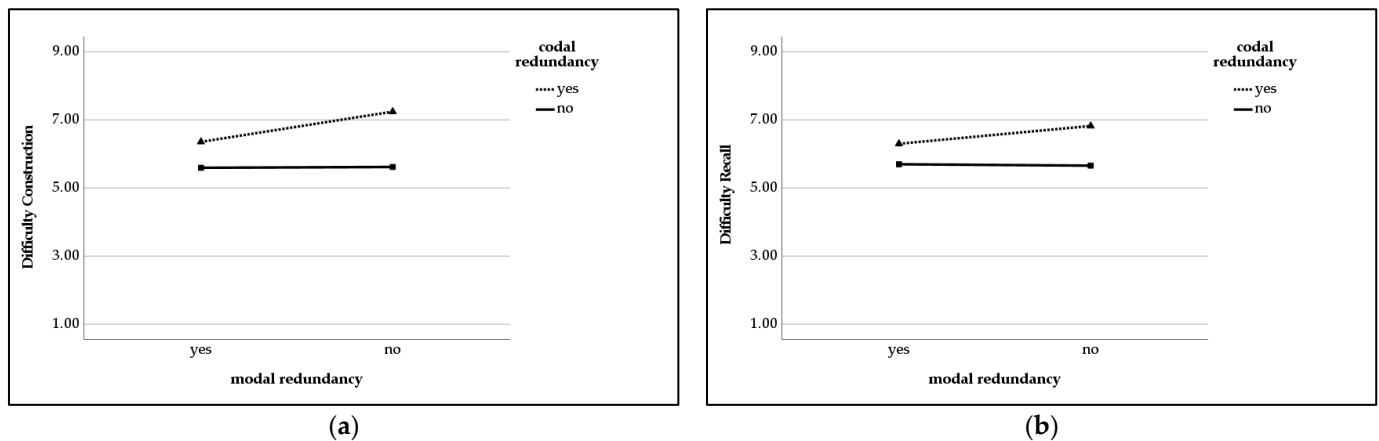
5.3. Perceived Task Difficulty

To investigate how modal and codal redundancy affect perceived task difficulty, we examined the ratings for construction and recall tasks (Table 3 and Figure 3). Results indicate that codal redundancy significantly increased perceived task difficulty for both construction, $F(1,154) = 24.349$, $p < 0.001$, $\eta p^2 = 0.137$, and recall tasks, $F(1,154) = 11.974$, $p < 0.001$, $\eta p^2 = 0.072$. In contrast, modal redundancy did not significantly affect perceived task difficulty.

Table 3. Perceived Task Difficulty Scores for all groups ($N = 158$).

Groups	Difficulty Construction		Difficulty Recall		Overall Difficulty *	
	M	SD	M	SD	M	SD
IN ($n = 43$)	5.62	1.68	5.66	1.81	5.64	1.65
NT ($n = 35$)	7.24	1.25	6.82	1.62	7.03	1.39
IT ($n = 41$)	5.59	1.68	5.70	1.61	5.65	1.54
INT ($n = 39$)	6.36	1.34	6.30	1.30	6.33	1.22

* Minimum score is 1; Maximum score is 9. Redundancy (modal/codal): IN (no/no), NT (no/yes), IT (yes/no), INT (yes/yes).

**Figure 3.** (a) Difficulty Scores for modal and codal redundancy (construction tasks); (b) Difficulty Scores for modal and codal redundancy (recall tasks).

5.4. Mental Load

To investigate how modal and codal redundancy affect mental load, we examined the ratings for construction and recall tasks (Table 4 and Figure 4). For the construction tasks, modal redundancy reduced mental load, $F(1,154) = 4.501$, $p < 0.05$, $\eta^2 = 0.028$, and codal redundancy increased it, $F(1,154) = 20.283$, $p < 0.001$, $\eta^2 = 0.116$. Additionally, the interaction between modal and codal redundancy was also significant, $F(1,154) = 5.049$, $p < 0.05$, $\eta^2 = 0.032$, suggesting that the presence of modal redundancy can alleviate the increased mental load caused by codal redundancy. However, no significant effects were found for the recall tasks.

Table 4. Mental Load Scores for all groups ($N = 158$).

Groups	Mental Load Construction		Mental Load Recall		Overall Mental Load *	
	M	SD	M	SD	M	SD
IN ($n = 43$)	5.58	1.71	5.57	1.84	5.58	1.69
NT ($n = 35$)	7.24	1.26	6.76	1.67	7.00	1.43
IT ($n = 41$)	5.61	1.67	5.66	1.59	5.64	1.54
INT ($n = 39$)	6.17	1.42	6.17	1.48	6.17	1.40

* Minimum Score is 1; Maximum Score is 9. Redundancy (modal/codal): IN (no/no), NT (no/yes), IT (yes/no), INT (yes/yes).

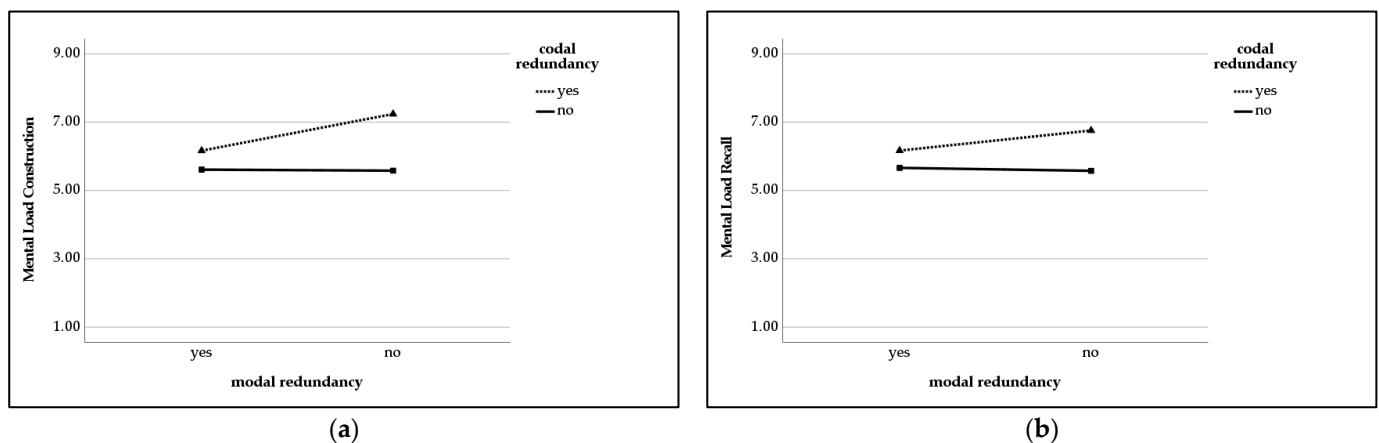


Figure 4. (a) Mental Load Scores for modal and codal redundancy (construction tasks); (b) Mental Load Scores for modal and codal redundancy (recall tasks).

6. Discussion

The results of this study contribute to our understanding of redundancy in multimedia learning, building upon existing research grounded in Cognitive Load Theory (CLT) and Cognitive Theory of Multimedia Learning (CTML). Our investigation explored how modal and codal redundancy impact learning outcomes and cognitive load.

The first hypothesis posited that presenting identical information through images accompanied by narration (no modal or codal redundancy) would result in superior learning outcomes and reduced cognitive load compared to the groups involving either codal or modal redundancy. Our results partially support this hypothesis. Specifically, the group with images and narration demonstrated higher scores in the construction tasks and reported lower cognitive load than the group with narration and written text (codal redundancy). However, contrary to expectations, no significant superiority was observed in learning and cognitive load between the group with images and narration and the group with images and written text (modal redundancy). This finding contrasts with prior research by Mayer and Moreno [34], who observed a modality effect where animation with narration was more effective than animation with written text when understanding the process of lightning formation and the car's braking system. One possible explanation for our findings is that the complexity of the learning tasks in our study may not have been sufficiently robust to highlight distinctions in learning outcomes and cognitive load. Unlike the specific scenarios investigated by Mayer and Moreno [34], our study focused on a setting where both narration and written text provided comparable support for cognitive processing. Therefore, it is conceivable that both groups in our study derived benefits from dual coding [35], which involves integrating visual and verbal information to enhance encoding and retrieval processes rather than overloading the verbal channel (codal redundancy). Given that both groups needed to integrate visual (images) and verbal (text or narration) information, it is plausible that the cognitive demands associated with integrating these types of information were sufficiently similar [36], resulting in comparable cognitive load and performance in construction tasks between the groups.

The second hypothesis posited that presenting identical information through narration and written text (codal redundancy) would result in lower learning outcomes and higher cognitive load compared to the other experimental groups. Our findings confirm this hypothesis. Codal redundancy was indeed associated with decreased learning outcomes and higher perceived task difficulty and mental load. This suggests that the concurrent engagement required to process information from both narration and written text increased cognitive load and impeded learning. This underscores that codal redundancy can overload the verbal processing channel, consistent with the verbal redundancy effect [2,37].

The results support Hypothesis 3, which posited that presenting identical information through images, narration, and written text (modal and codal redundancy) would enhance

learning outcomes and decrease cognitive load compared to groups with only one type of redundancy (group 2 codal redundancy and group 3 modal redundancy). The significant interaction observed between modal and codal redundancy on the construction tasks suggests that the combined use of both types of redundancy enhanced learning. Furthermore, this interaction appeared to mitigate the detrimental effects typically associated with codal redundancy alone, as indicated by the overall higher scores for construction tasks. This compensatory mechanism implies that while codal redundancy alone may increase cognitive load by requiring simultaneous processing of verbal and written information, the inclusion of modal redundancy serves to balance and distribute cognitive load more effectively across multiple channels. Consequently, learners benefit from improved learning and reduced cognitive load when information is presented through a combination of images, narration, and written text. These findings align with the reversed redundancy effect illustrated by many studies [23,38–40].

6.1. Answering the Research Question

How do modal and codal redundancy affect learning and cognitive load when the provided content is identical?

For learning, modal redundancy seems to have a positive effect on learning outcomes (construction tasks). Participants who received information in both visual and textual formats performed better on construction tasks compared to those who received the information through narration and written text (codal redundancy). The significant interaction between modal and codal redundancy shows that when both types of redundancy were present, the negative effects of codal redundancy were mitigated by the positive effects of modal redundancy, suggesting a compensatory mechanism. This mechanism can be explained by the findings from Mayer and Johnson [23], indicating that “redundant on-screen phrases may be particularly helpful for directing the learner’s attention with static illustrations” [23] (p. 385).

For cognitive load, modal redundancy did not significantly affect perceived task difficulty, indicating that participants did not find tasks more or less difficult when information was presented through images and written text. For mental load, modal redundancy had a significant effect on reducing mental load during construction tasks. We assume that the distribution of information across image and written text eased cognitive processing, leading to lower reported mental load. In contrast, codal redundancy significantly increased perceived task difficulty and mental load for construction tasks. Participants reported higher difficulty levels and placed a higher demand on the required cognitive resources to solve the tasks when they had to process both written and spoken information simultaneously. The significant interaction between modal and codal redundancy on mental load further demonstrates a compensatory mechanism similar to that observed for construction scores. When both types of redundancy were present, the negative impacts of codal redundancy were alleviated by the positive effects of modal redundancy.

But why do the effects only apply to the construction and not the recall tasks? When we refer to Chen et al. [41], we can assume that the construction tasks inherently involve a higher intrinsic cognitive load due to the need to understand the logical patterns to create the fourth rhombus (higher knowledge complexity). Since tasks with higher intrinsic cognitive load are more likely to be affected by additional extraneous cognitive load [4], changes in the manner of information presentation (which affect extraneous cognitive load) have a pronounced impact. In contrast, the recall tasks involve retrieving information from memory, typically resulting in a lower intrinsic cognitive load (lower knowledge complexity). Consequently, changes in how information is presented have a lesser effect on learning, perceived difficulty, and mental load for the recall tasks.

6.2. Limitations

While this study provides valuable insights into the effects of modal and codal redundancy, several limitations must be acknowledged. The study’s high internal validity is a

significant strength, owing to its controlled experimental design and meticulous management of variables such as redundancy type, presentation mode, and task difficulty. This control minimizes confounding factors, ensuring that observed differences in learning and cognitive load can be attributed to the manipulated variables rather than extraneous influences. Additionally, the use of system-paced tasks ensured standardized administration across all participants, reducing variability arising from individual pacing preferences. However, these strengths are counterbalanced by the study's limitations, particularly its low external validity. The tasks employed in this research were abstract and not representative of the complex, real-world learning scenarios encountered in educational settings. Authentic learning environments typically involve diverse materials that demand varied cognitive processes, which may not align with the simplified tasks used here. Consequently, the generalizability of findings to practical educational contexts remains uncertain. Another limitation concerns participant recruitment via platforms like Prolific, offering access to a broad participant pool but potentially introduces variability in engagement levels and task completion conditions. Despite efforts to mitigate these issues with attention checks, some participants may not have fully engaged or adhered to instructions consistently, which could affect data reliability. Furthermore, assessing participants' processing of learning materials was constrained by reliance on self-reporting, which may not always accurately reflect actual engagement or selective attention (we asked participants to report if they attempted to ignore some redundant information, but these self-reported data may not be entirely reliable). Future studies could benefit from integrating objective measures such as eye-tracking to provide more nuanced insights into how learners interact with and process the content. Such technologies can pinpoint where and for how long attention is directed, offering deeper insights into cognitive processing. Additionally, incorporating within-subjects designs and considering individual differences in cognitive abilities, such as working memory capacity, could enhance understanding of how learners benefit differently from various presentation formats. This approach would enable researchers to explore whether cognitive factors moderate the effects of redundancy on learning outcomes. Regarding the measurement of cognitive load, this study relied on perceived task difficulty and mental load. Since redundancy can increase extraneous cognitive load, employing more precise measures such as those proposed by Klepsch et al. [42] would be advantageous. Additionally, assessing active mental effort, as suggested by [28], could provide further insights into cognitive engagement. However, subjective measures are inherently prone to various biases, including social desirability and inaccuracies in self-perception [43], which can potentially undermine the validity and reliability of findings. To enhance the robustness of future studies, it is recommended to include complementary objective measures of cognitive load. Objective measures offer the advantage of providing quantifiable data that is less susceptible to individual biases. For instance, physiological indicators such as heart rate or pupil dilation can offer additional information by reflecting autonomic nervous system responses to cognitive demands [44]. However, expanding the number of dependent variables poses the risk of inducing a fatigue effect among participants, which should be carefully considered in future research designs.

7. Conclusions

The aim of this study was to investigate the effects of different types of redundancy—modal (images and written text) and codal (narration and written text)—on learning and cognitive load when the provided content is identical. An experimental study was conducted with participants completing tasks under varying redundancy conditions. The main findings indicate that modal redundancy enhances learning outcomes and reduces cognitive load, while codal redundancy increases cognitive load and hinders learning. When both types of redundancy are present, the positive effects of modal redundancy can mitigate the negative impacts of codal redundancy.

The key takeaway for educators and instructional designers is the importance of strategically combining visual and verbal information and the notion that repeated information,

when presented strategically, is helpful to learning. Therefore, it is essential to consider the types of redundancy and that a balanced combination of both types of redundancy can optimize learning outcomes and cognitive load by leveraging the strengths of each modality. However, researchers and educators must be aware of the degree of contentual overlap.

Author Contributions: Conceptualization, M.T. and J.W.; Investigation, M.T.; Methodology, M.T. and J.W.; Supervision, F.S. and J.W.; Writing—original draft, M.T.; Writing—review and editing, F.S. and J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was approved by the ethics committee of the Ruhr University Bochum, Germany (registration number: EPE-2023-012, March 2023).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data are available on request by contacting the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Mayer, R.E. Introduction to multimedia learning. In *The Cambridge Handbook of Multimedia Learning*, 2nd ed.; Mayer, R.E., Ed.; Cambridge Press: Cambridge, UK, 2014; pp. 1–24.
2. Mayer, R.E. *Multimedia Learning*, 2nd ed.; Cambridge Press: Cambridge, UK, 2009.
3. Low, R.; Sweller, J. The modality principle in multimedia learning. In *The Cambridge Handbook of Multimedia Learning*, 2nd ed.; Mayer, R.E., Ed.; Cambridge University Press: Cambridge, UK, 2014; pp. 227–246. [CrossRef]
4. Sweller, J.; Ayres, P.; Kalyuga, S. *Cognitive Load Theory*; Springer: Berlin/Heidelberg, Germany, 2011.
5. Chandler, P.; Sweller, J. Cognitive load theory and the format of instruction. *Cogn. Instr.* **1991**, *8*, 293–332. [CrossRef]
6. Trypke, M.; Stebner, F.; Wirth, J. Two types of redundancy in multimedia learning: A literature review. *Front. Psychol.* **2023**, *14*, 1–17. [CrossRef]
7. Adegoke, B.A. Integrating Animations, Narratives and Textual Information for Improving Physics Learning. *Electron. J. Res. Educ. Psychol.* **2017**, *8*, 725–748. [CrossRef]
8. Jamet, E.; Bohec, O.L. The effect of redundant text in multimedia instruction. *Contemp. Educ. Psychol.* **2007**, *32*, 588–598. [CrossRef]
9. Chu, S.L. Investigating the Effectiveness of Redundant Text and Animation in Multimedia Learning Environments. Ph.D. Thesis, University of Central Florida, Orlando, FL, USA, 2006.
10. Kalyuga, S.; Chandler, P.; Sweller, J. Incorporating learner experience into the design of multimedia instruction. *J. Educ. Psychol.* **2000**, *92*, 126–136. [CrossRef]
11. Sweller, J.; van Merriënboer, J.J.G.; Paas, F. Cognitive Architecture and Instructional Design: 20 Years Later. *Educ. Psychol. Rev.* **2019**, *31*, 261–292. [CrossRef]
12. Albers, F.; Trypke, M.; Stebner, F.; Wirth, J.; Plass, J.L. Different types of redundancy and their effect on learning and cognitive load. *Br. J. Educ. Psychol.* **2023**, 339–352. [CrossRef]
13. Bohec, O.L.; Jamet, E. Levels of verbal redundancy, note-taking and multimedia learning. In *Understanding Multimedia Documents*; Rouet, J.F., Lowe, R., Schnotz, W., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 79–101. [CrossRef]
14. Mayer, R.E.; Fiorella, L. 12 Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In *The Cambridge Handbook of Multimedia Learning*, 2nd ed.; Mayer, R.E., Ed.; Cambridge University Press: Cambridge, UK, 2014; pp. 279–315.
15. Harp, S.F.; Mayer, R.E. How seductive details do their damage: A theory of cognitive interest in science learning. *J. Educ. Psychol.* **1998**, *90*, 414–434. [CrossRef]
16. Bobis, J.; Sweller, J.; Cooper, M. Cognitive load effects in a primary school geometry task. *Learn. Instr.* **1993**, *3*, 1–21. [CrossRef]
17. Kalyuga, S.; Chandler, P.; Sweller, J. Levels of expertise and instructional design. *Hum. Factors* **1998**, *40*, 1–17. [CrossRef]
18. Sweller, J.; Chandler, P. Why Some Material Is Difficult to Learn. *Cogn. Instr.* **1994**, *12*, 185–233. [CrossRef]
19. Dowell, J.; Shmueli, Y. Blending speech output and visual text in the multimodal interface. *Hum. Factors* **2008**, *50*, 782–788. [CrossRef] [PubMed]
20. Gellevis, M.; van der Meij, H.; de Jong, T.; Pieters, J.M. Multimodal Versus Unimodal Instruction in a Complex Learning Context. *J. Exp. Educ.* **2002**, *70*, 215–239. [CrossRef]
21. McCrudden, M.T.; Hushman, C.J.; Marley, S.C. Exploring the boundary conditions of the redundancy principle. *J. Exp. Educ.* **2014**, *82*, 537–554. [CrossRef]
22. Van Gog, T. The signaling (or cueing) principle in multimedia learning. In *The Cambridge Handbook of Multimedia Learning*, 2nd ed.; Mayer, R.E., Ed.; Cambridge University Press: Cambridge, UK, 2014; pp. 263–278. [CrossRef]
23. Mayer, R.E.; Johnson, C.I. Revising the redundancy principle in multimedia learning. *J. Educ. Psychol.* **2008**, *100*, 380–386. [CrossRef]

24. Moreno, R.; Mayer, R.E. Verbal redundancy in multimedia learning: When reading helps listening. *J. Educ. Psychol.* **2002**, *94*, 156–163. [CrossRef]
25. Diao, Y.; Sweller, J. Redundancy in foreign language reading comprehension instruction: Concurrent written and spoken presentations. *Learn. Instr.* **2007**, *17*, 78–88. [CrossRef]
26. Faul, F.; Erdfelder, E.; Lang, A.-G.; Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **2007**, *39*, 175–191. [CrossRef] [PubMed]
27. Kalyuga, S.; Ayres, P.; Chandler, P.; Sweller, J. The expertise reversal effect. *Educ. Psychol.* **2003**, *38*, 23–31. [CrossRef]
28. Krell, M.; Hui, S.K.F. Evaluating an instrument to measure mental load and mental effort considering different sources of validity evidence. *Cogent Educ.* **2017**, *4*, 1280256. [CrossRef]
29. Paas, F.; Tuovinen, J.E.; Tabbers, H.; van Gerven, P.W.M. Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educ. Psychol.* **2003**, *38*, 63–71. [CrossRef]
30. Bratfisch, O.; Borg, G.; Dornic, S. Perceived item-difficulty in three tests of intellectual performance capacity. *Rep. Inst. Appl. Psychol.* **1972**, *29*, 1–14.
31. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload*; Hancock, P.A., Meshkati, N., Eds.; North-Holland: Amsterdam, The Netherlands, 1988; pp. 139–183. [CrossRef]
32. Richardson, J.T. Eta Squared and Partial Eta Squared as Measures of Effect Size in Educational Research. *Educ. Res. Rev.* **2011**, *6*, 135–147. [CrossRef]
33. Glass, G.V.; Peckham, P.D.; Sanders, J.R. Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance. *Rev. Educ. Res.* **1972**, *42*, 237–288. [CrossRef]
34. Mayer, R.E.; Moreno, R. A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *J. Educ. Psychol.* **1998**, *90*, 312–320. [CrossRef]
35. Paivio, A. *Mental Representations: A Dual Coding Approach*; Oxford University Press: Oxford, UK, 1986.
36. Schnotz, W. Integrated model of text and picture comprehension. In *The Cambridge Handbook of Multimedia Learning*; Mayer, R.E., Ed.; Cambridge University Press: Cambridge, UK, 2014; pp. 72–103.
37. Adesope, O.O.; Nesbit, J.C. Verbal redundancy in multimedia learning environments: A meta-analysis. *J. Educ. Psychol.* **2012**, *104*, 250–263. [CrossRef]
38. Ari, F.; Raymond, F.; Fethi, A.I.; Cheon, J.; Crooks, S.M.; Paniukov, D.; Kurucay, M. The effects of verbally redundant information on student learning: An instance of reverse redundancy. *Comput. Educ.* **2014**, *76*, 199–204. [CrossRef]
39. de Koning, B.B.; van Hooijdonk, C.M.J.; Lagerwerf, L. Verbal redundancy in a procedural animation: On-screen labels improve retention but not behavioral performance. *Comput. Educ.* **2017**, *107*, 45–53. [CrossRef]
40. Samur, Y. Redundancy effect on retention of vocabulary words using multimedia presentation. *Br. J. Educ. Technol.* **2012**, *43*, E166–E170. [CrossRef]
41. Chen, O.; Paas, F.; Sweller, J. A cognitive load theory approach to defining and measuring task complexity through element interactivity. *Educ. Psychol. Rev.* **2023**, *35*, 63. [CrossRef]
42. Klepsch, M.; Schmitz, F.; Seufert, T. Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Front. Psychol.* **2017**, *8*, 1997. [CrossRef]
43. Minkley, N.; Xu, K.M.; Krell, M. Analyzing relationships between causal and assessment factors of cognitive load: Associations between objective and subjective measures of cognitive load, stress, interest, and self-concept. *Front. Educ.* **2021**, *6*, 632907. [CrossRef]
44. Ayres, P.; Lee, J.Y.; Paas, F.; van Merriënboer, J.J.G. The validity of physiological measures to identify differences in intrinsic cognitive load. *Front. Psychol.* **2021**, *12*, 702538. [CrossRef] [PubMed]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

The Relationship between Interleaving and Variability Effects: A Cognitive Load Theory Perspective

Ouhao Chen ^{1,*}, Endah Retnowati ², Juan Cristobal Castro-Alonso ³, Fred Paas ⁴ and John Sweller ⁵¹ Department of Mathematics Education, Loughborough University, Loughborough LE11 3TU, UK² Department of Mathematics Education, Universitas Negeri Yogyakarta, Daerah Istimewa Yogyakarta 55281, Indonesia; e.retno@uny.ac.id³ School of Education, University of Birmingham, Birmingham B15 2TT, UK; j.castroalonso@bham.ac.uk⁴ Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, 3062 PA Rotterdam, The Netherlands; paas@essb.eur.nl⁵ School of Education, University of New South Wales, Sydney, NSW 2052, Australia; j.sweller@unsw.edu.au

* Correspondence: o.chen@lboro.ac.uk

Abstract: The interleaving effect indicates that students learn better from multiple areas that are interleaved rather than blocked. Two experiments tested the hypothesis that the effect is because interleaving facilitates comparisons between areas and is a variation of the variability effect that increases intrinsic cognitive load. Experiment 1 used an interleaved design with two obviously different topics and found no interleaving effect. Experiment 2 used a similar design but used topics that were more difficult to discriminate between, resulting in a clear advantage for the interleaved group associated with an increase in cognitive load. These results support the hypothesis that the interleaving and variability effects are closely related.

Keywords: cognitive load theory; interleaving effect; variability effect; discrimination hypothesis; working memory resources and intrinsic cognitive load

Citation: Chen, O.; Retnowati, E.; Castro-Alonso, J.C.; Paas, F.; Sweller, J. The Relationship between Interleaving and Variability Effects: A Cognitive Load Theory Perspective. *Educ. Sci.* **2023**, *13*, 1138. <https://doi.org/10.3390/educsci13111138>

Academic Editor: Dorothe Kienhues

Received: 16 October 2023

Revised: 7 November 2023

Accepted: 9 November 2023

Published: 14 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Students prefer to practice a single concept or procedure repeatedly before switching to a new concept or procedure (i.e., AAABBBCCC) compared to interleaving concepts or procedures (i.e., ABCABCABC) as they rate interleaving to be less enjoyable and more difficult, as well as being less common [1]. Counterintuitively, under some circumstances, interleaving two or more concepts during practice leads to better learning than practicing a single concept repeatedly, indicating the interleaving effect [2–4].

We designed two experiments to investigate the hypothesis that the interleaving effect is due to learning to categorise problems. Previously, using cognitive load theory [5], the variability effect was also hypothesised to be due to learning to categorise problems. If the two effects have similar causes, we further suggest that they are closely related and both may be combined under a cognitive load theory umbrella.

2. Cognitive Load Theory

Cognitive load theory [6,7] aims to generate novel instructions to facilitate learning. The following points summarise the theory.

1. Information can be categorised as either biologically primary if we have specifically evolved to acquire it or biologically secondary if we have not specifically evolved to acquire it [8–10].
2. Novel, secondary information can be acquired either by randomly generating it during problem solving and testing it for effectiveness or, more efficiently, by obtaining it from other people.

3. Irrespective of how the secondary information is obtained, it must be processed by a limited-capacity [11,12] and limited-duration [13] working memory.
4. If that information is to be used subsequently, it can be stored in long-term memory for subsequent use. Long-term memory has no known capacity or duration limits [14].
5. When environmental signals indicate that it is needed, information can be transferred from long-term memory back to working memory to generate appropriate action. Unlike when processing novel information, working memory has no known limits when processing familiar information retrieved from long-term memory [15].
6. Intrinsic cognitive load is imposed by the nature of learning materials, which is determined by the levels of element interactivity. Element interactivity is a concept determining the complexity of learning materials [16,17] in which an element is a concept or procedure.
7. Extraneous cognitive load is imposed by suboptimal instructional designs that artificially and unnecessarily increase element interactivity. Altering the design of instruction can reduce or eliminate extraneous cognitive load.
8. Germane cognitive load refers to the working-memory resources that are devoted to dealing with the element interactivity associated with intrinsic cognitive load [17,18].

Hence, intrinsic cognitive load and germane cognitive load represent beneficial cognitive loads that enhance learning. Conversely, extraneous cognitive load stems from suboptimal instructional designs that interfere with learning. In the realm of teaching, our objective is to optimise the presence of intrinsic and germane cognitive loads while minimising the presence of extraneous cognitive load [19–21].

3. Explanations of the Interleaving Effect

Any explanation of the interleaving effect must also explain why the effect cannot be demonstrated under some conditions. In other words, it must explain the limits of the effect. Cognitive load theory can be used as a construct to explain both the interleaving and variability effects based on their common characteristics.

3.1. Interleaving and the Discrimination Hypothesis

Discrimination Learning. When students learn multiple concepts that are similar to each other, a greater degree of similarity increases learning difficulty [22]. Failing to distinguish similar concepts is known as discrimination failure, and learning to discriminate similar concepts is called discrimination learning [2]. Discrimination learning has been found in a very wide range of domains, ranging from discriminating concepts outside the classroom, such as face recognition, to concepts in the classroom, such as learning language, mathematics, or science. Discrimination learning has been suggested to provide a theoretical foundation for explaining the interleaving effect, particularly in the domains of mathematics and physics, which is beneficial for learning in a very wide range of domains [2]. From a cognitive load theory perspective, an increase in similarity that needs to be distinguished will cause an increase in intrinsic cognitive load. Accordingly, as indicated below, the theory can potentially be used to predict when the interleaving effect is likely to occur.

Discrimination Hypothesis. An interleaved design was shown to be superior to a blocked design by several experimenters who used the discrimination hypothesis as an explanation for their results [23–27]. The authors in [25] investigated learning concepts or categories by observing exemplars, with participants randomly allocated to either the interleaved or blocked group. Six paintings were selected from each of 12 artists. During the intervention, the six paintings from each of the artists were either blocked or interleaved. Learners were informed of the correct response on each occasion and were tested on their ability to identify the artist of an unfamiliar painting from one of the 12 artists. The results favoured the interleaved design, with the effect explained by better learning to discriminate different artists' painting styles.

The study [24] replicated this finding and included a third group in which multiple paintings from multiple artists were shown simultaneously, with participants tested to identify the artist of a new painting that was from one of the artists. The results replicated the interleaving effect found before [25] and also found that the interleaved design performed equally as well as the simultaneously presented group, which further supported the discrimination hypothesis as an explanation of the interleaving effect.

The authors in [26] tested the interleaving effect on teaching mathematics in real classrooms. Four mathematics problems—the face, corner, edge, and angle of a prism—were taught in interleaved or blocked groups. Although the interleaved design impaired mathematics performance during the learning phase, performance on the post-test indicated that it significantly improved learning. The experiment suggested again that the interleaving effect depended on task similarity requiring learners to discriminate between different problems with similar solutions.

In the study, ref. [26] controlled the spacing factor for their interleaved design, suggesting that the spacing and the interleaving effects might be distinct. The interleaving effect might depend on learning to discriminate [25], whereas the spacing effect might depend on other factors such as the depletion of working memory after cognitive effort and recovery during rest [28,29]. Both explanations depend on different facets of cognitive load theory. Importantly, the explanation of the interleaving effect is identical to cognitive load theory's explanation of the variability effect. Both effects depend on people learning which concepts are similar and which are different.

3.2. Relations between the Interleaving Effect and the Variability Effect

The variability effect suggests that engaging in tasks with high surface variability enhances learning compared to tasks with low surface variability [30]. Tasks with higher variability impose a higher intrinsic cognitive load during learning due to more interactive elements being processed simultaneously in working memory [5]. With an increase in variability, the number of interactive elements is increased as learners need to distinguish more varied tasks requiring similar solutions. Learning to categorise problems is central to both the variability and interleaving effects.

The experimental procedures used to demonstrate the variability effect and the interleaving effect differ [5,26,31], but they have the same goal; teaching students to distinguish problem categories. The interleaved sequence in effect increases variability. Learners see a greater variety of problems or tasks in rapid succession with multiple changes between episodes compared to a single change from one block to another when the information is presented in a blocked sequence.

While the interleaving and variability effects have a common goal of teaching students to appropriately categorise problems, one difference between the two effects is that there is a common experimental design for all demonstrations of the variability effect but there are two designs for the interleaving effect. The variability effect always requires comparing highly variable examples of the same concept or procedure with less-variable examples. The interleaving effect can similarly compare interleaved examples of different surface structure versions of the same concept or procedure, but alternatively, one can use interleaved examples of different concepts or procedures that have similar surface structures [2,32]. In other words, interleaving can teach students to treat in the same way problems that look different but are in fact the same or to distinguish between different problems that look the same. The variability effect only teaches students to treat in the same way problems that look different but are the same. The current work tested the former version of the interleaving effect by using problems that appeared different but were structurally the same.

4. Present Study

There were two main goals for this study. Firstly, two experiments aimed to test that the interleaving effect is due to discrimination, with its increased intrinsic cognitive load leading to an increase in the depletion of working memory resources. Based on previous

results [28], measuring working memory resource depletion is positively correlated with measuring the intrinsic cognitive load, namely, the more working memory resources that are depleted, the higher is the level of intrinsic cognitive load imposed. Therefore, using working memory tests to measure working memory resource depletion can be used as a proxy for the level of intrinsic cognitive load, assuming all other factors such as levels of extraneous cognitive load are equal across groups.

Secondly, the study aimed to provide evidence for and extend the current literature by indicating that the interleaving effect is a variation of the variability effect that can be explained from a cognitive load theory perspective. If the interleaving effect is not obtained by using obviously different problems that really are different but is instead obtained using obviously different materials that actually require identical solutions, that result suggests that the increase in variability caused by interleaving provides the explanation of the effect.

Experiment 1 compared an interleaved design with a blocked design, using two dissimilar materials: mathematics and language materials. Learners do not need to learn to discriminate between mathematics and language materials because the differences are obvious. Accordingly, there should be no advantages to interleaving with no evidence for an interleaving effect. The following hypotheses were tested. There would be no significant differences between the interleaved design and the blocked design upon conducting a post-test (H1), and no significant differences between the two groups on working memory resource depletion (H2) as students do not have to learn to discriminate between mathematics and language materials because the differences are obvious.

Experiment 2 compared the interleaved design with the blocked design but used two different sets of mathematics problems solved by similar solutions. Unlike Experiment 1, students must learn that these problems that appear very different on the surface are in fact, similar. It was hypothesised that the interleaved group would perform better than the blocked group on the post-test (H3) when learning that two visually distinct problem categories require identical solutions, and that the interleaved group would deplete more working memory resources compared to the blocked group (H4) due to an increased intrinsic cognitive load due to having to discriminate between learning materials.

In summary, the two experiments were designed to test the general hypothesis that when dissimilar-looking information really is dissimilar because it requires unrelated problem solutions, it is less likely to lead to the interleaving effect (Section 5), while dissimilar-looking information that really is similar is more likely to lead to the interleaving effect (Section 6). This issue is important to counter the view that interleaving is always beneficial irrespective of the relations between interleaved tasks. To improve ecological validity, the two experiments were designed and conducted in real classrooms and based on a real national curriculum for school children.

Power Analysis and Ethics Approval

To obtain the interleaving effect, a sample of 120 participants was recommended considering an a priori effect size $f(V) = 0.27$ [large size effect, based on data from 26], α error probability = 0.05, power ($1 - \beta$ error probability) = 0.80.

All the experiments were conducted following the University ethics policy, with ethics approval numbers for Experiment 1 (IRB-2018-11-044) and 2 (160/UN.34.13/M/TU/2022).

5. Experiment 1

This experiment compared an interleaved with a blocked design using Mathematics (M) and Language (L) learning materials. Because students do not have to be taught to distinguish between the two topic areas, it was predicted that an interleaving effect would not be obtained.

5.1. Method

5.1.1. Participants

One hundred and forty-three Year-7 students (mean age: 13 years old; 75 females) were recruited. They were randomly assigned to four groups, namely, the Mathematics–Language blocked group (MMLL), the Language–Mathematics blocked group (LLMM), the Mathematics–Language interleaved group (MLML), and the Language–Mathematics interleaved group (LMLM). Each group had four learning episodes. All participants were novices concerning the mathematics and language topics taught. The experiment was conducted during regular class periods with the supervision of teachers.

5.1.2. Materials

Two mathematics booklets about measuring the angle made by parallel lines and transversal lines that cross them were designed for the two mathematics learning episodes for each of the four groups. Three worked example–problem solving pairs were designed for teaching the topic in each booklet (see Appendix A).

Two language booklets about coconuts were designed for the two language learning episodes for each of the four groups. Each booklet included a passage of three paragraphs followed by summaries of each of the three paragraphs that were designed to demonstrate how to write summaries of the ideas in each paragraph. The booklet also included information on how to answer a comprehension question based on the given passage (see Appendix B). This information was followed by another similar-length passage on the same topic with three paragraphs that the students had to summarise themselves before answering a comprehension question. In this way, a worked example was followed by a similar problem to solve.

Two versions of the working memory test were designed to measure working memory resources after every two learning episodes. The *language working memory test* included a storage task (memorising the final word of each sentence) and an information-processing task (assessing the logical coherence of each sentence) designed to measure the information storage and information processing functions of working memory [33]. The test began with a practice trial containing two sentences (Level 2), followed by Level 2, Level 3, and Level 4 trials, with each level containing 3 trials with 2, 3, or 4 sentences per trial, respectively. During each trial, the sentences were serially presented to participants, each appearing for 3 s. After every sentence, the participants selected either the ‘Made sense’ or ‘Did not make sense’ responses (processing function) before being shown the next sentence. At the end of each trial, the participants had to type the final word of each of the presented sentences (storage function).

The *mathematics working memory test* applied the same design principles but with equations instead of sentences, such as $3 + 2 - 1 = 6$. Participants needed to judge the validity of each equation, indicating ‘right’ or ‘wrong’, and memorise the first digit of the equation, for example, “3”. The internal consistency of the working memory test (including mathematics and language), estimated using Cronbach’s alpha, was 0.69.

Two delayed post-tests measuring the understanding of mathematics and language were designed. For the mathematics test, four questions that were similar to those taught were included. For the language test, two similar-length passages discussing coconuts were designed. Each passage had three paragraphs to be summarised and a comprehension question about this passage to answer. The internal consistency of the post-test, estimated using Cronbach’s alpha, was 0.85 for the language test and 0.61 for the mathematics test. For analyses, one of the questions was deleted from the mathematics test because it failed to discriminate between students.

5.1.3. Procedure

Before the experiment, the schoolteachers introduced the researcher to the students. Consent forms were distributed and collected. For mathematics learning, the relevant pre-requisite theorems were re-visited, whereas for language learning, the structure of a

passage and a paragraph for summarising it were explained. This preparation phase was 5 min.

The experiment was conducted over two days (see Table 1). During Day 1, participants were randomly assigned to one of the four experimental groups. For the first block of 60 min, they either studied two mathematics problems, two language activities, or one mathematics problem with one language activity. Teachers instructed students who finished the task early to review their answers, thus equating learning periods. This procedure was followed by the mathematics working memory test (15 min). For the second block of 60 min, the procedure was similar to the first block. Next, the language working memory test replaced the mathematics working memory test. On the second day, the mathematics and language post-tests were conducted for two 30 min periods.

Table 1. Experiment 1: procedure.

Day 1	MMLL	LLMM	MLML	LMLM
30 min	Mathematics A	Language A	Mathematics A	Language A
30 min	Mathematics B	Language B	Language A	Mathematics A
15 min	Mathematics Working Memory Test for all groups			
30 min	Language A	Mathematics A	Mathematics B	Language B
30 min	Language B	Mathematics B	Language B	Mathematics B
15 min	Language Working Memory Test for all groups			
Day 2	All groups			
30 min	Mathematics Post-Test			
30 min	Language Post-Test			

5.1.4. Scoring

For the language working memory test, one point was assigned for each correct judgement of grammatical coherence (i.e., the score for information processing), and another point was assigned for each correctly recalled last word (i.e., the score for information storage). The maximum score was 27 for both information processing and storage. The mathematics working memory test was scored in the same manner as the language working memory test. For the mathematics working memory test, one point was assigned for correctly memorising the first digit and one point was assigned for a correct judgment of equation equivalence.

For the mathematics post-test, each question could be solved via 4 steps. Each correct step was allocated 1 point. The total score for the mathematics test was 16 (4 questions).

For the language post-test, 1 point was allocated for correctly summarising a paragraph (maximum 3), and 1 point was allocated for the comprehension question, providing a maximum of 4 points for each passage. The total score for the language test was 8 (two passages).

5.2. Results

The means and standard deviations of the mathematics and language working memory test scores can be found in Table 2. The means and standard deviations of the mathematics and language post-test scores can be found in Table 3.

Table 2. Experiment 1: means (and standard deviations) of mathematics and language working memory scores.

Group	Processing	Storage
Mathematics Working Memory		
MMLL	22.75 (2.98)	19.67 (5.39)
LLMM	22.94 (2.61)	19.14 (6.12)
MLML	22.26 (4.02)	19.59 (5.98)
LMLM	22.70 (3.24)	18.76 (7.39)
Language Working Memory		
MMLL	20.72 (3.15)	21.89 (5.01)
LLMM	21.31 (1.55)	21.03 (3.72)
MLML	21.21 (3.05)	22.85 (5.21)
LMLM	21.54 (1.92)	22.51 (3.47)

Note. MMLL, $n = 36$; LLMM, $n = 36$; MLML, $n = 34$; LMLM, $n = 37$.

Table 3. Experiment 1: means (and standard deviations) of mathematics and language post-test scores.

Group	Mathematics Post-Test	Language Post-Test
MMLL	8.14 (3.78)	6.72 (2.25)
LLMM	8.00 (4.38)	7.28 (1.75)
MLML	7.64 (4.60)	6.97 (2.28)
LMLM	8.00 (4.32)	6.41 (2.43)

Note. MMLL, $n = 36$; LLMM, $n = 36$; MLML, $n = 34$; LMLM, $n = 37$.

5.2.1. Mathematics Working Memory Test

Two ANOVAs were conducted to analyse the processing and storage scores separately. For both the processing scores, $F(3, 139) = 0.27$, $MSE = 10.50$, $p = 0.85$, $\eta_p^2 = 0.006$, and the storage scores, $F(3, 139) = 0.17$, $MSE = 39.42$, $p = 0.92$, $\eta_p^2 = 0.004$, the effect of group was not significant. Following these non-significant ANOVAs, Bayes analyses were calculated to provide grounds for accepting the null hypothesis. For the storage scores, $BF_{01} = 22.2$, suggesting the data were about 22 times more likely under the null hypothesis compared to the alternative hypothesis. Similarly, for the processing scores, $BF_{01} = 19.6$, suggesting the data were 20 times more likely under the null hypothesis compared to the alternative hypothesis. These results indicate that there was no evidence of working memory resource depletion differences between the groups.

5.2.2. Language Working Memory Test

Identical analyses were carried out on the language working memory test data. For the processing scores, the effect of group was not significant, $F(3, 139) = 0.69$, $MSE = 6.27$, $p = 0.56$, $\eta_p^2 = 0.015$, and $BF_{01} = 12.2$, suggesting the data were about 12 times more likely under the null hypothesis compared to the alternative hypothesis. For the storage scores, the effect of group was again not significant, $F(3, 139) = 1.18$, $MSE = 19.37$, $p = 0.32$, $\eta_p^2 = 0.025$, and $BF_{01} = 6.90$, suggesting the data were about seven times more likely under the null hypothesis compared to the alternative hypothesis. Again, there were no significant working memory resource depletion differences found between the groups.

5.2.3. Post-Tests

A mixed 4 (Groups: MMLL, LLMM, MLML, LMLM) $\times$ 2 (Mathematics vs. Language) ANOVA with repeated measures on the last factor was used to analyse the data for the post-tests. For the mathematics post-test, the effect of group was not significant, $F(3, 139) = 0.31$, $MSE = 18.03.02$, $p = 0.82$, $\eta_p^2 = 0.010$, and $BF_{01} = 10$, suggesting the data were about 10 times more likely under the null hypothesis than the alternative hypothesis. For the language post-test, the effect of group also was not significant, $F(3, 139) = 1.04$, $MSE = 4.81$, $p = 0.38$,

$\eta_p^2 = 0.022$, and $BF_{01} = 10$, suggesting the data were about 10 times more likely under the null hypothesis than the alternative hypothesis.

5.3. Discussion

In Experiment 1, there was no evidence of either an interleaving effect (*H1*) or of working memory resource depletion due to an increased intrinsic cognitive load for the interleaved group (*H2*), which supported our hypotheses for Experiment 1. The failure to find an interleaving effect or working memory resource depletion is likely to have been caused by the absence of a need to learn to discriminate between the materials. Because they were so dissimilar, the difference between the language and mathematics information was obvious. We hypothesised that an interleaving effect would not be found when interleaving two dissimilar topics that students can easily discriminate between.

Interestingly, these results do not support the suggestion that the interleaving effect is due to spacing. For half of the groups, the topics were interleaved and therefore spaced (i.e., the MLML and LMLM groups). If spacing had been relevant in this context, these groups should have obtained higher learning outcomes on the post-test and less working memory resources depleted on the working memory test than the two blocked groups (i.e., the MMLL and LLMM groups), based on the framework of Chen and colleagues [29,34]. However, neither of these tests revealed evidence indicating the superiority of the interleaved design. The failure to find an effect can be attributed to students not needing to learn to discriminate between the mathematics- and language-based content.

6. Experiment 2

This experiment further investigated the discrimination hypothesis of the interleaving effect. There are two ways in which this hypothesis can be tested. Students may need to learn that seemingly similar-looking materials are in fact distinct and must be treated as being distinct, or they may need to learn that specific types of materials that look different are in fact the same and should be treated as being the same. In Experiment 2, different = appearing mathematics materials that in fact required the same solution were used. It was hypothesised that interleaving such materials would assist students to learn to recognise them and learn to treat them as being functionally identical.

6.1. Method

6.1.1. Participants

A 2 (Design: Interleaved vs. Blocked) $\times$ 2 (Testing Time: Immediate vs. Delayed) between-subject design was used. One hundred and fifteen Year-7 students were recruited for this experiment. They were randomly assigned to one of the four groups: interleaved with immediate testing, interleaved with delayed testing, blocked with immediate testing, and blocked with delayed testing. The mathematics topic of creating auxiliary lines to solve an unknown angle was chosen. No participants had been taught this topic and so it was new to them. The experiment was conducted during regular class periods.

6.1.2. Materials

Sixteen slides were created to introduce basic concepts associated with the learning objective of how to create auxiliary lines in a geometrical figure in such a way that angle theorems can be used to measure the unknown angles. Four booklets were designed with two types of problems. For type A problems, the unknown angle was formed between two parallel lines (see Appendix C), whereas for type B problems, the unknown angle was formed by two squares within two parallel lines (see Appendix D).

The two types of problems could both be solved by creating auxiliary lines. For each of type A and B problems, three worked example–problem solving pairs were designed, totalling six worked example–problem solving pairs. They were presented in either interleaved or blocked form to teach the two types of problems for each group. For the interleaved group, one worked example–problem solving pair teaching the type A category

of problems was followed by a worked example–problem solving pair teaching the type B category of problems, and so the six pairs were arranged using an ABABAB format. For the blocked group, all the type A problems were taught before any type B problems were seen, followed by all the type B problems, resulting in the six pairs following an AAABBB format.

The same mathematics working memory test in Experiment 1 was used in Experiment 2. The internal reliability of the working memory test was 0.86. For the post-test, we designed six questions that were similar to those taught during the instruction period. The internal reliability of the post-test was 0.84.

6.1.3. Procedure

The procedure is summarised in Table 4. Initially, all students were presented and taught the introductory materials for 20 min. Next, during the first learning phase, students in each group studied three worked example pairs for 3 min/pair. For groups with delayed testing, students then had a 20 min break after 9 min of studying the three worked example pairs. During the second learning phase, another 9 min were used to study the remaining three worked example–problem solving pairs. For groups with delayed testing, they had another 20 min break. Teachers instructed students who finished tasks earlier to review their answers, thus ensuring that the learning and resting periods were controlled and equal irrespective of group. The working memory test (15 min) was conducted after these two learning phases. Finally, the post-test lasted 20 min.

Table 4. Experiment 2: procedure.

	Interleaved Design Immediate Testing	Delayed Testing	Blocked Design Immediate Testing	Delayed Testing
20 min	Introduction of basic concepts			
3 min	Mathematics A	Mathematics A	Mathematics A	Mathematics A
3 min	Mathematics B	Mathematics B	Mathematics A	Mathematics A
3 min	Mathematics A	Mathematics A	Mathematics A	Mathematics A
20 min		Rest		Rest
3 min	Mathematics B	Mathematics B	Mathematics B	Mathematics B
3 min	Mathematics A	Mathematics A	Mathematics B	Mathematics B
3 min	Mathematics B	Mathematics B	Mathematics B	Mathematics B
20 min		Rest		Rest
15 min	Mathematics Working Memory Test for all groups			
20 min	Mathematics Post-Test for all groups			

6.1.4. Scoring

The scoring system used in this experiment was the same as for Experiment 1. The maximum scores for the processing and storage parts of the working memory test were both 27. The maximum score for the post-test was 6.

6.2. Results

The means and standard deviations of the working memory test scores can be found in the Table 5. The means and standard deviations of the post-test scores are displayed in Table 6.

Table 5. Experiment 2: means (and standard deviations) of working memory scores.

Group	Processing	Storage
1. Interleaved with immediate testing	42.39 (6.86)	44.32 (6.92)
2. Interleaved with delayed testing	42.77 (6.36)	45.17 (4.11)
3. Blocked with immediate testing	41.39 (6.84)	51.07 (2.83)
4. Blocked with delayed testing	39.17 (9.50)	44.00 (7.86)

Note. Group 1, $n = 28$; Group 2, $n = 30$; Group 3, $n = 28$; Group 4, $n = 29$.

Table 6. Experiment 2: means (and standard deviations) of post-test scores.

Groups	Post-Test
1. Interleaved with immediate testing	3.07 (1.82)
2. Interleaved with delayed testing	2.50 (1.87)
3. Blocked with immediate testing	2.45 (1.85)
4. Blocked with delayed testing	1.67 (1.59)

Note. Group 1, $n = 28$; Group 2, $n = 30$; Group 3, $n = 28$; Group 4, $n = 29$.

6.2.1. Working Memory Test: Processing

A 2 (Design: Interleaved vs. Massed) $\times$ 2 (Testing Time: Immediate vs. Delayed) ANOVA was conducted on the processing scores. The effect of design was not significant, $F(1, 111) = 2.70$, $MSE = 56.15$, $p = 0.10$, partial $\eta^2 = 0.024$, and $BF_{01} = 1.43$, suggesting the data were slightly above one time more likely under the null hypothesis compared to the alternative hypothesis. The effect of testing time was not significant, $F(1, 111) = 0.44$, $MSE = 56.15$, $p = 0.51$, partial $\eta^2 = 0.004$, and $BF_{01} = 5$, suggesting the data were about five times more likely under the null hypothesis compared to the alternative hypothesis. The Design $\times$ Testing Time interaction was also not significant, $F(1, 111) = 0.86$, $MSE = 56.15$, $p = 0.36$, partial $\eta^2 = 0.008$, and $BF_{01} = 20$, suggesting the data were about 20 times more likely under the null hypothesis compared to the alternative hypothesis.

6.2.2. Working Memory Test: Storage

A 2 (Design: Interleaved vs. Massed) $\times$ 2 (Testing Time: Immediate vs. Delayed) ANOVA on the storage scores was conducted. The effect of design was significant, $F(1, 111) = 6.66$, $MSE = 33.61$, $p = 0.011$, partial $\eta^2 = 0.057$, and $BF_{01} = 5.2$. The effect of testing time was significant, $F(1, 111) = 8.28$, $MSE = 33.61$, $p = 0.005$, partial $\eta^2 = 0.069$, and $BF_{01} = 1.4$. The interaction Design $\times$ Testing Time was also significant, $F(1, 111) = 13.39$, $MSE = 33.61$, $p < 0.001$, partial $\eta^2 = 0.108$, and $BF_{01} = 4.8$. The Bayes factors were more than 1, which was due to the large sample size and relatively small effect size.

The significant interaction was followed by simple effects tests. For the immediate testing groups, interleaving depleted working memory resources to a greater extent than blocking, $t(54) = -4.77$, $SED_{diff} = 1.41$, $p < 0.001$, $d = 1.08$, and $BF_{01} = 0.01$ (suggesting the data were about 100 times more likely under the alternative hypothesis). For the delayed testing groups, there was no significant difference between the interleaved and blocked designs, $t(57) = 0.72$, $SED_{diff} = 1.63$, $p = 0.48$, $d = 0.56$, and $BF_{01} = 142.9$ (suggesting the data were about 143 times more likely under the null hypothesis than the alternative hypothesis).

6.2.3. Post-Test

A 2 (Design: Interleaved vs. Blocked) $\times$ 2 (Testing Time: Immediate vs. Delayed) ANOVA was used to analyse the post-test scores. The effect of design was significant, $F(1, 111) = 4.77$, $MSE = 3.19$, $p = 0.03$, partial $\eta^2 = 0.041$, and $BF_{01} = 1.3$, indicating that the interleaved design outperformed the blocked design, showing an interleaving effect. The effect of testing time was significant, $F(1, 111) = 4.09$, $MSE = 13.07$, $p = 0.046$, partial $\eta^2 = 0.036$, and $BF_{01} = 1.7$, indicating that groups with immediate testing outperformed those with delayed testing. Again, the Bayes factor values above 1 are due to the large sample size and relatively small effect size. The interaction between design and testing time was not significant, $F(1, 111) = 0.09$, $MSE = 0.31$, $p = 0.76$, $\eta_p^2 = 0.001$, and $BF_{01} = 3.3$ (suggesting the data were about three times more likely under the null hypothesis than the alternative hypothesis).

6.3. Discussion

Using different learning materials requiring a similar solution generated a conventional interleaving effect ($H3$). In addition, the storage part of the working memory test indicated that the interleaved groups depleted more working memory resources than the blocked groups ($H4$), which was shown with immediate testing but not with delayed

testing. According to the working memory resources depletion hypothesis [28], that difference in working memory scores disappeared for the delayed memory scores since the rest period should have permitted the recovery of working memory resources, eliminating any differences. The increased depletion of working memory resources by the interleaved group when measured immediately without rest suggests an increased cognitive load in this group compared to the equivalent blocked group. An increase in cognitive load should generate an increase in working memory resource depletion [28]. Therefore, the results supported the hypotheses of Experiment 2.

As working memory resources depletion is positively correlated with the levels of intrinsic cognitive load [28], therefore, a higher depletion of working memory resources indicates a higher level of intrinsic cognitive load. We suggest that interleaving increases intrinsic cognitive load (as evidenced by the increased depletion of working memory resources) for the same reason that variability increases cognitive load. Learners needed to learn to recognise the differences and similarities between the two types of problems and learn which differences did not matter when solving the problems and which similarities did matter because both problem types required the same solution. Attempting to learn to detect the critical similarities and irrelevant differences as well as learning how to generate a similar solution for problems that do not look similar are activities that are likely to be engaged in when faced with interleaving but less likely to be engaged in when faced with blocking, resulting in an increase in intrinsic cognitive load when interleaving. If the same factors result in the variability effect, interleaving may result in the same increase in intrinsic cognitive load for the same reason as increasing variability. If so, the interleaving effect is related to the variability effect, and both occur for the same cognitive load theory reasons.

7. General Discussion

The primary aim of this research was to investigate the discrimination hypothesis used to explain the interleaving effect. In addition, using a cognitive load theory perspective, we wished to link the interleaving and variability effects.

Experiment 1 interleaved two tasks that were clearly distinct, with one task involving mathematics and the other task involving language content. Students did not have to learn the important characteristics that distinguished the two tasks because they were very familiar with the relevant distinguishing characteristics. As hypothesised, the interleaving effect was not found. Because of the tendency to assume that all interleaving is beneficial, we believe the failure to find an effect in Experiment 1 is important, especially in conjunction with the results of Experiment 2.

By using two similar mathematics tasks in Experiment 2, the interleaving effect was obtained: we observed higher post-test scores (learning) and more depletion of working memory resources associated with interleaving than with blocking. Learners in Experiment 2 had to learn to identify the characteristics of the mathematics problems that were relevant to the solution and learn to detect those critical characteristics despite the different surface structures of the two problem types. Testing problems appearing different but in fact the same in Experiment 2 also echoed the mechanism of the variability effect that could aid in generalisation within a concept due to more elaborate and differentiated encoding during learning.

Interleaving allowed students to compare two different structures and learn to identify the specific structures that must be present to allow for the solution while also learning which structures to ignore. For the blocked groups, comparing and contrasting the structures was less important, thus reducing the intrinsic cognitive load and learning. The study by [28] introduced working memory depletion and recovery as concepts relevant to cognitive load theory and suggested that at least under some circumstances, the spacing effect could be treated as a cognitive load theory effect. In contrast, in [29] it was suggested that the interleaving effect is due to learning to discriminate and is not a cognitive load theory effect. That conclusion may be erroneous, with the two effects caused by different cognitive load theory concepts. The variability effect is a cognitive load theory effect [5,30], and a

comparison between the variability and interleaving effects reveals very close similarities. We suggest that learning to discriminate increases intrinsic cognitive load, as is the case for the variability effect [5]. Provided that the increase does not exceed working memory limits, more may be learned by interleaving than by blocking.

Limitations and Future Directions

There are two main limitations of this study: Firstly, although working memory resources are positively correlated with levels of intrinsic cognitive load [28], future research might use cognitive load scales, such as recent multidimensional instruments [20,35,36], to directly measure the level of intrinsic cognitive load. Secondly, for Experiment 2, we only used visually distinct problems that required identical solutions but did not use visually identical problems that required distinct solutions, providing a possible follow-up for the current study.

8. Conclusions

Based on the current experiments, the results support the suggestion that the interleaving effect is due to discrimination and relies on interleaving leading to a higher intrinsic cognitive load, thus providing learners with more to learn. The same cognitive mechanism suggests that the interleaving effect is related to the variability effect. The failure to find a spacing effect embedded in the interleaved design in Experiment 1 provides some evidence supporting the framework in [29] that the spacing effect may be due to working memory resource depletion, while the interleaving effect is due to the discrimination of learning items.

Author Contributions: Conceptualization, O.C.; methodology, O.C., J.S. and F.P.; formal analysis, O.C. and E.R.; investigation, O.C. and E.R.; data curation, E.R.; writing—original draft preparation, O.C.; writing—review and editing, J.C.C.-A., J.S. and F.P.; funding acquisition, J.C.C.-A. All authors have read and agreed to the published version of the manuscript.

Funding: Funding from ANID/PIA/Basal Funds for Centers of Excellence FB0003 is gratefully acknowledged.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Yogyakarta State University (IRB-2018-11-044 and 160/UN.34.13/M/TU/2022 and August 2020).

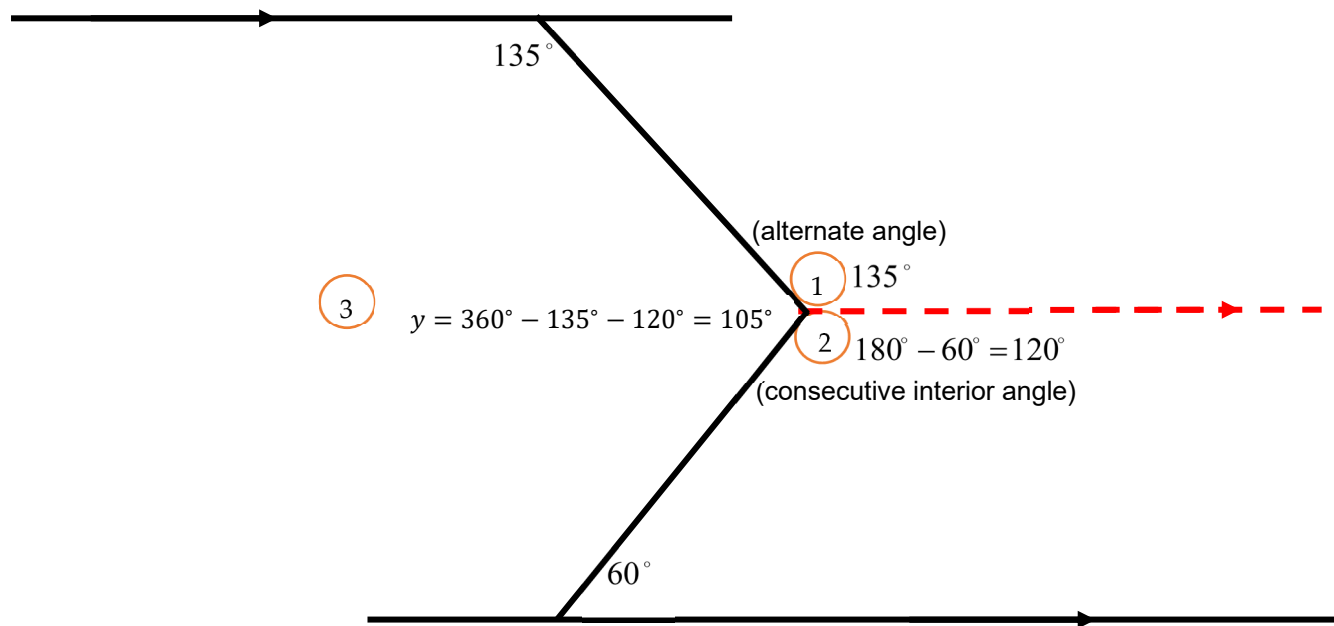
Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data could be shared by contacting with the first and the second authors.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Sample of Mathematics Material Used in Experiment 1

Please calculate the value of angle y



Appendix B. Sample of Language Material Used in Experiment 1 (Translated from the Original Bahasa Indonesia)

Reading 1. (15 min)

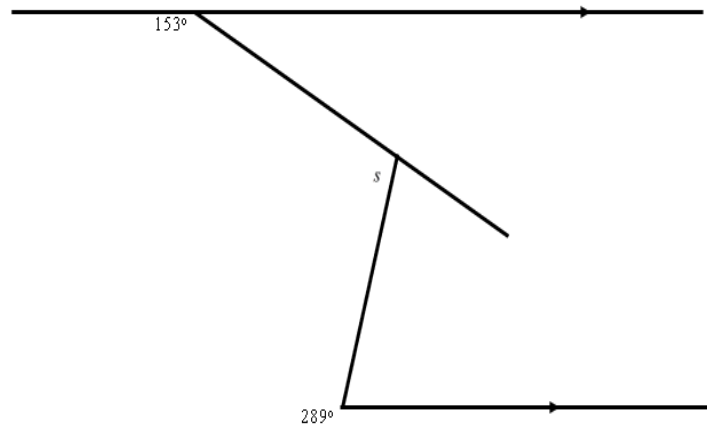
Read the following three paragraphs about the benefits of consuming coconut water and learn how to write a summary for each paragraph and how to answer a comprehension question about this reading.

Paragraph	Summary
One of the benefits of coconut is to prevent premature aging and relieve diarrhoea. The relevant part is coconut water. The water in coconut fruit contains cytokinin hormones that prevent premature aging, while the minerals, amino acids, and enzymes it contains are useful for relieving diarrhoea.	This paragraph describes the benefits of coconut water to prevent premature aging and relieve diarrhoea.
The content of cytokinin in coconut water serves to regulate the growth, development, and aging of cells. The cytokinin content in coconut water has anti-aging, anti-carcinogenic, and anti-thrombotic effects.	This paragraph describes the content of cytokinin as anti-aging, anti-carcinogenic, and anti-thrombotic effects.
Coconut water is useful when you have diarrhoea because it can replace the fluid lost from the gastrointestinal tract. This is because coconut water contains amino acids, enzymes, minerals, and fatty acids that result in coconut water having high osmolarity. In addition, coconut water also contains low amounts of sodium chloride as well as high amounts of sugars and amino acids. It has a balanced composition of fluids to prevent dehydration during diarrhoea.	This paragraph describes the content of coconut water in the form of minerals with the right composition that can restore body fluids lost during diarrhoea.

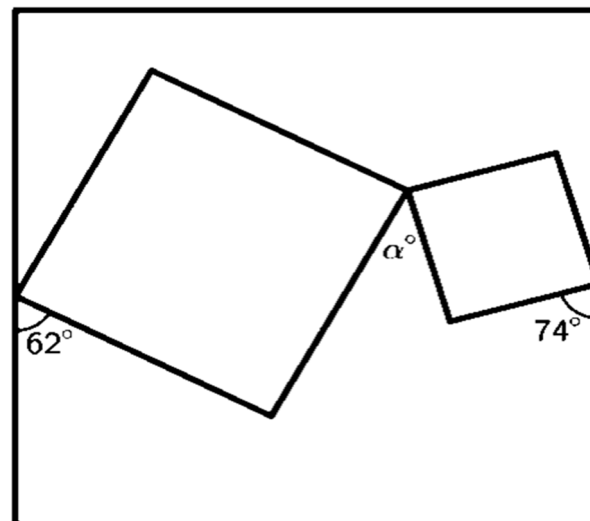
What are the benefits of coconut water for the body?

Answer	Explanation
Coconut fruit has benefits for the body. One part is that coconut water can prevent premature aging and relieve diarrhoea. Coconut water contains cytokinins that have anti-aging, anti-carcinogenic, and anti-thrombotic benefits. The content of coconut water in the form of minerals with the right composition can restore body fluids lost during diarrhoea.	The answer consists of four sentences, namely: The main idea of all three paragraphs; Summary of the first paragraph; Summary of the second paragraph; Summary of the third paragraph

Appendix C. Sample of Type A Problem Used in Experiment 2



Appendix D. Sample of Type B Problem Used in Experiment 2



References

1. Hartwig, M.K.; Rohrer, D.; Dedrick, R.F. Scheduling math practice: Students' underappreciation of spacing and interleaving. *J. Exp. Psychol. Appl.* **2022**, *28*, 100–113. [CrossRef] [PubMed]
2. Rohrer, D. Interleaving helps students distinguish among similar concepts. *Educ. Psychol. Rev.* **2012**, *24*, 355–367. [CrossRef]
3. Sana, F.; Yan, V.X. Interleaving retrieval practice promotes science learning. *Psychol. Sci.* **2022**, *33*, 782–788. [CrossRef]
4. Brunmair, M.; Richter, T. Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychol. Bull.* **2019**, *145*, 1029–1052. [CrossRef]
5. Likourezos, V.; Kalyuga, S.; Sweller, J. The variability effect: When instructional variability is advantageous. *Educ. Psychol. Rev.* **2019**, *31*, 479–497. [CrossRef]
6. Sweller, J.; Sweller, S. Natural information processing systems. *Evol. Psychol.* **2006**, *4*, 434–458. [CrossRef]
7. Sweller, J.; van Merriënboer, J.; Paas, F. Cognitive architecture and instructional design: 20 years later. *Educ. Psychol. Rev.* **2019**, *31*, 261–292. [CrossRef]
8. Geary, D.C. An evolutionarily informed education science. *Educ. Psychol.* **2008**, *43*, 179–195. [CrossRef]
9. Geary, D.C. Evolutionary educational psychology. In *APA Educational Psychology Handbook, Vol 1: Theories, Constructs, and Critical Issues*; Harris, K.R., Graham, S., Urdan, T., McCormick, C.B., Sinatra, G.M., Sweller, J., Eds.; American Psychological Association: Worcester, MA, USA, 2012; pp. 597–621. [CrossRef]
10. Geary, D.C.; Berch, D.B. Evolution and children's cognitive and academic development. In *Evolutionary Perspectives on Child Development and Education*; Geary, D.C., Berch, D.B., Eds.; Springer: Berlin/Heidelberg, Germany, 2016; pp. 217–249. [CrossRef]
11. Cowan, N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behav. Brain Sci.* **2001**, *24*, 87–185. [CrossRef]

12. Miller, G.A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.* **1956**, *63*, 81–97. [CrossRef]
13. Peterson, L.R.; Peterson, M.J. Short-term retention of individual verbal items. *J. Exp. Psychol.* **1959**, *58*, 193–198. [CrossRef] [PubMed]
14. Simon, H.A.; Gilmarin, K. A simulation of memory for chess positions. *Cogn. Psychol.* **1973**, *5*, 29–46. [CrossRef]
15. Ericsson, K.A.; Kintsch, W. Long-term working memory. *Psychol. Rev.* **1995**, *102*, 211–245. [CrossRef] [PubMed]
16. Chen, O.; Paas, F.; Sweller, J. A cognitive load theory approach to defining and measuring task complexity through element interactivity. *Educ. Psychol. Rev.* **2023**, *35*, 63. [CrossRef]
17. Sweller, J. Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educ. Psychol. Rev.* **2010**, *22*, 123–138. [CrossRef]
18. Kalyuga, S. Cognitive load theory: How many types of load does it really need? *Educ. Psychol. Rev.* **2011**, *23*, 1–19. [CrossRef]
19. Klepsch, M.; Seufert, T. Making an effort versus experiencing load. *Front. Educ.* **2021**, *6*, 645284. [CrossRef]
20. Kriegelstein, F.; Beege, M.; Rey, G.D.; Sanchez-Stockhammer, C.; Schneider, S. Development and validation of a theory-based questionnaire to measure different types of cognitive load. *Educ. Psychol. Rev.* **2023**, *35*, 9. [CrossRef]
21. Paas, F.; van Merriënboer JJ, G. Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks. *Curr. Dir. Psychol. Sci.* **2020**, *29*, 394–398. [CrossRef]
22. Skinner, B.F. The rate of establishment of a discrimination. *J. Gen. Psychol.* **1933**, *9*, 302–350. [CrossRef]
23. Carvalho, P.F.; Goldstone, R.L. The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychon. Bull. Rev.* **2015**, *22*, 281–288. [CrossRef] [PubMed]
24. Kang SH, K.; Pashler, H. Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Appl. Cogn. Psychol.* **2012**, *26*, 97–103. [CrossRef]
25. Kornell, N.; Bjork, R.A. Learning concepts and categories: Is spacing the “enemy of induction”? *Psychol. Sci.* **2008**, *19*, 585–592. [CrossRef]
26. Taylor, K.; Rohrer, D. The effects of interleaved practice. *Appl. Cogn. Psychol.* **2010**, *24*, 837–848. [CrossRef]
27. Wahlheim, C.N.; Dunlosky, J.; Jacoby, L.L. Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Mem. Cogn.* **2011**, *39*, 750–763. [CrossRef] [PubMed]
28. Chen, O.; Castro-Alonso, J.C.; Paas, F.; Sweller, J. Extending cognitive load theory to incorporate working memory resource depletion: Evidence from the spacing effect. *Educ. Psychol. Rev.* **2018**, *30*, 483–501. [CrossRef]
29. Chen, O.; Paas, F.; Sweller, J. Spacing and interleaving effects require distinct theoretical bases: A systematic review testing the cognitive load and discriminative-contrast hypotheses. *Educ. Psychol. Rev.* **2021**, *33*, 1499–1522. [CrossRef]
30. Paas, F.; van Merriënboer JJ, G. Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *J. Educ. Psychol.* **1994**, *86*, 122–133. [CrossRef]
31. Pan, S.C.; Tajran, J.; Lovelett, J.; Osuna, J.; Rickard, T.C. Does interleaved practice enhance foreign language learning? The effects of training schedule on Spanish verb conjugation skills. *J. Educ. Psychol.* **2019**, *111*, 1172–1188. [CrossRef]
32. Birnbaum, M.S.; Kornell, N.; Bjork, E.L.; Bjork, R.A. Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Mem. Cogn.* **2013**, *41*, 392–402. [CrossRef]
33. Daneman, M.; Carpenter, P.A. Individual differences in working memory and reading. *J. Verbal Learn. Verbal Behav.* **1980**, *19*, 450–466. [CrossRef]
34. Chen, O.; Paas, F.; Sweller, J. Reply to Sana et al.’s (2022) Commentary on rest-from-deliberate-learning as a mechanism for the spacing effect. *Educ. Psychol. Rev.* **2022**, *34*, 1851–1858. [CrossRef]
35. Leppink, J.; Paas, F.; van der Vleuten CP, M.; van Gog, T.; van Merriënboer JJ, G. Development of an instrument for measuring different types of cognitive load. *Behav. Res. Methods* **2013**, *45*, 1058–1072. [CrossRef] [PubMed]
36. Klepsch, M.; Schmitz, F.; Seufert, T. Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Front. Psychol.* **2017**, *8*, 1997. [CrossRef] [PubMed]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Procedural Learning in Mixed Reality: Assessing Cognitive Load and Performance

Ghislain Mugisha ^{1,2} and Amael Arguel ^{2,*}

¹ G-SCOP, University Grenoble Alpes, CNRS, Grenoble INP, 38400 Grenoble, France; ghislain.mugisha@grenoble-inp.fr

² CLLE, University of Toulouse, CNRS, 31400 Toulouse, France

* Correspondence: amael.arguel@univ-tlse2.fr

Abstract: Immersive technologies offer promising advancements in medical education, particularly in procedural skill acquisition. However, their implementation often lacks a foundation in learning theories. This study investigates the application of the split-attention principle, a multimedia learning guideline, in the design of knot-tying procedural content using a mixed reality (MR) technology, specifically Microsoft HoloLens 2. A total of 26 participants took part in a between-group design experiment comparing integrated and split-source formats for learning arthroscopic knots, with the performance and the cognitive load assessed. The initial hypotheses were not confirmed, as results did not show significant differences in performance during recall, nor in extraneous and germane cognitive load. However, the findings on intrinsic cognitive load highlight the complexity of participant engagement and the cognitive demands of procedural learning. To better capture the split-attention effect, future research should address the high element interactivity in MR representations. The study provides some foundation for designing procedural simulation training that considers both learners' needs and cognitive processes in highly immersive environments. It contributes to the ongoing exploration of instructional design in MR-based medical education, emphasizing both the potential and challenges of multimedia learning principles in advanced technological contexts.

Keywords: cognitive load theory; surgical gesture learning; arthroscopic; mixed reality; split-attention effect; learning outcome

Academic Editors: Fred Paas, Andre Tricot, Kim Ouwehand and Florence Lespiau

Received: 22 January 2025

Revised: 22 February 2025

Accepted: 5 March 2025

Published: 10 March 2025

Citation: Mugisha, G., & Arguel, A. (2025). Procedural Learning in Mixed Reality: Assessing Cognitive Load and Performance. *Education Sciences*, 15(3), 339. <https://doi.org/10.3390/educsci15030339>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Technological advancements have propelled immersive technologies into vital roles across various fields, including education (Chen, 2023). Extended reality (XR) encompasses (Milgram & Kishino, 1994) virtual reality (VR), augmented reality (AR), and mixed reality (MR), each offering distinct levels of digital integration into real-world environments. VR creates fully virtual spaces (Allcoat et al., 2021), AR overlays digital elements onto physical settings, and MR enables seamless interaction between the two (Park et al., 2021; Pimentel et al., 2022). Given the discussions on MR's definition (Milgram & Kishino, 1994; Parong, 2021), we separated VR, AR, and MR. These technologies, particularly MR tools, are increasingly recognized for their potential to enhance healthcare systems (Masson, 2023) and may improve medical training (Silvero Isidre et al., 2023).

Advantages presented by MR are abounding. Some scholars assume that surgical training should no longer rely on patient-based learning (Mayo, 1927), since MR tools can enhance traditional training methods (de Sá et al., 2022) by enabling surgical technique practices (Grantcharov, 2008; Mergen et al., 2024) including high-risk procedures (Asoodar

et al., 2024; Forgione & Guraya, 2017; Nagayo et al., 2022). In fact, procedural learning (such as suturing and knot-tying) may benefit from MR especially in countries where certain techniques like arthroscopic knot-tying are absent from curricula (Lacroix et al., 2021). Overall, these technologies offer some responses to ethical, economic, and time-related issues in medical training (Forgione & Guraya, 2017; Nagayo et al., 2022).

Nevertheless, their integration into the surgical education must tackle some challenges, sparking discussions on their effectiveness and appropriate implementation (Co et al., 2023). Some educators, it is argued, adopt these tools without effectively leveraging them to enhance and support meaningful learning experiences (Sandars et al., 2015). Others noted that despite the abundance of learning theories, few are designed to help educators integrate procedural simulations into their curricula (Rivière et al., 2018).

Thus, researchers have emphasized the importance of grounding the educational designs of MR techniques in empirical evidence, especially in medical training (Issa et al., 2011; Mayer, 2020). For example, studies have shown that learning was more effective when instructional materials included both words and images, compared to text-only formats (Mayer, 2002). Poorly designed educational content can hinder learning by increasing extraneous cognitive load (Parong, 2021). This underscores the significance of applying multimedia learning principles and the Cognitive Theory of Multimedia Learning (CTML) (Mayer, 2021) to minimize the unnecessary cognitive load and manage task complexity (Rivière et al., 2018). Additionally, the importance of monitoring the learners' cognitive load during procedural simulations, such as suturing (McInnis et al., 2021), is highlighted, with cognitive load theory (CLT) emerging as a particularly relevant framework (Fraser et al., 2015).

Against this backdrop, this paper investigated the application of the split-attention principle in the design of knot-tying procedural content using a specific MR technology, HoloLens 2 (Palumbo, 2022). The principle is anchored in CLT (Ayres & Sweller, 2021) and CTML, and has shown potential for designing procedural simulations in a MR medium. In the following sections we succinctly present this principle, which is the focus of our attention.

1.1. Design of Learning Scenarios: The Case of the Split-Attention Principle

Multimedia learning principles guide the design of learning content to support cognitive processes. These principles can be categorized into three groups: (1) minimizing the extraneous cognitive load (ECL) by limiting unnecessary information, (2) focusing the learners' attention on essential information, and (3) promoting generative processing to help the learners synthesize information into meaningful representations (Mayer & Fiorella, 2021).

For instance, the split-attention principle reduces ECL by presenting essential learning elements spatially close to each other, thereby minimizing the need for learners to switch between different information sources (Ayres & Sweller, 2021). Studies have shown that integrated content formats, like those used to learn the kidney structure (Cierniak et al., 2009) and orthopedic physical therapy (Pociask & Morrison, 2008), led to a better performance than split-source formats. Split-source formats increased the ECL (Cierniak et al., 2009), making tasks more difficult and requiring more concentration, but had no impact on task completion times (Pociask & Morrison, 2008).

The recent studies in immersive environments explored the application of this principle in physics lab experiments, and the results of these pioneering studies were mixed. Krüger and Bodemer (2022) found no significant effects of spatial contiguity (a form of split-attention principle) on the cognitive load and the declarative knowledge in an AR-based botanical garden course, although there was a reduction in temporal demand and an increase in perceived performance. In contrast, Thees et al. (2020) found a significant

reduction in the ECL with an MR-adapted integrated condition, leading to better conceptual knowledge compared with a traditional setup. However, another study comparing a HoloLens (integrated condition) with an iPad (separate-display condition) did not show the expected benefits (Thees et al., 2022).

Despite these variations, Ginns (2006) reported a large effect size ($d = 0.85$) in a meta-analysis of spatial contiguity, which includes the split-attention effect, suggesting that integrated formats generally improve learning by reducing the ECL. The split-attention and spatial contiguity principles are essentially the same (Schroeder & Cenkci, 2018), with the latter being more specific, as emphasized by the CLT proponents (Fiorella & Mayer, 2021).

Moreover, there is evidence in the literature that the split-attention principle has yet to be explored in the context of MR-based medical, procedural training, especially as, to the best of our knowledge, no studies have been conducted regarding its implementation.

It is essential to briefly address two key theories that underpin the split-attention (or spatial contiguity) principle: CLT and its counterpart, CTML.

1.2. Cognitive Load Theory (CLT)

CLT (Sweller et al., 1998, 2011) aligns with classical cognitive architecture models, which propose that the cognitive system includes a limited-capacity working memory (WM) for learning and conscious thought, and an unlimited long-term memory (LTM) that stores automated schemas (Chanquoy et al., 2007; Tricot, 1998).

During learning, two key mechanisms are involved—the acquisition of schemas and the transfer of learned procedures from controlled to automated processing. Initially, knowledge is organized into schemas, which guide how the new information is processed. Over time, the handling of this information shifts to automatic processing, and without this automation performance remains slow, clumsy, and error-prone (Sweller, 1994). The CLT offers a theoretical framework to design surgical learning environments that account for the cognitive processes involved in learning (Tokuno et al., 2023).

CLT identifies three types of the cognitive load (CL) that affect working memory: (1) intrinsic cognitive load (ICL), related to the processed information, (2) extraneous cognitive load (ECL), caused by the information presentation, and (3) germane cognitive load (GCL), which relates to the effort needed to process and integrate relevant information into meaningful schemas (Chanquoy et al., 2007). These distinctions stress the importance of designing instructional scenarios that consider the learners' CL, particularly in complex environments where an excessive load can hinder learning (Fraser et al., 2015).

In their systematic review, Çeken and Taşkın (2022) reported that most CL assessments relied on subjective measures, such as questionnaires, or indirect objective measurements. Thus, some subjective scales assessed the intensity of mental effort perceived by learners (Paas, 1992), while others attempted to capture all three types of CL (Leppink et al., 2013, 2014). However, those studies raised concerns about the psychometric and theoretical validity of these approaches in measuring and managing CL in educational contexts (Schroeder & Cenkci, 2020).

1.3. Cognitive Theory of Multimedia Learning (CTML)

We can argue that the CLT considers the relevant aspects of human cognitive architecture when designing good instructional materials (Paas & Sweller, 2021), shedding some light on how the different cognitive loads affect working memory capacity.

The CTML (Mayer, 2021), partly shaped by the CLT (Camp et al., 2021), focuses on how people can learn meaningfully from multimedia materials and asserts three types of demands on the limited cognitive capacity during learning (extraneous, essential, and generative processing), leading to the following three instructional goals: to reduce ex-

traneous processing, to manage essential processing, and to foster generative processing (Mayer, 2024). Indeed, the CLT has helped provide the CTML with a solid foundation in instructional design research, and some CTML principles (such as the spatial contiguity principle) are subsets of the split-attention effect, originally developed by CLT's tenets, which have evolved (Camp et al., 2021). Furthermore, the concept of multimedia has evolved (Camp et al., 2021), with recent works illustrating how the CTML is being applied in extended reality learning technologies (Altmeyer et al., 2020, 2021; Laumann et al., 2024; Strzys et al., 2018; Thees et al., 2020, 2022).

The deployment of immersive technologies in education underscores the need to address CL and optimize the learning content (Buchner et al., 2022; Poupard et al., 2024), yet many studies still overlook the established learning theories or frameworks (Radianti et al., 2020; Wu et al., 2020). Recently, Makransky and Petersen (2021) introduced the cognitive affective model of immersive learning (CAMIL), emphasizing the integration of cognitive and emotional factors to enhance the learning outcomes in virtual environments through thoughtfully designed immersive experiences.

1.4. Learning Surgical Procedures in Medical Education

Chiniara et al. classified simulations like suturing and knot-tying as procedural simulations, aimed at developing technical skills for real-life applications (Chiniara et al., 2013). For these critical tasks, researchers explore suitable learning technologies, ranging from traditional to innovative.

Previous studies on the procedural learning of knot-tying have compared 2D laptop-based presentations with 3D VR or MR environments (Nagayo et al., 2021; Yoganathan et al., 2018), but these often lacked of grounding in educational theories. Furthermore, the prevailing medium comparison approach has been criticized for the difficulty of setting up experimental conditions that are comparable in all aspects, and also focus on global effects rather than the nuanced features and interactions inherent in specific technologies (Buchner et al., 2022; Gonnermann-Müller & Krüger, 2024). Beyond these limitations, most past research has been conducted in less immersive 2D environments, highlighting the need for new investigations into the emerging educational contexts (Çeken & Taşkın, 2022).

Immersive technologies are widely used in medical education, particularly for anatomy and surgical training (Asoodar et al., 2024), with features like immersion and interaction (Makransky & Petersen, 2021). MR appears to be particularly well-suited for surgical training, as it offers a higher degree of interaction with 3D information within a real-world environment, compared with AR and VR (Sánchez-Margallo et al., 2021; Bautista et al., 2023). Recent studies have highlighted the growing use of MR in operating rooms, where it offers benefits such as real-time visualization of anatomical data, information overlay, and medical training applications (Magalhães et al., 2024).

The Microsoft HoloLens 2 is predominant in these contexts, and is used in various surgical procedures (Gregory et al., 2018; Liu et al., 2019). HoloLens 2, along with the Microsoft Dynamics Guide 365 software, enables the creation of personalized training guides using holograms, videos, images, and 3D models, making it valuable for medical education and simulation (Park et al., 2021; White, 2019).

A randomized controlled study conducted in Japan assessed the usability of an MR system for learning (Nagayo et al., 2022), in which 43 medical students with no prior suturing experience performed a subcuticular suture on synthetic skin using either a HoloLens 2 or a laptop. Performance scores showed no significant differences, but participants preferred the HoloLens 2 for its ability to merge virtual information with real objects, thereby aiding in surgical instrument manipulation. These findings suggest that the novelty effect

(Parong, 2021) of technologies like the HoloLens 2 may hinder procedural learning (Nagayo et al., 2022).

Despite their growing popularity, MR technologies must overcome potential pitfalls (Pimentel et al., 2022). A too technocentric approach, focusing on technology over the learners' cognitive needs (Mayer & Fiorella, 2021), often creates a gap between expectations and realities, sometimes leading to rejection.

1.5. Current Study

This research may be classified as a preliminary investigation aimed at addressing the various limitations raised, and functioning as a proof of concept to explore the split-attention effect within the context of knot-tying simulations utilizing MR technology. In alignment with the prior studies to which we alluded previously, we postulated that:

H1: *CL scores of participants learning knot-tying in the non-split-attention condition will be significantly lower than those in the split-attention condition.*

H2: *Performance scores of participants learning knot-tying in the non-split-attention condition will be significantly higher than those in the split-attention condition.*

H3: *Execution time scores of participants learning knot-tying in the non-split-attention condition will be significantly lower than those in the split-attention condition.*

H4: *Average learning time of participants learning knot-tying in the non-split-attention condition will be lower than those in the split-attention condition.*

2. Materials and Methods

2.1. Participants, Sampling, and Measures

The inclusion criteria of the participants were as follows: individuals aged 18 and more, a minimum B2 proficiency in French, low and no experience in knot-tying, and new users of mixed-reality headsets. The exclusions criteria were physical or mental health issues, motor or visual impairments, epilepsy, heart conditions, phobias, and cybersickness risks. To counter potential bias, we also excluded individuals pursuing medical education, as this could have exposed them to knot-tying and suturing practice. Indeed, the study plan was adapted to different issues which will be addressed in the limitations section.

A total of 26 participants (women = 14, men = 12; $M_{age} = 23$, $SD_{age} = 2.735$; 11 pursuing bachelor's degrees, 15 pursuing master's degrees; 24 advanced and 2 intermediate French speakers, with low expertise in knot-tying: $M_{expertise} = 1.8$, $SD_{expertise} = 0.962$) were recruited for the experiment, primarily through online advertisements, social media associated with the university, and direct contact. The results of one participant were excluded due to interruptions during the session. All participants were students, mainly in psychology, who received credits for their participation. Some students were in the science, technology and mathematics (STM) field. Experimental sessions were conducted from May to June 2023 in France, on the campus of a public university focusing on humanities and social sciences.

Participants were randomly assigned to either of the two conditions [independent variable (IV)]: 13 in the non-split-attention/integrated condition (C) and 13 in the split-attention/separated condition (NC). A counterbalancing method minimized the transfer effects in knot recall by using computer-generated sequences for up to 100 participants (Appendix I). Participants completed the informed consent and socio-demographic Limesurvey questionnaires using a laptop. The entire procedure is summarized (Figure 1);

the experiment was planned to last 45 to 60 min per participant, depending on the time each participant chose to spent on learning knots.

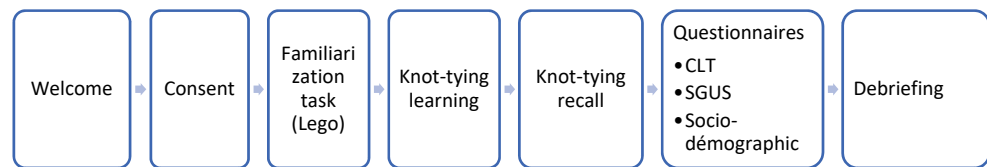


Figure 1. Experimental procedure.

The separated condition (Figure 2) required the participants to alternate their visual attention between a holographic learning screen positioned behind them (at 180°) and the knot-tying task in front of them, while the integrated condition (Figure 2) enabled simultaneous viewing of both elements in front of them.



Figure 2. Illustration of the MR view of a participant knot-tying in the following conditions: the separated condition (**left**), and integrated condition (**right**).

The CL (measured using a 7-item Likert scale, with a 10-point scale for each item ranging from 1 (very low) to 10 (very high)) and performance (knot quality and execution time) were the dependent variables (DVs). The CL was assessed using a subjective Likert scale (Leppink et al., 2013, 2014; Ouwehand et al., 2021) adapted for the present study (Appendix F: Table A5). Leppink et al. (2013) developed a 10-item scale to assess ICL, ECL, and GCL, initially linked to statistical knowledge but applicable across other domains, including language learning (Leppink et al., 2014). Overall, the modifications made, such as adding items and language translation, did not challenge its internal structure (Thees et al., 2020). For this study, the adaptations included retaining only the following 7 items from the original 10 items (previously 13 items) and their translation into French: ICL (items 2: Complexity and 4: Mental Effort-Complexity), ECL (items 5: Clarity, 7: Efficiency, and 8: Mental Effort-Clarity & Efficiency), and GCL (items 12: Knowledge & Understanding, 13: Mental Effort-Knowledge & Understanding). We conducted an exploratory factor analysis to assess the internal structure (Appendix D: Table A4) of the instrument modified. For the three subscales, excellent reliability was obtained for the ICL and ECL ($\alpha_{ICL} = 0.826$, $\alpha_{ECL} = 0.735$) but not for the GCL ($\alpha_{GCL} = 0.0665$). Overall, the instrument demonstrated acceptable reliability for this research ($\alpha_{ICL,ECL,GCL} = 0.744$).

For the performance assessment, the quality of knot execution was evaluated using a rating grid created by authors, after the assessment of the video recordings of knot-tying (each knot was tied in 3 scenarios, divided into 3 sub-scenarios). Each correctly performed task in a sub-scenario earned one point (otherwise it earned zero). The execution time was recorded using a stopwatch.

The user satisfaction with the HoloLens 2 was assessed with an 11-item (rated on a 10-point Likert scale (1 = very low to 10 = very high) Smart Glass User Satisfaction

(SGUS) questionnaire (Helin et al., 2018; Xue et al., 2019) (Appendix G: Table A6). It evaluated the user experience with smart glasses, focusing on MR perception, interaction, and usability. Grounded in web-based learning and mobile MR evaluation frameworks, it assesses interface design, MR-specific features, and learning effectiveness (Xue et al., 2019). An exploratory factor analysis was used to assess its internal structure (Appendix H: Table A7) and had an excellent reliability: $\alpha = 0.860$.

2.2. Materials and Experimental Design

Participants viewed three knot-tying scenarios using a HoloLens 2. The knots chosen were of the following sliding types: Samsung Medical Center knot (SMC) (Kim & Ha, 2000), Double Twist knot (Rolla & Surace, 2002), and Duncan loop (Nottage & Lieurance, 1999). The knots were selected for their relevance in surgical training (Lacroix et al., 2021) and for their varying difficulty levels (Baumgarten & Wright, 2007). We based our design of the knot videos on traditional step-by-step instructions (Figure 3), dividing each video into three key moments before uploading them to the Dynamics 365 Guide. The application had 2 screens: one of the virtual screens displayed instructional reminders, reiterating the previously provided guidelines to help participants progress through the learning process (e.g., “Click the arrow to move to the next sub-step of the knot-tying procedure”). On the other screen, the video of the procedure was projected. Notably, audio guidance was deliberately omitted from the instructional videos.

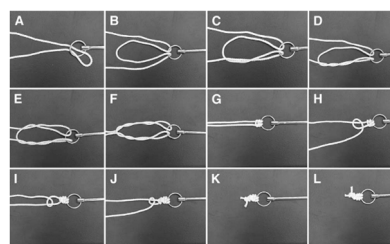


Figure 3. An example of a step-by-step instructional method for the Double Twist knot developed by Rolla and Surace (2002) used to design the Double Twist video which was divided into three key moments (A–L).

To begin, each participant was required to scan a QR code related to a task by looking directly at it and staying about (12–20 inches) away from it. To navigate through different steps (stop, move forward, go back), the participant used eye-tracking interactions. Before starting the knot-tying task, the participants completed a familiarization task consisting of a Lego assembly procedure to learn how to navigate the HoloLens interface for about 15–20 min (Figure 3).

The Lego task and watching instructional videos independently helped the participants practice using the HoloLens commands. Knot-tying was then conducted using multicolored cords attached to adhesive plastic hooks. To prevent confusion, each type of knot was assigned a specific cord color: pink (Duncan), green (SMC), or blue (Double Twist). The videos were designed, recorded using an iPhone 13 Mini, and then edited for instructional purposes. The mixed-reality setup allowed for the real-time monitoring of the participant’s perspective, which also enabled capturing photos/videos (Figure 4).

During the learning stage of knot-tying, participants had up to three rounds for each knot, with the time recorded but no limits set. Specifically, after completing the first round, participants could perform two additional rounds—only then could they use the option to go back. However, during the knot-tying learning task (with each knot consisting of three sub-steps), they could not go back to view a sub-step that had already been completed. Statistical analyses were conducted using Jamovi and LimeSurvey for questionnaires.

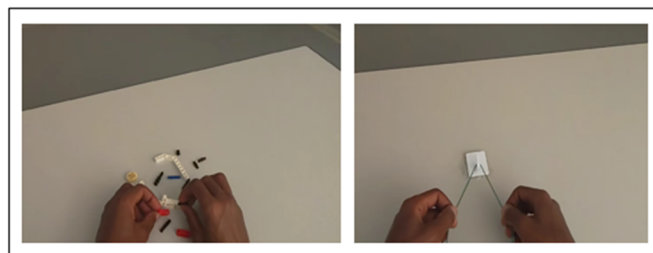


Figure 4. Monitoring of the Lego familiarization task (left) and the knot-tying task during learning time (right)—the computer’s view screenshot.

Globally, the analyses included *t*-tests for independent samples, with the normality verified using the Shapiro–Wilk test and the homogeneity of variances using Levene’s test. If the assumptions were not met, the Mann–Whitney U test was used instead. A pilot study with 2 volunteers (1 Man, 1 Woman; $M_{age} = 22$; bachelor’s degree; no prior experience in knot-tying; novice with an MR tool; from psychology and STM fields) allowed the assessment of questionnaire clarity, image quality, and experiment duration.

3. Results

3.1. Cognitive Load

The descriptive results of each item are reported (Appendix A: Table A1), for each category of CL (Table 1). The results of the *t*-test or the Mann–Whitney test are provided, considering $p < 0.05$ as the threshold for statistical significance.

Table 1. Descriptive statistics of ICL, ECL, GCL in different conditions.

		95% CI				
	Conditions	Mean	Lower	Upper	Median	SD
ICL	C	4.846	3.775	5.917	5.500	1.772
	NC	6.269	4.843	7.695	6.500	2.360
ECL	C	2.795	1.663	3.927	2.333	1.873
	NC	3.846	2.804	4.888	4.000	1.725
GCL	C	7.000	6.240	7.760	7.500	1.258
	NC	6.692	5.585	7.799	7.500	1.832

Overall, although the mean CL scores may suggest some benefit of the integrated condition for ECL, with the confidence interval CI 95% $[-\infty; 0.257]$, no statistically significant difference was detected, $t(24) = -1.489$, $p = 0.075$, Cohen’s $d = -0.584$. To obtain a clear estimate of CI, we applied bootstrap resampling (1000 iterations) (Appendix E) which provided CI $[2.333; 4.167]$. For the GCL, $U = 77.5$, $p = 0.651$, no statistically significant difference was noted (CI 95% $[-\infty; 1.5]$ effect size $r = 0.083$).

However, with 95% CI $[-\infty; -0.023]$ after bootstrap resampling $[4.5; 6.5]$ for the ICL, the study found a statistically significant difference in favor of the integrated condition: $t(24) = -1.423$, $p = 0.047$, with a moderate effect size, Cohen’s $d = -0.682$.

3.2. Overall Performance Related to Knot Quality During Recall Phase

We began to assess the quality of knots in each sub-step of the recall phase. From this we calculated the overall recall scores for each participant, for each knot, which were as follows: Qt SMC (overall score for the SMC knot), Qt Double Twist (overall score for the Double Twist knot), and Qt Duncan (overall score for the Duncan). The descriptive results are provided below (Table 2). We detected no statistically significant differences with Qt

SMC, $U = 72$, $p = 0.762$; Qt Double Twist, $U = 78$, $p = 0.326$; Qt Duncan, $U = 60$, $p = 0.082$, with, respectively, $r = 0.1479$, $r = 0.652$, and $r = 0.0769$.

Table 2. Descriptive statistics of variables related to knot quality in different conditions of presentation.

		95% CI						
	Conditions	Mean	Lower	Upper	Median	SD	Minimum	Maximum
Qt SMC	C	1.231	0.728	1.73	1	0.832	0	2
	NC	1.462	0.781	2.14	2	1.127	0	3
Qt Double Twist	C	1.231	0.966	1.50	1	0.439	1	2
	NC	1.154	0.927	1.38	1	0.376	1	2
Qt Duncan	C	2.538	2.069	3.01	3	0.776	1	3
	NC	2.000	1.396	2.60	2	1.000	1	3

The overall performance was assessed by summing the points each participant earned across all three sequences: Quality = Qt SMC + Qt Double Twist + Qt Duncan. The results indicate no significant difference: $U = 75$, $p = 0.316$.

3.3. Performance Related to Knot Execution Time During Recall Phase

We measured the time taken by each participant to complete each of the three knots. The descriptive statistics for these results are presented (Table 3). No statistically significant differences were detected for SMC (CI 95% [7; ∞] after bootstrap resampling (Appendix E) [31; 66]) $U = 37$, $p = 0.993$, with a moderate effect size ($r = 0.562$). Similarly, for, respectively, Double Twist (CI 95% [−42; ∞] after bootstrap resampling (1000 iterations) [102, 158] and Duncan (CI 95% [2.94; ∞] gave [30; 39]), no significant differences were observed: $t(24) = 0.141$, $p = 0.555$, $d = 0.055$ (small effect); and $t(24) = 2.142$, $p = 0.979$, $d = 0.84$ (large effect). In contrast, mean scores were in favor of the integrated condition.

Table 3. Descriptive statistics of recall time variable according to presentation conditions.

		95% CI						
	Conditions	Mean	Lower	Upper	Median	SD	Minimum	Maximum
Recall SMC	C	79.9	50.4	109.4	67	48.8	23	177
	NC	38.4	28.1	48.7	31	17.0	17	69
Recall Double Twist	C	144.1	97.4	190.8	125	77.3	32	270
	NC	140.3	105.4	175.2	125	57.8	79	288
Recall Duncan	C	48.2	36.5	59.8	39	19.2	23	81
	NC	33.5	24.2	42.8	32	15.4	17	71

3.4. Learning Time for Knot-Tying

Participants were allowed to complete up to three rounds of learning. All participants completed the 1st round (T1); however, most chose not to proceed with the 2nd and/or the 3rd round. Descriptive statistics for all rounds are elaborated in (Appendix B: Table A2).

For T1, we conducted statistical analyses on the SMC, Double Twist, and Duncan knots by applying the Mann–Whitney U test (Table 4). Given the one extreme unclear value of CI, we also applied bootstrap resampling (1000 iterations) to estimate 95% CIs (Appendix E). The participants in the integrated condition completed the tasks in less time than those in the separated condition. Specifically, the different times required to complete the knots were [95% CI: SMC [4, 8.5], Double Twist Knot [8, 17], and Duncan Knot [2.513, 4.0]]. Indeed, the results revealed a statistically significant difference between the conditions, with moderate effect sizes (Table 4).

Table 4. Inferential statistics relating to the first round learning time under different conditions.

		Statistic	<i>p</i>	Effect Size
T1 SMC	Mann–Whitney U	47.5	0.030	0.438
T1 Double Twist	Mann–Whitney U	51.0	0.044	0.396
T1 Duncan	Mann–Whitney U	47.0	0.025	0.444

3.5. User Satisfaction with the Use of HoloLens 2

The descriptive statistics for the satisfaction scores are presented (Appendix C: Table A3), with the unexpected observation that mean scores for certain SGUS items (participants' satisfaction) were slightly higher in favor of the separated condition.

Overall, the participants reported general satisfaction with the use of the HoloLens 2 for the task across the two conditions. There were no significant differences observed, and the results were as follows: GL1: $U = 72.0$, $p = 0.750$; GL2: $U = 82.0$, $p = 0.458$; GL3: $U = 79.5$, $p = 0.405$; GL4: $U = 64.0$, $p = 0.862$; GL5: $U = 76.0$, $p = 0.332$; GL6: $U = 76.5$, $p = 0.675$; GL7: $t(24) = -0.927$, $p = 0.819$; GL8: $U = 77.5$, $p = 0.363$; GL9: $t(24) = -0.519$, $p = 0.696$; GL10: $U = 70.5$, $p = 0.235$; GL11: $U = 76.0$, $p = 0.681$).

4. Discussion

Our study aimed to examine the impact of two presentation formats of knot-tying procedures using MR HoloLens 2. Grounded in CLT and CTML, expectations were that the integrated/non-split-attention format would prove benefits pertaining to cognitive load and performance. The findings did not reveal significant differences for the ECL, nor the GCL; notably, a significant effect was detected for the ICL. Although no significant differences were observed in terms of the quality of knot realization during recall and the time needed to perform the task, some significant differences were noted for the first learning round of all of the knots.

4.1. Empirical Contributions

These results contrasted with the previous studies conducted in traditional learning environments, which demonstrated benefits for integrated formats in terms of performance and CL (Ginns, 2006; Schroeder & Cenkci, 2018). Before proceeding further, it is necessary to address the surprisingly significant difference in intrinsic cognitive load (ICL) observed over the course of the experiment.

We noted a marked difference in how the participants engaged with the complexity of the procedures studied. This finding was unexpected, as most prior research has not reported such discrepancies, except for Laumann et al. (2024), who identified significant ICL differences attributed to the learners' prior knowledge. In the context of this study, it is important to highlight that declarative knowledge appears to be the primary focus of the studies implementing the split-attention principle. Declarative knowledge is largely associated with declarative memory, which stores facts and events. One proposed explanation is that declarative knowledge requires directed attention (ten Berge & van Hezewijk, 1999). During the experiment, we observed that procedural knowledge, although reliant to some extent on declarative knowledge (e.g., user manuals, cooking recipes) in various forms (verbal, visual), is closely tied to skill acquisition and pattern recognition for goal completion. Repeatedly practicing a sequence of actions enhances performance (Wit et al., 2023), as over time the skills become automatic and require less attention (ten Berge & van Hezewijk, 1999). This observation raises three key considerations regarding the ICL and the ECL.

First, the differences in the ICL may be attributed to the number of training rounds participants chose to complete in order to master the procedures. Our analysis of participant

engagement shows that although the number of repetitions decreased over time, the participants in the integrated condition were more likely to complete the second and third rounds of training. This may have contributed to perceiving the task's complexity and the associated mental effort as lower.

Second, concerning ECL, the procedural learning characteristics facilitated by the HoloLens may have mitigated the split-attention effect, in contrast to other studies primarily focused on declarative knowledge and conducted in a medium-comparison format. Participants could complete the procedure without frequently shifting attention between the screen, their hands, and the cords used to tie the knots. This differs from situations requiring continuous attention shifts between multiple essential sources of information, such as manuals, texts, or virtual representations, to successfully complete a task, such as recalling knowledge (Pociask & Morrison, 2008) or responding to heat and temperature conduction (Thees et al., 2020). This considers the necessity of a high element interactivity (Leahy & Sweller, 2019) if any effect is to be detected. Certainly, even with a large sample size, if there are few connections between the sources of information, even when the distance between the elements is there (Ayres & Sweller, 2021), the research may come to null findings.

Regarding the results for ECL, a certain degree of split-attention was implemented; as we observe the mean scores of the participants, we may presume that for longer or more complex tasks, the split-attention effect might become more pronounced. However, the lack of significant differences in the performance and cognitive load suggests the possibility of a Type II error, where we fail to reject the null hypothesis despite there being a true effect. This occurs if the study is underpowered, resulting in insufficient data to detect a significant difference when one exists. The near-threshold p -value and effect size ($p = 0.075$; $d = -0.584$) could indicate that a larger sample size and improved design might have revealed significant effects.

Third, a challenge arises when comparing the prior research findings to our study, which focuses on procedural learning, whereas most prior studies measure declarative knowledge. Cierniak et al. (2009) observed a learning benefit for an integrated presentation, with the split-attention effect noted only when measuring declarative knowledge about nephron facts, but not when performing inference tests. This issue also applies to Krüger and Bodemer's study (Krüger & Bodemer, 2022) on the declarative learning of botanical material, which showed no reduction in the cognitive load but an increase in perceived performance. We suggest that this reflects the broader trend in immersive environment studies, typically focused on medium comparisons and declarative learning tasks.

Meanwhile, the results of the study align with the recent studies on the application of multimedia learning principles in immersive environments, which also failed to find conclusive results on reducing cognitive load or improving performance (Strzys et al., 2018; Altmeyer et al., 2020; Thees et al., 2020; Altmeyer et al., 2021; Thees et al., 2022).

Having discussed how the study relates to the previous theoretical and empirical research, it seems necessary to note some other points that may account for the results obtained.

The variability in the amount of learning rounds selected by the participants could be an indirect indicator of perceived task difficulty or a reflection of intrinsic motivation. As noted, intrinsic motivation can be driven by interest or challenge (Ryan & Deci, 2000). When participants complete the first round successfully, they may perceive the task as sufficiently mastered or uninteresting, which reduces their motivation to continue. Conversely, those who find the task more challenging may engage in additional rounds to reinforce their skills. As highlighted in the motivation research, beliefs such as self-efficacy and the perceived value of a task significantly influence effort investment (Feldon et al., 2019). These factors,

combined with the cognitive appraisal of task complexity and utility, appear to shape the participants' decisions to repeat or abandon the subsequent learning stages.

Moreover, the self-regulation strategies of the participants likely influenced the results (Zimmerman, 2002). During the experiment, some participants mentally rehearsed the steps or mimicked them before deciding whether to proceed, demonstrating metacognitive skills. These strategies may have reduced the ECL by enabling the learners to focus on essential procedural aspects, thereby minimizing the complexity of the control condition. This raises questions about the interplay between self-regulation strategies and the cognitive load (de Bruin et al., 2020; Seufert, 2018).

Exploratory findings indicated that the participants were generally satisfied with the presentation and interaction with the content, aligning with (Xue et al., 2019). However, certain items, such as GL1 and GL7, showed higher mean scores in the control condition suggesting areas that warrant further exploration.

4.2. Practical Implications

Given the amount of research that promotes the value of mixed reality for surgical training (Barsom et al., 2016; Lu et al., 2020; Nagayo et al., 2021; Magalhães et al., 2024; Gonnermann-Müller & Krüger, 2024), the question pertains of whether our study calls into question its benefits. One should be balanced, and avoid extremes such as a technocentric approach and a non-informed learner-centered approach. Avoiding the latter extreme prevents the study from being naïve and ignoring various limitations, and then trying to fix them by using educational theories.

The cost and development of MR tools are crucial in surgical training, requiring institutions to assess the value of their financial investment (Parong, 2021). Effective MR development involves collaborations among educators, technologists, and healthcare professionals to develop relevant learning scenarios. While MR tools like the HoloLens offer much affordance, their battery time limits varying with the brightness levels can hinder practicality. A balanced approach is essential to integrate MR tools with traditional training methods.

The study explored the contribution of cognitive and educational psychology theories on surgical training (Sullivan, 2020), specifically by proposing a basic study to adapt the split-attention principle, which is currently underexplored in VR/AR/MR environments (Çeken & Taşkın, 2022), and in the design of a surgical course scenario. Using the Dynamics Guide 365 on the HoloLens 2, we developed procedural learning guides incorporating text and video demonstrations. While not utilized in this study, the possibility of integrating 3D holographic models offers a promising direction for the enhancement of knot-tying instructions.

4.3. Limitations and Future Directions

One of the main limitations of our study is the sample size and the lack of medical students among the participants, which can hinder the generalization of results.

During the study planning phase, we had to address the challenge of not being able to reach the intended sample size. Various constraints prevented us from achieving this goal, despite considering key works such as Ginns' meta-analysis (Ginns, 2006). Barriers included the conditions of the controlled experiment, which required individual testing sessions, as well as material constraints (a single HoloLens device that required recharging, especially in high light conditions). Also, as it was necessary to control some environmental factors (such as lighting issues and noise disturbances) by managing the logistical limitations (availability of a testing room in the university laboratory) and the time required for each session (at least one hour), an adjustment of our initial approach was imperative. To cope with the fact that the study was

not conducted in the context of a classroom or laboratory session (which is achieved in most studies, and provides the possibility to recruit more participants), the decision was to rely on somewhat of a convenience sample, to ensure relevance and feasibility.

As a pilot study, the current research provides food for thought. To address this, one could, for instance, rely on Ginns' major meta-analysis (Ginns, 2006) and perform sample size calculations (effect size: 0.72, error threshold: 0.05, statistical power: 0.8), which indicate that a minimum of 50 participants are required. It provides perspective on how to assess the split-attention, not only by considering the high element interactivity but also the prior experiences of participants.

The participants' prior experiences with XR tools were not accounted for, despite some reporting familiarity with such technologies. This potentially influenced the results through a novelty effect (Parong, 2021). Additionally, the influence of prior knowledge is crucial if any effect is to be observed, otherwise there is a risk of reverse results (Ayres & Sweller, 2021).

The absence of a maximum learning time likely introduced variability, affecting performance and cognitive load. Additionally, the absence of a standardized evaluation grid for knot-tying may have affected the reliability and comparability.

Future research should address these limitations and explore the following areas to optimize methodology and enrich findings. First, it is of utmost importance to conduct the research with a larger sample size, and to consider medical students as the population of interest, since they are the ones concerned with such education as future surgeons. Second, the use of standardized and task-relevant medical materials can enhance study comparability. Third, including a broader range of knots, both sliding and non-sliding, would better represent the diversity of surgical skills. Fourth, integrating 3D holograms via the Dynamics 365 Guide software could provide clearer and more interactive demonstrations and guidance in the learning knot-tying. Fifth, objective measures of knot quality, such as tension, could be incorporated. Sixth, a multimodal approach combining subjective and objective data (behavioral and physiological) could provide deeper insights into attention and mental effort, and potentially respond to the critics of the self-reported scales while assessing the cognitive load. Finally, examining the relationship between the cognitive load and self-regulated learning would clarify how the learners adjust their pace, repeat steps, or use metacognitive strategies to manage tasks. It seems that exploring the connection between the cognitive load theory and motivation is of great importance.

5. Conclusions

This study contributes valuable insights into the use of MR for knot-tying simulations. While the anticipated benefits of the integrated formats on the cognitive load were not fully realized concerning the ECL and the GCL, the significant findings related to the intrinsic cognitive load underscore the complexity of participant engagement and the learning peculiarity of procedures. To effectively observe the split-attention effect, future research should address high element interactivity in XR representations. Moreover, the integration of a MR in medical training should consider the cost and development of that technology, and balance it with some traditional methods. Ultimately, further exploration of the interplay between the cognitive load and self-regulated learning is necessary.

Author Contributions: Conceptualization, G.M. and A.A.; methodology, G.M. and A.A.; software, G.M. and A.A.; validation, G.M. and A.A.; formal analysis, G.M. and A.A.; investigation, G.M. and A.A.; resources, G.M. and A.A.; data curation, G.M. and A.A.; writing—original draft preparation, G.M. and A.A.; writing—review and editing, G.M. and A.A.; visualization, G.M. and A.A.; supervision, A.A. and G.M.; project administration, G.M. and A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was approved by the Ethics Committee for Research at the University of Toulouse, approval date 31 March 2023, under the reference number 2023_651.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Additional data will be provided upon request.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Descriptive statistics of the variables relating to the CL perceived according to the conditions of presentation.

		95% CI						
	Conditions	Mean	Lower	Upper	Median	SD	Minimum	Maximum
Complexity	C	4.15	2.83	5.48	5	2.19	1	7
	NC	5.77	4.33	7.21	6	2.39	1	10
Mental Effort-Complexity	C	5.54	4.39	6.69	6	1.90	2	8
	NC	6.77	5.21	8.33	8	2.59	2	10
Clarity	C	2.31	1.11	3.50	2	1.97	1	8
	NC	3.08	1.93	4.22	3	1.89	1	7
Efficiency	C	2.92	1.56	4.28	2	2.25	1	8
	NC	3.69	2.27	5.12	3	2.36	1	8
Mental Effort-Clarity & Efficiency	C	3.15	1.62	4.69	3	2.54	1	10
	NC	4.77	3.31	6.23	5	2.42	1	9
Knowledge & understanding	C	7.85	6.93	8.76	8	1.52	6	10
	NC	6.08	4.38	7.78	6	2.81	1	10
Mental Effort-Knowledge & Understanding	C	6.15	5.00	7.31	7	1.91	3	9
	NC	7.31	6.04	8.57	8	2.10	4	10

Appendix B

Table A2. Descriptive statistics of the variable relating to the learning time per turn, according to the conditions.

	Conditions	N	Mean	Median	SD	Min	Max
T1 SMC	C	13	5.85	4	4.413	2	18
	NC	13	8.92	8	4.591	3	16
T2 SMC	C	12	2.33	2.00	1.155	1	5
	NC	8	5.13	4.00	3.796	2	14
T3 SMC	C	4	1.50	1.50	0.577	1	2
	NC	4	3.00	2.00	2.828	1	7
T1 Double Twist	C	13	12.23	10	9.057	3	32
	NC	13	17.08	17	8.108	7	30
T2 Double Twist	C	5	4.40	4	1.140	3	6
	NC	4	5.25	4.00	3.304	3	10
T3 Double Twist	C	0	NaN	NaN	NaN	NaN	NaN
	NC	0	NaN	NaN	NaN	NaN	NaN
T1 Duncan	C	13	2.85	3	1.573	1	6
	NC	13	3.92	4	1.656	2	8
T2 Duncan	C	9	1.44	1	0.527	1	2
	NC	7	1.43	1	0.535	1	2
T3 Duncan	C	2	1.50	1.50	0.707	1	2
	NC	1	3.00	3	NaN	3	3

Appendix C

Table A3. Descriptive statistics of user satisfaction.

	Conditions	N	Mean	Median	SD	Min	Max
GL1	C	13	6.92	7	2.47	1	10
	NC	13	7.23	8	2.98	2	10
GL2	C	13	7.77	8	2.62	2	10
	NC	13	7.77	8	2.42	3	10
GL3	C	13	7.69	8	2.53	3	10
	NC	13	7.77	8	1.96	3	10
GL4	C	13	6.69	7	2.66	1	10
	NC	13	7.69	8	2.32	2	10
GL5	C	13	9.08	9	1.04	7	10
	NC	13	8.54	9	1.81	5	10
GL6	C	13	7.54	8	1.98	2	10
	NC	13	7.85	8	2.08	4	10
GL7	C	13	7.46	8	2.18	3	10
	NC	13	8.15	8	1.57	6	10
GL8	C	13	7.77	8	2.39	3	10
	NC	13	8.08	8	1.38	5	10
GL9	C	13	6.00	6	1.83	3	9
	NC	13	6.46	7	2.63	1	10
GL10	C	13	8.54	9	1.39	7	10
	NC	13	7.69	8	2.32	1	10
GL11	C	13	6.85	8	3.08	1	10
	NC	13	7.46	8	2.54	3	10

Appendix D

We analyzed the internal structure of our 7-item instrument for the self-reported cognitive load scale. Basic assumptions were checked: $KMO = 0.528$, Bartlett's $\chi^2 (21) = 64.8$, $p < 0.001$. Given the small sample size ($N = 26$), a principal component analysis was conducted with an oblique rotation, specifically using “oblimin,” as in previous studies (Leppink et al., 2013; Thees et al., 2020).

Table A4. KMO, reliability of adapted self-reported CLS.

CL Type/Item	KMO	Reliability α
ICL		0.826
Item 1	0.509	
Item 2	0.560	
ECL	0.393	0.735
Item 3	0.393	
Item 4	0.749	
Item 5	0.477	
GCL		0.067
Item 6	0.550	
Item 7	0.698	

Appendix E. Bootstrap Process

The Jamovi software was used to conduct the data analysis, utilizing the *t*-test or Mann–Whitney test when the assumptions were not met. When the confidence intervals from these tests were unclear, the Rj Editor module was employed, which allows for the integration and execution of R code directly within Jamovi version 2.3.28.0. This enabled the estimation of confidence intervals for various variables of interest. Different codes used are available in the supplementary material provided.

Appendix F. Adaptation and Translation of the CLS

For our adapted CLS, the corresponding items retained are: ICL (items 2: Complexity, 4: Mental Effort-Complexity), ECL (items 5: Clarity, 7: Efficiency, 8: Mental Effort-Clarity & Efficiency), and GCL (items 12: Knowledge & Understanding, 13: Mental Effort-Knowledge & Understanding).

Table A5. Original, adapted, and translated items of the CLS.

#	Original Item (English)	# New	Adapted Item (English)	Adapted Selected Item (French)
1	The content of this activity was very complex.	1	-	-
2	The problem/s covered in this activity was/were very complex.	2	The procedures covered in these activities were very complex.	Les procédures abordées dans ces activités étaient très complexes.
3	In this activity, very complex terms were mentioned.	3	-	-
4	I invested a very high mental effort in the complexity of this activity.	4	I invested a very high mental effort in the complexity of these activities.	J'ai investi un effort mental très important dans la complexité de ces activités.
5	The explanations and instructions in this activity were very unclear.	5	The instructions for these activities were very unclear.	Les instructions de ces activités n'étaient pas très claires
6	The explanations and instructions in this activity were full of unclear language.	6	-	-
7	The explanations and instructions in this activity were, in terms of learning, very ineffective.	7	The instructions for these activities were, in terms of learning very ineffective.	Les instructions de ces activités étaient, en termes d'apprentissage, très inefficaces.
8	I invested a very high mental effort in unclear and ineffective explanations and instructions in this activity.	8	I invested a very high mental effort in unclear and ineffective instructions in these activities.	J'ai investi un effort mental très important dans les instructions peu claires et inefficaces de ces activités.
9	This activity really enhanced my understanding of the content that was covered.	9	-	-
10	This activity really enhanced my understanding of the problem/s that was/ were covered.	10	-	-
11	This activity really enhanced my knowledge of the terms that were mentioned.	11	-	-
12	This activity really enhanced my knowledge and understanding of how to deal with the problem/s covered.	12	These activities really enhanced my knowledge and understanding of how to realize the procedures covered.	Ces activités ont vraiment amélioré ma connaissance et ma compréhension de la manière de réaliser les procédures abordées.
13	I invested a very high mental effort during this activity in enhancing my knowledge and understanding.	13	I invested a very high mental effort during these activities in enhancing my knowledge and understanding of the procedures.	J'ai investi un effort mental très important au cours de ces activités pour améliorer ma connaissance et ma compréhension des procédures.

Appendix G. Translation of SGUS Questionnaire

Table A6. Translation of SGUS Questionnaire.

#	Original Item (English)	Translated Item (French)
1	GL1 With AR-glasses I could access information at the most appropriate place and moment.	GL1 Avec les lunettes AR, j'ai pu accéder à l'information à l'endroit et au moment les plus appropriés.
2	GL2 Content displayed on the AR-glasses made sense in the context I used it.	GL2 Le contenu affiché sur les lunettes AR avait un sens dans le contexte où je l'ai utilisé.
3	GL3 AR-glasses provided me with the most suitable amount of information.	GL3 Les lunettes AR m'ont fourni la quantité d'informations la plus appropriée.
4	GL4 AR-glasses allowed a natural way to interact with information displayed.	GL4 Les lunettes AR permettent d'interagir de manière naturelle avec les informations affichées.
5	GL5 I had a good conception of what is real and what is augmented when using AR-glasses.	GL5 J'ai une bonne conception de ce qui est réel et de ce qui est augmenté lorsque j'utilise des lunettes AR.
6	GL6 The interaction with content on AR-glasses captivated my attention in a positive way.	GL 6 L'interaction avec le contenu des lunettes AR a captivé mon attention de manière positive.
7	GL7 The instructions given by AR-glasses helped me to accomplish the task.	GL7 Les instructions données par les lunettes AR m'ont aidé à accomplir la tâche.
8	GL8 I understood what is expected from me in each phase of the task with the help of AR-glasses.	GL8 J'ai compris ce que l'on attendait de moi à chaque phase de la tâche avec l'aide des lunettes AR.
9	GL9 Performing the task with the help of AR-glasses was natural to me.	GL9 L'exécution de la tâche avec l'aide des lunettes AR était naturelle pour moi
10	GL10 While using AR-glasses, I was aware of the phase of the task at all times during the execution of the task.	GL10 En utilisant les lunettes AR, j'étais conscient de la phase de la tâche à tout moment pendant l'exécution de la tâche.
11	GL11 While using AR-glasses, I was able to pay attention to the essential aspects of the task all the time.	GL11 En utilisant les lunettes AR, j'ai pu prêter attention aux aspects essentiels de la tâche à tout moment.

Appendix H. Smart Glass User Satisfaction

We analyzed the internal structure of our 11-item Smart Glass User Satisfaction (SGUS) scale. Basic assumptions were checked: $KMO = 0.661$, Bartlett's $\chi^2 (55) = 167.019$, $p < 0.001$. Given the sample size ($N = 26$), a principal component analysis was conducted with an oblique rotation, specifically using "oblimin", in line with previous studies. The results demonstrate an adequate internal structure for the majority of the items, indicating that the SGUS scale is a promising instrument for measuring user satisfaction with smart glasses.

Table A7. KMO Measure of Sampling Adequacy per Item of SGUS questionnaire.

Items	MSA
Overall	0.661
GL1	0.760
GL2	0.735
GL3	0.874
GL4	0.519
GL5	0.193
GL6	0.678
GL7	0.667
GL8	0.703
GL9	0.581
GL10	0.412
GL11	0.726

Appendix I. Counterbalancing Method for Knot Recall

Counterbalancing Procedure for 26 participants : An example of a list of randomly generated (by a small program written in Java) order for knot recall

Participant 1 task order:

Task 1: SMC Knot

Task 2: Duncan Knot

Task 3: Double Twist Knot

Participant 2 task order:

Task 1: Double Twist Knot

Task 2: SMC Knot

Task 3: Duncan Knot

Participant 3 task order:

Task 1: Double Twist Knot

Task 2: SMC Knot

Task 3: Duncan Knot

Participant 4 task order:

Task 1: Duncan Knot

Task 2: Double Twist Knot

Task 3: SMC Knot

Participant 5 task order:

Task 1: SMC Knot

Task 2: Duncan Knot

Task 3: Double Twist Knot

Participant 6 task order:

Task 1: SMC Knot

Task 2: Duncan Knot

Task 3: Double Twist Knot

Participant 7 task order:

Task 1: Double Twist Knot

Task 2: SMC Knot

Task 3: Duncan Knot

Participant 8 task order:

Task 1: SMC Knot

Task 2: Double Twist Knot

Task 3: Duncan Knot

Participant 9 task order:

Task 1: Double Twist Knot

Task 2: SMC Knot

Task 3: Duncan Knot

Participant 10 task order:

Task 1: Double Twist Knot

Task 2: Duncan Knot

Task 3: SMC Knot

Participant 11 task order:

Task 1: Duncan Knot

Task 2: SMC Knot

Task 3: Double Twist Knot

Participant 12 task order:

Task 1: SMC Knot

Task 2: Double Twist Knot

Task 3: Duncan Knot

Participant 13 task order:

Task 1: Double Twist Knot

Task 2: Duncan Knot

Task 3: SMC Knot

Participant 14 task order:

Task 1: Duncan Knot

Task 2: SMC Knot

Task 3: Double Twist Knot

Participant 15 task order:

Task 1: SMC Knot

Task 2: Double Twist Knot

Task 3: Duncan Knot

Participant 16 task order:

Task 1: Double Twist Knot

Task 2: Duncan Knot

Task 3: SMC Knot

Participant 17 task order:

Task 1: Double Twist Knot

Task 2: Duncan Knot

Task 3: SMC Knot

Participant 18 task order:

Task 1: SMC Knot

Task 2: Double Twist Knot

Task 3: Duncan Knot

Participant 19 task order:

Task 1: Double Twist Knot

Task 2: Duncan Knot

Task 3: SMC Knot

Participant 20 task order:

Task 1: Duncan Knot

Task 2: Double Twist Knot

Task 3: SMC Knot

Participant 21 task order:

Task 1: Double Twist Knot

Task 2: Duncan Knot

Task 3: SMC Knot

Participant 22 task order:

Task 1: SMC Knot

Task 2: Duncan Knot

Task 3: Double Twist Knot

Participant 23 task order:

Task 1: SMC Knot

Task 2: Duncan Knot

Task 3: Double Twist Knot

Participant 24 task order:

Task 1: Duncan Knot

Task 2: SMC Knot

Task 3: Double Twist Knot

Participant 25 task order:

Task 1: Double Twist Knot

Task 2: SMC Knot

Task 3: Duncan Knot

Participant 26 task order:

Task 1: Duncan Knot

Task 2: SMC Knot

Task 3: Double Twist Knot

References

- Allcoat, D., Hatchard, T., Azmat, F., Stansfield, K., Watson, D., & von Mühlennen, A. (2021). Education in the digital age: Learning experience in virtual and mixed realities. *Journal of Educational Computing Research*, 59(5), 795–816. [CrossRef]
- Altmeyer, K., Kapp, S., Thees, M., Malone, S., Kuhn, J., & Brünken, R. (2020). The use of augmented reality to foster conceptual knowledge acquisition in STEM laboratory courses—Theoretical background and empirical results. *British Journal of Educational Technology*, 51(3), 611–628. [CrossRef]
- Altmeyer, K., Malone, S., Kapp, S., Barz, M., Lauer, L., Thees, M., Kuhn, J., Peschel, M., Sonntag, D., & Brünken, R. (2021, September 20–22). *The effect of augmented reality on global coherence formation processes during STEM laboratory work in elementary school children*. [Conference session] (pp. 20–22), International Cognitive Load Theory Conference, Kingston, ON, Canada. Available online: <https://www.dfki.de/en/web/research/projects-and-publications/publication/11870> (accessed on 6 December 2024).
- Asoodar, M., Janesarvatan, F., Yu, H., & de Jong, N. (2024). Theoretical foundations and implications of augmented reality, virtual reality, and mixed reality for immersive learning in health professions education. *Advances in Simulation*, 9(1), 36. [CrossRef]
- Ayres, P., & Sweller, J. (2021). The split-attention principle in multimedia learning. In L. Fiorella, & R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 199–211). Cambridge University Press. [CrossRef]
- Barsom, E. Z., Graafland, M., & Schijven, M. P. (2016). Systematic review on the effectiveness of augmented reality applications in medical training. *Surgical Endoscopy*, 30(10), 4174–4183. [CrossRef]
- Baumgarten, K. M., & Wright, R. W. (2007). Ease of tying arthroscopic knots. *Journal of Shoulder and Elbow Surgery*, 16(4), 438–442. [CrossRef]
- Bautista, L., Maradei, F., & Pedraza, G. (2023). Strategies to reduce visual attention changes while learning and training in extended reality environments. *International Journal on Interactive Design and Manufacturing*, 17(1), 17–43. [CrossRef]

- Buchner, J., Buntins, K., & Kerres, M. (2022). The impact of augmented reality on cognitive load and performance: A systematic review. *Journal of Computer Assisted Learning*, 38(1), 285–303. [CrossRef]
- Camp, G., Surma, T., & Kirschner, P. A. (2021). Foundations of multimedia learning. In L. Fiorella, & R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 17–24). Cambridge University Press. [CrossRef]
- Çeken, B., & Taşkın, N. (2022). Multimedia learning principles in different learning environments: A systematic review. *Smart Learning Environments*, 9(1), 19. [CrossRef]
- Chanquoy, L., Tricot, A., & Sweller, J. (2007). Chapitre 3. La théorie de la charge cognitive. *Collection U*, 131–188. Available online: <https://shs-cairn-info.gorgone.univ-toulouse.fr/la-charge-cognitive--9782200347246-page-131> (accessed on 23 November 2024).
- Chen, P. (2023, June 2). How immersive technology is transforming education, healthcare and beyond. *World Economic Forum*. Available online: <https://www.weforum.org/stories/2023/06/immersive-technology-transform-education-healthcare/> (accessed on 23 November 2024).
- Chiniara, G., Cole, G., Brisbin, K., Huffman, D., Cragg, B., Lamacchia, M., & Norman, D. (2013). Simulation in healthcare: A taxonomy and a conceptual framework for instructional design and media selection. *Medical Teacher*, 35(8), e1380–e1395. [CrossRef]
- Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior*, 25(2), 315–324. [CrossRef]
- Co, M., Chiu, S., & Billy Cheung, H. H. (2023). Extended reality in surgical education: A systematic review. *Surgery*, 174(5), 1175–1183. [CrossRef]
- de Bruin, A. B. H., Roelle, J., Carpenter, S. K., Baars, M., & EFG-MRE. (2020). Synthesizing cognitive load and self-regulation theory: A theoretical framework and research agenda. *Educational Psychology Review*, 32(4), 903–915. [CrossRef]
- de Sá, V. H. L. C., Pazin, G. S., Elias, P. E., Achar, E., & Pereira Filho, G. V. (2022). How to do it: Teaching surgical skills to medical undergraduates. *Annals of Medicine and Surgery*, 82, 104617. [CrossRef]
- Feldon, D. F., Callan, G., Juth, S., & Jeong, S. (2019). Cognitive load as motivational cost. *Educational Psychology Review*, 31(2), 319–337. [CrossRef]
- Fiorella, L., & Mayer, R. E. (2021). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In L. Fiorella, & R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 185–198). Cambridge University Press. [CrossRef]
- Forgione, A., & Guraya, S. Y. (2017). The cutting-edge training modalities and educational platforms for accredited surgical training: A systematic review. *Journal of Research in Medical Sciences: The Official Journal of Isfahan University of Medical Sciences*, 22(1), 51. [CrossRef] [PubMed]
- Fraser, K. L., Ayres, P., & Sweller, J. (2015). Cognitive load theory for the design of medical simulations. *Simulation in Healthcare*, 10(5), 295. [CrossRef]
- Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, 16(6), 511–525. [CrossRef]
- Gonnermann-Müller, J., & Krüger, J. M. (2024). Unlocking augmented reality learning design based on evidence from empirical cognitive load studies—A systematic literature review. *Journal of Computer Assisted Learning*, 41(1), e13095. [CrossRef]
- Grantcharov, T. P. (2008). Is virtual reality simulation an effective training method in surgery? *Nature Clinical Practice Gastroenterology & Hepatology*, 5(5), 232–233. [CrossRef]
- Gregory, T. M., Gregory, J., Sledge, J., Allard, R., & Mir, O. (2018). Surgery guided by mixed reality: Presentation of a proof of concept. *Acta Orthopaedica*, 89(5), 480–483. [CrossRef]
- Helin, K., Kuula, T., Vizzi, C., Karjalainen, J., & Vovk, A. (2018). User experience of augmented reality system for astronaut’s manual work support. *Frontiers in Robotics and AI*, 5, 106. [CrossRef] [PubMed]
- Issa, N., Schuller, M., Santacaterina, S., Shapiro, M., Wang, E., Mayer, R. E., & DaRosa, D. A. (2011). Applying multimedia design principles enhances learning in medical education. *Medical Education*, 45(8), 818–826. [CrossRef]
- Kim, S. H., & Ha, K. I. (2000). The SMC knot—A new slip knot with locking mechanism. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, 16(5), 563–565. [CrossRef]
- Krüger, J. M., & Bodemer, D. (2022). Application and investigation of multimedia design principles in augmented reality learning environments. *Information*, 13(2), 74. [CrossRef]
- Lacroix, P.-M., Commeil, P., Chauveaux, D., & Fabre, T. (2021). Learning and optimizing arthroscopic knot-tying by surgery residents using procedural simulation. *Orthopaedics & Traumatology: Surgery & Research*, 107(8), 102944. [CrossRef]
- Laumann, D., Schlummer, P., Abazi, A., Borkamp, R., Lauströer, J., Pernice, W., Schuck, C., Schulz-Schaeffer, R., & Heusler, S. (2024). Analyzing the effective use of augmented reality glasses in university physics laboratory courses for the example topic of optical polarization. *Journal of Science Education and Technology*, 33(5), 668–685. [CrossRef]
- Leahy, W., & Sweller, J. (2019). The centrality of element interactivity to cognitive load theory. In *Advances in cognitive load theory* (pp. 221–232). Routledge.

- Leppink, J., Paas, F., Gog, T., Van der Vleuten, C., & Van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction*, 30, 32–42. [CrossRef]
- Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, 45(4), 1058–1072. [CrossRef] [PubMed]
- Liu, H., Wu, J., Tang, Y., Li, H., Wang, W., Li, C., & Zhou, Y. (2019). Percutaneous placement of lumbar pedicle screws via intraoperative CT image-based augmented reality-guided technology. *Journal of Neurosurgery: Spine*, 32(4), 542–547. [CrossRef]
- Lu, S., Sanchez Perdomo, Y. P., Jiang, X., & Zheng, B. (2020). Integrating eye-tracking to augmented reality system for surgical training. *Journal of Medical Systems*, 44(11), 192. [CrossRef]
- Magalhães, R., Oliveira, A., Terroso, D., Vilaça, A., Veloso, R., Marques, A., Pereira, J., & Coelho, L. (2024). Mixed reality in the operating room: A systematic review. *Journal of Medical Systems*, 48(1), 76. [CrossRef] [PubMed]
- Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review*, 33(3), 937–958. [CrossRef]
- Masson, J.-B. (2023). Qu’attendre de la réalité virtuelle et augmentée pour les applications médicales. *Annales des Mines-Enjeux Numériques*, 22(2), 42–50. [CrossRef]
- Mayer, R. E. (2002). Multimedia learning. In *Psychology of learning and motivation* (Volume 41, pp. 85–139). Academic Press. [CrossRef]
- Mayer, R. E. (2020). Designing multimedia instruction in anatomy: An evidence-based approach. *Clinical Anatomy*, 33(1), 2–11. [CrossRef]
- Mayer, R. E. (2021). Cognitive theory of multimedia learning. In L. Fiorella, & R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 57–72). Cambridge University Press. [CrossRef]
- Mayer, R. E. (2024). The past, present, and future of the cognitive theory of multimedia learning. *Educational Psychology Review*, 36(1), 8. [CrossRef]
- Mayer, R. E., & Fiorella, L. (2021). Introduction to multimedia learning. In L. Fiorella, & R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 3–16). Cambridge University Press. [CrossRef]
- Mayo, W. J. (1927). Medical education for the general practitioner. *Journal of the American Medical Association*, 88(18), 1377–1379. [CrossRef]
- McInnis, C., Asif, H., Ajzenberg, H., Wang, P., Mosa, A., Dang, F., Savage, T., Vo, T. X., Wang, J., Zevin, B., Mann, S., & Winthrop, A. (2021). The next surgical skills and technology elective program: The “surgical skills and technology elective program” decreases cognitive load during suturing tasks in second year medical students. *Journal of Surgical Research*, 267, 598–604. [CrossRef] [PubMed]
- Mergen, M., Graf, N., & Meyerheim, M. (2024). Reviewing the current state of virtual reality integration in medical education—A scoping review. *BMC Medical Education*, 24(1), 788. [CrossRef]
- Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, 77(12), 1321–1329.
- Nagayo, Y., Saito, T., & Oyama, H. (2021). A novel suture training system for open surgery replicating procedures performed by experts using augmented reality. *Journal of Medical Systems*, 45(5), 60. [CrossRef]
- Nagayo, Y., Saito, T., & Oyama, H. (2022). Augmented reality self-training system for suturing in open surgery: A randomized controlled trial. *International Journal of Surgery*, 102, 106650. [CrossRef]
- Nottage, W. M., & Lieurance, R. K. (1999). Arthroscopic knot typing techniques. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, 15(5), 515–521. [CrossRef]
- Ouwehand, K., van der Kroef, A., Wong, J., & Paas, F. (2021). Measuring cognitive load: Are there more valid alternatives to likert rating scales? *Frontiers in Education*, 6, 702616. Available online: <https://www.frontiersin.org/articles/10.3389/feduc.2021.702616> (accessed on 11 November 2024). [CrossRef]
- Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84(4), 429. [CrossRef]
- Paas, F., & Sweller, J. (2021). Implications of cognitive load theory for multimedia learning. In L. Fiorella, & R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 73–81). Cambridge University Press. [CrossRef]
- Palumbo, A. (2022). Microsoft HoloLens 2 in medical and healthcare context: State of the art and future prospects. *Sensors*, 22(20), 7709. [CrossRef] [PubMed]
- Park, S., Bokijonov, S., & Choi, Y. (2021). Review of microsoft hololens applications over the past five years. *Applied Sciences*, 11(16), 7259. [CrossRef]
- Parong, J. (2021). Multimedia learning in virtual and mixed reality. In L. Fiorella, & R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 498–509). Cambridge University Press. [CrossRef]
- Pimentel, D., Fauville, G., Frazier, K., McGivney, E., Rosas, S., & Woolsey, E. (2022). *An introduction to learning in the metaverse*. Meridian Treehouse.

- Pociask, F. D., & Morrison, G. R. (2008). Controlling split attention and redundancy in physical therapy instruction. *Educational Technology Research and Development*, 56(4), 379–399. [CrossRef]
- Poupard, M., Larrue, F., Sauzéon, H., & Tricot, A. (2024). A systematic review of immersive technologies for education: Learning performance, cognitive load and intrinsic motivation. *British Journal of Educational Technology*, 56(1), 5–41. [CrossRef]
- Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, 147, 103778. [CrossRef]
- Rivière, E., Saucier, D., Lafleur, A., Lacasse, M., & Chiniara, G. (2018). Twelve tips for efficient procedural simulation. *Medical Teacher*, 40(7), 743–751. [CrossRef]
- Rolla, P. R., & Surace, M. F. (2002). The double-twist knot: A new arthroscopic sliding knot. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, 18(7), 815–820. [CrossRef]
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. [CrossRef] [PubMed]
- Sánchez-Margallo, J. A., Plaza de Miguel, C., Fernández Anzules, R. A., & Sánchez-Margallo, F. M. (2021). Application of mixed reality in medical training and surgical planning focused on minimally invasive surgery. *Frontiers in Virtual Reality*, 2, 692641. [CrossRef]
- Sandars, J., Patel, R. S., Goh, P. S., Kokatailo, P. K., & Lafferty, N. (2015). The importance of educational theories for facilitating learning when using technology in medical education. *Medical Teacher*, 37(11), 1039–1042. [CrossRef]
- Schroeder, N. L., & Cenkci, A. T. (2018). Spatial contiguity and spatial split-attention effects in multimedia learning environments: A meta-analysis. *Educational Psychology Review*, 30(3), 679–701. [CrossRef]
- Schroeder, N. L., & Cenkci, A. T. (2020). Do measures of cognitive load explain the spatial split-attention principle in multimedia learning environments? A systematic review. *Journal of Educational Psychology*, 112, 254–270. [CrossRef]
- Seufert, T. (2018). The interplay between self-regulation in learning and cognitive load. *Educational Research Review*, 24, 116–129. [CrossRef]
- Silvero Isidre, A., Friederichs, H., Mütther, M., Gallus, M., Stummer, W., & Holling, M. (2023). Mixed reality as a teaching tool for medical students in neurosurgery. *Medicina*, 59(10), 1720. [CrossRef]
- Strzys, M. P., Kapp, S., Thees, M., Klein, P., Lukowicz, P., Knierim, P., Schmidt, A., & Kuhn, J. (2018). Physics holo.lab learning experience: Using smartglasses for augmented reality labwork to foster the concepts of heat conduction. *European Journal of Physics*, 39(3), 035703. [CrossRef]
- Sullivan, M. E. (2020). Applying the science of learning to the teaching and learning of surgical skills: The basics of surgical education. *Journal of Surgical Oncology*, 122(1), 5–10. [CrossRef]
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312. [CrossRef]
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. [CrossRef]
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296. [CrossRef]
- ten Berge, T., & van Hezewijk, R. (1999). Procedural and declarative knowledge: An evolutionary perspective. *Theory & Psychology*, 9(5), 605–624. [CrossRef]
- Thees, M., Altmeyer, K., Kapp, S., Rexigel, E., Beil, F., Klein, P., Malone, S., Brünken, R., & Kuhn, J. (2022). Augmented reality for presenting real-time data during students' laboratory work: Comparing a head-mounted display with a separate display. *Frontiers in Psychology*, 13, 804742. [CrossRef]
- Thees, M., Kapp, S., Strzys, M. P., Beil, F., Lukowicz, P., & Kuhn, J. (2020). Effects of augmented reality on learning and cognitive load in university physics laboratory courses. *Computers in Human Behavior*, 108, 106316. [CrossRef]
- Tokuno, J., Carver, T. E., & Fried, G. M. (2023). Measurement and management of cognitive load in surgical education: A narrative review. *Journal of Surgical Education*, 80(2), 208–215. [CrossRef] [PubMed]
- Tricot, A. (1998). Charge cognitive et apprentissage. Une présentation des travaux de John Sweller. *Revue de Psychologie de L'Education*, 3, 37–64.
- White, J. 2019 February 24. *Microsoft at MWC barcelona: Introducing microsoft HoloLens 2*. The Official Microsoft Blog. Available online: <https://blogs.microsoft.com/blog/2019/02/24/microsoft-at-mwc-barcelona-introducing-microsoft-hololens-2/> (accessed on 21 November 2024).
- Wit, L. D., Kessels, R. P. C., Kurasz, A. M., Sr, P. A., O'Shea, D., Marsiske, M., Chandler, M. J., Piai, V., Lambertus, T., & Smith, G. E. (2023). Declarative learning, priming, and procedural learning performances comparing individuals with amnesic mild cognitive impairment, and cognitively unimpaired older adults. *Journal of the International Neuropsychological Society*, 29(2), 113–125. [CrossRef]
- Wu, B., Yu, X., & Gu, X. (2020). Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *British Journal of Educational Technology*, 51(6), 1991–2005. [CrossRef]
- Xue, H., Sharma, P., & Wild, F. (2019). User satisfaction in augmented reality-based training using microsoft HoloLens. *Computers*, 8(1), 9. [CrossRef]

- Yoganathan, S., Finch, D. A., Parkin, E., & Pollard, J. (2018). 360° virtual reality video for the acquisition of knot tying skills: A randomised controlled trial. *International Journal of Surgery*, 54, 24–27. [CrossRef] [PubMed]
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64–70. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Cognitive Reappraisal: The Bridge between Cognitive Load and Emotion

Rebecca B. Brockbank * and David F. Feldon

Department of Instructional Technology & Learning Sciences, Utah State University, Logan, UT 84322, USA;
david.feldon@usu.edu

* Correspondence: rebecca.brockbank@usu.edu

Abstract: Within this integrative review, cognitive load theory (CLT) is asserted as a powerful framework for conceptualizing human cognitive processes within learning. The relationship between cognition and emotion is then examined and further integrated within the scope of CLT. Emotion regulation strategies are discussed and adaptive strategies are proposed as being of particular relevance to broadening the theoretical and practical impacts of CLT. Central to the argument of this review is the use of cognitive reappraisal as a potential mitigator of cognitive load. Cognitive reappraisal involves reframing or reassessing understandings or beliefs that underlie an emotional response, which may mitigate cognitive load imposed by maladaptive emotion. It is proposed that effectively integrating adaptive emotion regulation strategies such as cognitive reappraisal in our pursuit of more effective cognitive functioning will aid in the development of a more integrated model of cognition and emotion within CLT.

Keywords: cognitive load theory; emotion regulation; cognitive reappraisal; emotion; cognition

Citation: Brockbank, R.B.; Feldon, D.F. Cognitive Reappraisal: The Bridge between Cognitive Load and Emotion. *Educ. Sci.* **2024**, *14*, 870.
<https://doi.org/10.3390/educsci14080870>

Academic Editors: Fred Paas,
Andre Tricot, Kim Ouwehand and
Florence Lespiau

Received: 30 June 2024
Revised: 2 August 2024
Accepted: 7 August 2024
Published: 9 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Since its inception, cognitive load theory (CLT) has become a powerful framework for conceptualizing human cognitive processes while learning. CLT research has historically focused solely on the cognitive impacts of learning environment, effective instructional design practices, and content within the learning experience. To expand the scope of the theory and its implications for practice, recent studies have initiated exploration of interactions between cognitive load and psychosocial factors impacting well-being, including cognitive effects (e.g., uncertainty), physiological effects (e.g., stress responses), and affective effects (e.g., emotion) [1–3]. However, despite an expansive body of literature surrounding how emotion affects cognition [1,4,5] and neuropsychological evidence that they use the same neural architecture [6–9], there remains a dearth of literature within CLT examining the impact of affective states on cognitive load (but see [10]). Indeed, more than a decade ago, Moreno [11] articulated the need for CLT to consider motivation (i.e., processes shaping goal direction, intensity, and persistence of behavior) as an essential component of effective and efficient learning. This gap within CLT exists despite studies identifying direct effects of emotion on both working memory and mental effort, including identifying certain types of emotion tending to foster more active versus less active cognitive processing [4]. Hence, examining the dynamic changes that occur during learning and acknowledging learners' agency for regulation may aid in expanding the scope of CLT [12].

Synthesizing the findings in the broader literature, this review argues that affective states (i.e., emotion) can have direct effects on cognitive load and therefore indirect effects on learning outcomes. Following a discussion of the major tenets of CLT and the impacts of emotion on cognition generally, this article will synthesize empirical findings implicating the impacts of emotion on cognitive load phenomena with the broader theoretical framework.

2. Cognitive Load Theory: Working Memory Limitations Affect Learning

At its core, CLT orients around a set of assumptions involving different types of cognitive processing demands as they relate to learning outcomes (e.g., comprehension, problem solving, and schema construction and automation) [13,14]. During learning, information is temporarily stored and processed within working memory [15]. Working memory involves controlled processing [16] and can be defined as “the ability to sustain goal-relevant information processing in the presence of alternative goals or other distraction”; hence, working memory capacity “serves to sustain goal-relevant processing despite the presence of competing response tendencies or distractions” [17] (p. 1527). Additionally, working memory is limited in both capacity and duration, processing only small amounts (or chunks) of information and retaining them for approximately 20–30 s [18–21]. This results in working memory gating out irrelevant information, as well as abstracting and selectively filtering what eventually becomes encoded as schemas into long-term memory [9,13,19,22]. Constraints in capacity and duration appear to be only relevant when dealing with novel information, however, as working memory limitations virtually disappear when information from long-term memory is transferred into working memory during cognitive tasks [3]. Hence, prior knowledge (i.e., long-term memory schemas or templates) effectively increases how much information can be held and processed in working memory at any given time, facilitating the manipulation of complex cognitive representations while avoiding excessive cognitive load [19,23,24]. These schemas or templates of domain-specific information grow and expand as new knowledge is integrated into larger structures. Such recursive integration forms the bedrock for the development of expertise, hence facilitating the organization of (and accessibility of) vast amounts of complex information [19]. Thus, “it is memory more than intelligence that underpins expert performance” [19] (p. 38).

Because the primary purpose of instruction is to facilitate the accumulation of critical information in long-term memory (i.e., learning), instruction that overloads and exceeds the limits of working memory—particularly relating to novel information—will ultimately be ineffective [3,10,25]. Central to CLT, therefore, is the premise that “the design of instructional materials should be aligned with learners’ limited cognitive processing resources in a way that unnecessary cognitive load is prevented and effective higher-level cognitive processes are supported whenever possible” [13] (p. 43). Because of this, inadequate instructional methods or environmental distractions during learning impose unnecessary demands on the cognitive system and hamper learning [3].

Cognitive load can be best conceptualized as a multidimensional construct involving a range of cognitive processes, including the effort and load associated with a learning task [26,27]. Mental effort involves the cognitive resources allocated and utilized for learning. Working memory capacity is depleted by mental effort but recovers after rest [28]. Mental (or cognitive) load, on the other hand, represents the total burden placed on working memory during the learning task [26,29,30].

Invested mental effort has been associated with learners’ motivation for and persistence within tasks, as well as with self-efficacy (i.e., belief in one’s ability to successfully complete a task) [10,31–35]. In addition to investment of mental effort, self-efficacy is also associated with cognitive load [36–39]. Specifically, intrinsic load (i.e., proceeding demands inherent to content complexity and/or productive for learning) may directly enhance self-efficacy, while extraneous load (i.e., processing unproductive for learning) may decrease it [37]. Additionally, learners’ prior knowledge mitigates cognitive load, reducing the burden placed on working memory and impacting motivational beliefs [10].

Types of Cognitive Load

Within CLT, cognitive load is typically categorized as either intrinsic or extraneous [40,41]. Both types of cognitive load are cumulative, or additive, in that any distribution of the two which in combination exceed working memory capacity results in ineffective learning [42].

Intrinsic cognitive load (ICL) refers to the demand inherent within the learning task itself (i.e., load imposed by effective instruction) *relative to the prior knowledge of the learner*.

Learners with higher levels of prior knowledge possess well-developed schemas which enhance learners' ability to process the information efficiently. According to Sweller, Merriënboer, and Paas [3] (p. 264), "complexity or element interactivity depends on a combination of both the nature of the information and the knowledge of the person processing the information" where interactivity is defined as "the level of interconnectedness between the information elements that need to be processed in working memory simultaneously to make sense of the learning tasks or materials" [5] (p. 342). Thus, the more challenging the information in terms of its novelty for a learner and the extent to which interactions among knowledge elements (i.e., element interactivity) must be understood for comprehension, the greater the level of ICL [41]. Effective instructional procedures, therefore, aim to reduce unnecessary levels of element interactivity [3,43].

Whereas ICL reflects the inherent knowledge complexity within a learning task for a given learner, extraneous cognitive load (ECL) reflects information processing imposed by instruction or the environment that does not contribute to learning (e.g., poorly designed instruction, distracting learning environment, etc.) [41,44–46]. This involves not only how information is presented but also what is required of the learner via instructional procedures. Because of this, ECL (unlike ICL) can be changed simply by changing instructional procedures.

Recent work by Klepsch and Seufert [4,12,47] further delineates cognitive load into passive and active states. Specifically, cognitive load experienced can be due either to the learner actively investing cognitive resources (i.e., mental effort) or passively experienced load (i.e., mental load). In their words, "While mental load reflects the task-centered dimension, which is determined externally by the task, mental effort is human-centered, and thus determined internally and dependent on learner's decisions and characteristics" [4] (p. 2). Hence, actively investing cognitive resources within the learning process (i.e., mental effort) requires a more self-regulated approach to learning. This additional perspective provides valuable insight into the interaction between emotion regulation within cognitive processes as it relates to cognitive load experienced.

3. Emotions Affect Learning and Cognition

Emotions are multidimensional phenomena involving "sets of coordinated psychological [or subjective] processes, including affective, cognitive, physiological, motivation, and expressive components" [48] (p. 110) (see also [49,50]), including reactions to both external stimuli and internal mental representations [51]. Indeed, emotional information/material "constitutes much of the information we process daily" [52] (p. 1). Because emotions have the capacity of promoting social and moral behavior [53] and are central to achievement strivings [48,54], emotions can also be conceptualized as action tendencies that link actions and behaviors [55,56]. Thus, emotions inherently involve subjective (or psychological) feeling experiences and physiological expressions, as well as behavioral tendencies [53,57]. They are dynamic in nature, encompassing complex motivational and behavioral characteristics/aspects [55]. In general, emotions arise from a cognitive appraisal of the object, person, or situation associated with the emotion, particularly in terms of how it relates to one's well-being [53,58].

Two core dimensions constitute the nature and function of emotion: valence (e.g., pleasant vs. unpleasant) and arousal (e.g., intensity of emotion; activating vs. deactivating) [55,56,59]. Positive states (i.e., pleasant valence) include emotions such as joy, excitement, and enjoyment, whereas negative states (i.e., unpleasant valence) are associated with emotions such as frustration, anger, and sadness. Further, higher arousal (i.e., activating) emotions, such as anger, are delineated from lower arousal (i.e., deactivating) emotions, such as boredom [60]. Emotions vary from individual to individual [61] yet can be influenced through social interactions [62] and can even be considered contagious within collaborative learning environments, eliciting similar valence and arousal among groups of people [63].

Depending on the context or situation, emotions can be either helpful (i.e., adaptive) or harmful (i.e., maladaptive) [64]. Yet, not all negative emotions are maladaptive or unproductive, and not all positive emotions are adaptive [65]. For example, Plass and Kaplan [55] assert that “moderate and manageable levels of stress and [state-related] anxiety can be adaptive as part of the motivation process of certain individuals in some learning tasks” (p. 148). Within academic contexts, such feelings are generally associated with fear of failure within performance and/or evaluation situations [66–71]. Academic anxiety within these contexts could potentially trigger a ‘fight’ rather than ‘flight’ response to academic challenges and setback. If this were to be the case, academic anxiety resulting in ‘fight’ responses to challenges could foster more determined engagement and perseverance, which would be positively associated with academic resilience [72]. The critical factor seems to be the emotion regulation strategy employed, as maladaptive emotion regulation strategies are generally more strongly associated with psychopathology than adaptive strategies [73]. In particular, while effective emotion regulation tends to be more problematic among mood related disorders such as clinical anxiety or depression, implementation of adaptive emotion regulation strategies can play a central role in the management of psychopathology. Hence, employing adaptive regulation strategies such as cognitive reappraisal may help reorient or moderate maladaptive emotions such as state-related anxiety in situations of stress [64].

Furthermore, emotional states can further be distinguished as either integral or incidental, with both states being induced via a stimulus [53,74–76]. Integral emotional states are associated with an object of decision or judgment. Incidental emotional states, on the other hand, are not related to the object of decision or judgment (e.g., pre-existing moods and/or chronic emotional dispositions [53]. Regarding integral and incidental emotions, Blanchette and Richards [74] hypothesize that “incidental emotion may focus attention away from task-relevant information while integral emotion may focus attention towards task-relevant information” (p. 580). Further, incidental emotion appears to impair logicity, while integral emotion appears to facilitate logicity [53,74].

3.1. *Emotion and Working Memory*

While emotion in general has not been extensively explored in relation to its impact on working memory, research has shown that emotions consume limited working memory resources via focusing attention on the object of the emotion and influence the storage and retrieval of information [48].

According to Plass and Kalyuga [5], the valence of emotion often matters. For example, emotions may affect memory by either broadening or narrowing cognitive resources. Whereas negative emotions may limit learning by increasing load on working memory via their propensity to narrow and focus attention, positive emotions may reduce load on working memory by expanding and consolidating cognitive resources, facilitating more efficient recall and application of information from long-term memory to the learning task at hand [1,5,77]. Similarly, negative emotions have been associated with task-irrelevant thinking and decreased fluency in performance, whereas positive emotions have been shown to increase fluency and reduce task-irrelevant thinking [48,78–80].

Further, valence of emotion can influence retrieval and impact motivation. For example, positive emotions tend to facilitate the retrieval of positive self- and task-related information. Likewise, negative emotions tend to facilitate the retrieval of negative information (e.g., mood-congruent retrieval) [48,81]. According to Blanchette and Richards [74], interpretation, judgment, reasoning, and decision-making are important cognitive tools we use to form coherent representations of the world. Each of these processes are reliant upon working memory resources—they help us anticipate what is ahead and make choices about courses of action. Interpretation, judgment, reasoning, and decision-making are not solely cognitive, however. Emotion affects each of these processes.

3.2. Influence of Emotion Regulation on Cognition and Behavior

Emotion—positive or negative—has the capacity to bias cognition and thereby bias behavior [82,83]. When emotions are unhelpful, emotion regulation in the form of cognitive reappraisal can be particularly useful. Thus, though emotion can be transitory in nature, individuals have the ability to exert some measure of control on emotion, particularly when it directly relates to a goal. Indeed, emotion regulation is heavily implicated within virtually all goal-directed behavior [64].

Emotion regulation refers to the processes and stages by which the spontaneous flow of emotional responses to everyday stimuli is redirected or modified to achieve an outcome or goal [64,84–86]. As such, emotion regulation inherently involves cognitive change via processes that range in intentionality from deliberate and effortful to relatively automatic [60,86–91]. Indeed, Koole [86] positions emotion regulation as “one of the most far-ranging and influential processes at the interface of cognition and emotion” (p. 4). Though emotions are often portrayed as “irresistible forces that exert a sweeping influence on behavior”, research indicates “virtually every aspect of emotional processing” can be controlled, “including how emotion directs attention, the cognitive appraisals that shape emotional experience, and the physiological consequences of emotion” [86] (p. 4) (see also [92–94]). This is accomplished via effective emotion regulation, a capacity positively associated with many beneficial outcomes on mental and physical health, social interactions and relationships, and task and/or work performance [95,96].

According to Gross [64], emotion regulation involves a series of substeps, namely (1) perceiving the emotional response accurately (i.e., adaptive versus maladaptive), (2) valuing the emotion regulation goal, and (3) implementing the emotion regulation strategy (i.e., behavioral response or action). Each of these components must be present for successful emotion regulation to occur. Failure to regulate emotion indicates a failure within one of these three components and results in maladaptive default behavior patterns [97,98].

Within the realm of emotion regulation, six primary strategies have been identified: cognitive reappraisal, expressive suppression, acceptance, avoidance, rumination, and problem solving [73]. Of these, expressive suppression and cognitive reappraisal have been more robustly studied [64]. For the purposes of this review, focus will be given to these two emotion regulation strategies, with particular emphasis on cognitive reappraisal (see Table 1). Expressive suppression occurs when an emotional response is inhibited or blocked expressively or behaviorally [60,64]. While expressive suppression can be effective at blocking positive emotion, it has not been shown to be as effective in reducing negative or maladaptive emotions. In fact, suppression can negatively impact physical health, including higher physiological arousal, poorer overall well-being, and poorer health outcomes [60,99–102]. Further, expressive suppression can also lead to reduced emotional self-awareness, poorer memory recall, increased intrusive thoughts, and negative social consequences, including diminished interpersonal connection [100,103–108].

Table 1. Emotion Regulation (ER) Strategies of focus [64,73,74].

Type	Definition	Context
Cognitive Reappraisal	Reframing or reassessing understandings or beliefs that underly emotional responses to make them more adaptive.	-Generally considered adaptive. -Associated with positive overall well-being. -Most effective ER strategy leading to cognitive change. -Most robustly studied ER strategy from a neuroimaging perspective.
Expressive Suppression	Inhibiting or blocking emotional responses expressively or behaviorally.	-Generally considered maladaptive. -Associated with poorer health and overall well-being, anxiety, and depression. -Most frequently examined ER strategy, often via observation and self-report.

Because emotional responses are a product or consequence of thought (i.e., appraisals), cognitive reappraisal involves subjective evaluations, including reframing or reassessing understandings or beliefs that underlie an emotional response to make it more adaptive to

the situation, bring it into better alignment with goals, or both [64,86,109]. The underlying premise of cognitive reappraisal is that if individuals can change the way they think about situations, experiences, etc., they can change the way they feel (i.e., change the emotional meaning) [109]. Indeed, cognitive reappraisal is associated with positive psychological states and mental health [110,111]. Additionally, cognitive reappraisal draws upon working memory resources [17] and includes the reappraisal of perceived control and/or value within a task [60,64,84]. As articulated by Huang et al. [60], “when it comes to emotions, [individuals] can modulate their cognitive responses to get a desired emotional experience” (p. 4). Cognitive reappraisal is generally more effective with conceptually generated emotions [109]. Except for select circumstances (e.g., high emotion intensity, particularly negative emotion), cognitive reappraisal is considered the most effective and adaptive emotion regulation strategy leading to cognitive change [64,84,109].

Because emotion regulation strategies generally fall under a self-regulated approach, using such strategies inherently involves the simultaneous use of both cognitive and metacognitive strategies. As such, emotion monitoring and regulating efforts rely on working memory resources and have thus been shown to impose cognitive load during learning [12].

4. Emotion Impacts Cognitive Load

Plass and Kalyuga [5] identify four ways in which emotion may influence or relate to cognitive load (see Table 2). First, emotion can be conceptualized as imposing ECL, competing for limited working memory resources with sources of ICL via the processing of task-irrelevant information (i.e., the emotion). Second, emotion may influence cognitive load by affecting motivation, and therefore, mental effort investment (i.e., active cognitive load) [4]. Third, emotion may affect the cognitive processes of encoding, storage, and retrieval of information, including impacting working memory through the broadening or narrowing of cognitive resources. Fourth, emotion may affect ICL, specifically when emotion regulation is an integral part of learning objectives and outcomes. For instance, some occupational training requires individuals to process or regulate emotions while performing key tasks (e.g., a physician informing a patient of a cancer diagnosis or a billing specialist dealing with an angry customer).

Table 2. Ways Emotion Influences Cognitive Load [5].

Emotional Impacts on Cognitive Load	Description
Emotion as Extraneous Cognitive Load	Emotion competes for limited working memory resources, imposing extraneous cognitive load as task-irrelevant information.
Emotion Impacts Motivation	Emotion influences cognitive load by affecting motivation and resulting mental effort investment.
Emotion Affects Cognitive Processes	Emotion affects the cognitive processes of encoding, storage and retrieval of information, thus impacting working memory.
Emotion as Intrinsic Cognitive Load	Emotion affects intrinsic cognitive load when emotion regulation is an integral part of learning outcomes & objectives.

4.1. Cognitive Load via Emotional Task-Irrelevant Information

Emotion may influence or relate to cognitive load when it imposes an unnecessary burden via the cognitive processing of task-irrelevant information [5], especially maladaptive emotions related to scarcity and anxiety. Such emotions are likely to lead to rumination (i.e., a maladaptive emotion regulation strategy) [112,113], which sustains extraneous load over an extended or recurring way.

Internal thoughts and feelings based in scarcity (e.g., perception of not having enough time, money, resources, sleep, etc.) can disrupt and diminish individuals’ cognitive capacity to maintain attention, persist, resist distractions, modulate impulses, and self-regulate via consuming limited cognitive resources within working memory, thus contributing to

task-unrelated cognitive load (e.g., ECL) [114–116]. In addition to impacting cognitive functioning, scarcity can also impact individuals' sense of well-being and is often associated with individuals experiencing acute stress [114–116].

Anxiety often involves worry, rumination, task-irrelevant thoughts, nervousness, fear of failure, and/or poor performance. This type of anxiety is state-related and delineated from trait-related clinical anxiety and may overload working memory with ECL via decreasing attentional control and increasing counter-productive strategies and task-irrelevant thoughts and/or behavior aimed at dealing with fear of failure [1,67,117,118]. For example, higher levels of test anxiety have been shown to increase both processing and performance time while also decreasing accuracy [113,119,120]. Impaired performance and efficiency is often a result of cognitive worry (including task-irrelevant thoughts), which unnecessarily increases cognitive demand. Such cognitive interference depletes working memory resources by reducing working memory processing and storage capacity [113,117,119,120]. Hence, some forms of stress and anxiousness may decrease availability of working memory resources [121–124].

Further, much evidence points to the robust effect that anxiousness has on cognitive processes heavily dependent on working memory resources, including attention, interpretation, judgment, decision making, and reasoning [74]. When such cognitive processes are compromised by anxiety, it impacts the process through which individuals extract meaning from information [74]. Thus, the fostering of anxious thoughts potentially depletes available cognitive resources and consumes limited working memory capacity, imposing ECL [74,125]. This impacts individuals' ability to identify, select, and maintain access to relevant information, as well as resist distractions competing for limited cognitive resources during tasks [126,127]. In sum, because anxiety and worry are attentionally demanding, they can interfere with an individual's ability to sustain controlled attention, as well as inhibit irrelevant stimuli and responses, producing deficits in attentional control and information filtering efficiency [117,127–130]. Such impaired inhibitory control often results in irrelevant information consuming limited working memory resources, impairing task performance, and impacting fluid cognition (i.e., domain-general abilities) [127,131,132].

Brockbank and Feldon [133] further found that stress, overwhelm, and worry were positively associated with higher levels of ECL, independent of content complexity in learning contexts (i.e., ICL). While ECL was associated with motivational cost, ICL was not [133]. Further, common sources of ECL were differentially associated with negatively experienced emotional aspects of motivational cost (see also [10,37]).

How emotion impacts motivational processes is influenced by both valence and intensity within both positive (e.g., joyful, excited) and negative (e.g., sad, angry) emotions. Emotion valence and intensity can potentially contribute to task-irrelevant thoughts compared to more neutral emotions (e.g., contented, bored) [5]. As the proportion of irrelevant thoughts increases, motivation and performance suffer [132]. Similarly, stress can lead to both task-relevant and task-irrelevant thoughts. Consequently, task-oriented thinking that directs one's thoughts to the task at hand helps mitigate anxiousness, unproductive worry and rumination, and poor performance [119,120]. As an adaptive strategy, cognitive reappraisal may encourage task-oriented thinking. Hence, thoughts (particularly emotionally based thoughts) can play an important role in the cognition–emotion relationship through either enhancing or impairing cognitive performance [5,119,120,132].

4.2. *Emotion, Motivation, and Cognition*

Relatedly, two additional ways in which emotion may influence or relate to cognitive load is when it impacts motivation, mental effort, and cognitive processes (including via the broadening or narrowing of cognitive resources) [5]. Indeed, learning how cognition, emotion, and motivation interact in influencing human behavior, while less understood, is a valuable endeavor [134]. Furthermore, that emotions contribute to working memory performance is also of particular relevance to motivationally based realms such as goal-directed cognitive task performance [135].

In its traditional formulation, CLT focused little on emotional or motivational constructs and “could not explain the performance of self-regulated learners who were capable of expanding on their effective cognitive capacity” [5] (p. 343) (see also [11,32]). Indeed, Moreno [11] articulated the need for CLT to consider motivation (i.e., the complex processes shaping goal direction, intensity, and persistence of behavior) as an essential component of effective and efficient learning. Research has shown how motivation and engagement—both emotionally based processes—can be enhanced by reducing cognitive load (e.g., via cognitive load reducing instructional practices), thereby positively impacting self-efficacy, valuing, planning, persistence, and mastery orientation [136].

Pessoa [134] further asserted that emotion and motivation have crucial roles in either enhancing or impairing behavior, depending on how they interact with cognitive control functions, particularly in terms of competing for limited cognitive processing resources. Because these subcomponents of cognition are mutually interacting, such ‘executive competition’ results in ‘capacity sharing’ within cognitive information processing systems, including working memory [134]. Additionally, emotion may either increase or decrease motivation for an individual to expend cognitive effort, and in this way, the relationship between emotion and cognitive load may be mediated by motivation [5].

4.3. Emotion, Regulation, and Reappraisal

The final way in which emotion may influence or relate to cognitive load is via emotion regulation [5]. Russell [56] asserts emotional episodes are psychologically constructed, rather than being biologically or socially determined. Because of this, there is substantial intrapersonal and interpersonal variation in emotional experiences, including changes over time [5]. An emotional episode is therefore dynamic, “constructed anew each time to fit its specific circumstances” [56] (p. 151).

Such episodic construction supports the notion that individuals have a certain level of control over and choice in how they cognitively appraise, reappraise, and construct emotion schemas, independent of circumstance or context. Indeed, cognitive appraisals may support individuals’ sense of autonomy: “the individual has to learn how to adapt to situational demands while preserving individual autonomy—inevitably a process guided by appraisals” [79] (p. 241). And the dynamic cognition–emotion interactions inherent within such appraisals and self-regulation may “serve as motivating forces that guide human adaptation and learning” [55] (p. 149). It is essential, therefore, to “take into account interpersonal variability, namely, the variable emotional and motivational needs and demands” of individuals, including contextual and environmental factors [55] (p. 148). Consequently, self-regulation is an important means through which effective functioning, learning, and motivation may occur, and adaptive emotion regulation is integral to this process. Such regulation is heavily influenced by both long-term and working memory functions [19]. Cognitive regulation and behavior regulation represent two key forms of self-regulation [1,137–139]. Yet, cognitive regulation and behavior regulation, while sharing overlapping aspects, involve important distinctions. Behavior regulation is related to simple emotion regulation or suppression of emotional responses (e.g., expressive suppression) [73,140] and, as such, suppresses expressive behavior. In contrast, cognitive regulation (e.g., cognitive reappraisal related to cognitive control) [73,141] attends to or interprets emotion-eliciting situations in ways that influence emotional responding [1]. Such cognitive regulation (e.g., cognitive reappraisal) includes reappraising maladaptive emotions or perceived costs associated with tasks, particularly difficult ones, instead of allowing such emotions to influence behavior [137,138,142]. When negative or maladaptive emotions influence behavior, focus shifts from regulating cognition to regulating behavior [1,143]. Hence, in the face of everyday challenges, individuals have the ability to cognitively regulate via reappraisal strategies.

Furthermore, while positive emotions potentially enhance motivation and cognitive functioning, negative emotions such as frustration, boredom, anxiety, and hopelessness may reduce effective working memory functioning, resulting in poorer performance, re-

duced learning, and longer times needed to reach mastery levels [5,65,79,144,145]. Hence, promoting emotion regulation and cognitive reappraisal of maladaptive emotions can potentially enhance cognitive functioning [55]. Thus, when dealing with the challenges and difficulties common to everyday life, individuals who reappraise emotions may have lower levels of ECL than those who do not due to their ability to cognitively regulate rather than behaviorally regulate their responses to maladaptive emotions [1]. Such a reframing strategy can be utilized as many times as is needed. For example, from a cognitive perspective, stress can be conceptualized and viewed in terms of a call for action enhancing one's awareness of the need to do something about a given state [119,146].

5. Discussion

Of the four ways emotion impacts cognitive load discussed above [5], the self-regulation strategy of cognitive reappraisal may be the most crucial in terms of conceptualizing the direct relationship between emotion and cognitive load (i.e., the cognitive–affective relationship). Indeed, Plass and Kalyuga [5] encourage the need to conceptualize adaptive emotion regulation—particularly cognitive reappraisal—within the context of CLT because any influence that disturbs information processing flow interferes with working memory [9,19,135]. Related to such flow disturbances, Kellerman et al. [147] found higher levels of cognitive load were associated with decreased activation of emotion processing centers in the brain, potentially negatively impacting ability to effectively emotionally regulate. Because emotion has been posited as being one of the main interfering factors within working memory due to the preferential processing of emotionally loaded information within working memory [9,19,135,148], utilizing adaptive emotion regulation strategies such as cognitive reappraisal to mitigate the negative effects of maladaptive emotion is of particular importance to managing cognitive load.

However, because cognitive reappraisal is an emotion regulation strategy involving the metacognitive process of monitoring, it draws on additional cognitive resources, potentially impacting working memory capacity [12]. Allocating working memory resources to metacognitive processes may result in ICL if it complements the learning process but may also impose ECL when it hinders the learning task, particularly when complex tasks are undertaken by novice learners [12,149]. Further, because cognitive reappraisal is a form of self-control, when individuals experience additional cognitive demands conducive to cognitive load conditions, reappraisal may be less likely to be effective if such a strategy is ‘new’ to the cognitive system (i.e., not regularly integrated into cognitive processing) because these demands may compromise the ability to utilize cognitive resources [109]. Therefore, explicit guidance regarding cognitive reappraisal strategies (including instruction, prompts, cues, and feedback) is important for creating well-equipped learners able to effectively regulate without incurring unmanageable or unhelpful levels of cognitive load [12].

6. Implications

Practicing and integrating adaptive cognitive reappraisal strategies in a way that leads to automaticity (i.e., becomes habit and imposes no appreciable cognitive load) [150] is essential [52,96]. Overall, this suggests that cognitive reappraisal strategies could potentially become productive and worthwhile theoretically, as well as in practice within CLT in terms of adaptively dealing with emotion during instruction, as well as within other contexts and interactions.

Historically, CLT research has worked under the premise that emotions “may restrict the capacity of working memory by competing with task-relevant processes” thus increasing cognitive load, decreasing transfer, and adversely impacting the learning process [3] (p. 285). If this were to be the case, what logically follows is the notion that learning is most effective when emotion is minimized or potentially eliminated from the learning environment. However, humans are fundamentally emotional beings, and as such, human

learning cannot be devoid of emotion. Indeed, “emotions, stress, and uncertainty are often an integral part” of educational and professional tasks [3] (p. 285).

In some cases, it would be inherently unproductive to prevent emotion during learning [3]. For example, in many instances, vocational and professional education should be “carefully designed in such a way that learners develop professional competencies enabling them to perform professional tasks up to the standards, including the ability to deal with emotion, stress and uncertainty and to maintain overall wellbeing” [3] (p. 285). Therefore, broadening the theoretical scope of CLT to directly address emotion in cognition, including the study and integration of adaptive strategies with potential to mitigate cognitive load (particularly ECL) during learning, would indeed be a valuable endeavor.

Recent evidence toward this end indicates that targeted intervention may enhance the processing of emotionally based cognition [52,96]. Schweizer et al. [52] found that while working memory training led to gains on other working memory tasks, only working memory training with emotional material led to transferable gains with improved affective control (i.e., emotion regulation). This led to Schweizer et al. [96] asserting that “working memory capacity and the capacity to use working memory successfully in emotional contexts are partly separable cognitive abilities” (p. 5301) (see also [151,152]). Further, evidence from their study demonstrates significantly improved adaptive emotion regulation capacity as a result of targeted emotion-related working memory training. Thus, findings indicate enhancing emotion-related working memory capacity subsequently enhances emotion regulation ability due to working memory and emotion regulation sharing underlying neural circuitry, and hence “training in one context would benefit performance in the other by augmenting the efficiency of this common neural network” [96] (p. 5310). Given these promising findings, it is plausible that targeted instruction and training may not only enhance working memory and adaptive emotion regulation but potentially reduce overall cognitive load and benefit learning.

7. Conclusions

While CLT strategies have traditionally focused on modifying task complexity, instruction design, and/or learner knowledge, relatively little attention has been given within CLT research concerning how emotions and cognitive load influence one another [153]. Research indicates that “emotion modulates learning through numerous cognitive processes including motivation, attention, working memory, and long-term memory”, thus potentially significantly influencing cognitive load experience by learners [153] (p. S221). Indeed, gaining a more complete understanding of higher-level cognition will inherently involve a more in-depth examination of the integral relationship between cognition and emotion [74].

This review has provided strong evidence illustrating a dynamic and important interplay between emotion and cognition despite prior traditional distinctions categorizing them as separate and relatively unrelated functions [3,5,74]. Indeed, sparse evidence exists indicating emotion does not have an effect on higher level cognitive processes [74]. Thus, because emotion and cognition dynamically influence each other over time [65], the relationship between cognition and emotion may be better characterized in terms of a continuously interconnected yet dynamic feedback loop that is contextually influenced [55,154,155]. It is of great importance to the theoretical scope of CLT, therefore, to better understand the fundamental connectedness between emotion and cognition, including what might be included in an instructional context to facilitate adaptive emotion regulation strategies—particularly cognitive reappraisal—within a cognitive load framework.

Author Contributions: Conceptualization, R.B.B. and D.F.F.; investigation, R.B.B.; data curation, R.B.B.; writing—original draft preparation, R.B.B.; writing—review and editing, D.F.F.; supervision, D.F.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Hawthorne, B.S.; Vella-Brodrick, D.A.; Hattie, J. Well-being as a cognitive load reducing agent: A review of the literature. *Front. Educ.* **2019**, *4*, 121. [CrossRef]
- Kalyuga, S. Evolutionary perspective on human cognitive architecture in cognitive load theory: A dynamic, emerging principle approach. *Educ. Psychol. Rev.* **2023**, *35*, 91. [CrossRef]
- Sweller, J.; van Merriënboer, J.J.; Paas, F. Cognitive architecture and instructional design: 20 years later. *Educ. Psychol. Rev.* **2019**, *31*, 261–292. [CrossRef]
- Klepsch, M.; Seufert, T. Making an effort versus experiencing load. *Front. Educ.* **2021**, *6*, 645284. [CrossRef]
- Plass, J.L.; Kalyuga, S. Four ways of considering emotion in cognitive load theory. *Educ. Psychol. Rev.* **2019**, *31*, 339–359. [CrossRef]
- Gray, J. Emotional modulation of cognitive control: Approach-withdrawal states double-dissociate spatial from verbal two-back task performance. *J. Exp. Psychol. Gen.* **2001**, *130*, 436–452. [CrossRef] [PubMed]
- Gray, J.R. Integration of emotion and cognitive control. *Curr. Dir. Psychol. Sci.* **2004**, *13*, 46–48. [CrossRef]
- Gray, J.R.; Braver, T.S.; Raichle, M.E. Integration of emotion and cognition in the lateral prefrontal cortex. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 4115–4120. [CrossRef]
- Kensinger, E.A.; Corkin, S. Effect of negative emotional content on working memory and long-term memory. *Emotion* **2003**, *3*, 378–393. [CrossRef]
- Feldon, D.F.; Callan, G.; Juth, S.; Jeong, S. Cognitive load as motivational cost. *Educ. Psychol. Rev.* **2019**, *31*, 319–337. [CrossRef]
- Moreno, R. Cognitive load theory: More food for thought. *Instr. Sci.* **2010**, *38*, 135–141. [CrossRef]
- Seufert, T. The interplay between self-regulation in learning and cognitive load. *Educ. Res. Rev.* **2018**, *24*, 116–129. [CrossRef]
- Gerjets, P.; Scheiter, K.; Cierniak, G. The scientific value of cognitive load theory: A research agenda based on the structuralist view of theories. *Educ. Psychol. Rev.* **2009**, *21*, 43–54. [CrossRef]
- Ginns, P.; Leppink, J. Special issue on cognitive load theory. *Educ. Psychol. Rev.* **2019**, *31*, 255–259. [CrossRef]
- Repovs, G.; Baddeley, A. The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience* **2006**, *139*, 5–21. [CrossRef]
- Yang, H.; Yang, S.; Isen, A.M. Positive affect improves working memory: Implications for controlled cognitive processing. *Cogn. Emot.* **2013**, *27*, 474–482. [CrossRef]
- Schmeichel, B.J.; Volokhov, R.N.; Demaree, H.A. Working memory capacity and the self-regulation of emotional expression and experience. *J. Personal. Soc. Psychol.* **2008**, *95*, 1526–1540. [CrossRef] [PubMed]
- Adams, E.J.; Nguyen, A.T.; Cowan, N. Theories of working memory: Differences in definition, degree of modularity, role of attention, and purpose. *Lang. Speech Hear. Serv. Sch.* **2018**, *49*, 340–355. [CrossRef] [PubMed]
- Chassy, P.; Lahaye, R.; Gobet, F. Templates but not emotions facilitate the information flow between long-term and working memory: A Sternberg study with chess experts. *J. Expert. March* **2023**, *6*.
- Cowan, N. Metatheory of storage capacity limits. *Behav. Brain Sci.* **2001**, *24*, 154–176. [CrossRef]
- Martin, S. A critical analysis of the theoretical construction and empirical measurement of cognitive load. In *Cognitive Load Measurement and Application: A Theoretical Framework for Meaningful Research and Practice*; Zheng, R.Z., Ed.; Routledge: New York, NY, USA, 2018; pp. 29–44.
- Kane, M.J.; Engle, R.W. The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychon. Bull. Rev.* **2002**, *9*, 637–671. [CrossRef]
- Norris, D.; Kalm, K. Chunking and data compression in verbal short-term memory. *Cognition* **2021**, *208*, 104534. [CrossRef]
- Thalmann, M.; Souza, A.S.; Oberauer, K. How does chunking help working memory? *J. Exp. Psychol. Learn. Mem. Cogn.* **2019**, *45*, 37–55. [CrossRef]
- Sweller, J. Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instr. Sci.* **2004**, *32*, 9–31. [CrossRef]
- Paas, F.G. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *J. Educ. Psychol.* **1992**, *84*, 429–434. [CrossRef]
- Paas, F.; Renkl, A.; Sweller, J. Cognitive load theory and instructional design: Recent developments. *Educ. Psychol.* **2003**, *38*, 1–4. [CrossRef]
- Chen, O.; Castro-Alonso, J.C.; Paas, F.; Sweller, J. Extending cognitive load theory to incorporate working memory resource depletion: Evidence from the spacing effect. *Educ. Psychol. Rev.* **2018**, *30*, 483–501. [CrossRef]
- Ayres, P. Subjective measures of cognitive load: What can they reliability measure? In *Cognitive Load Measurement and Application: A Theoretical Framework for Meaningful Research and Practice*; Zheng, R.Z., Ed.; Routledge: New York, NY, USA, 2018; pp. 9–28.
- Kalyuga, S.; Plass, J.L. Cognitive load as a local characteristic of cognitive processes: Implications for measurement approaches. In *Cognitive Load Measurement and Application: A Theoretical Framework for Meaningful Research and Practice*; Zheng, R.Z., Ed.; Routledge: New York, NY, USA, 2018; pp. 59–74.

31. Bandura, A.; Barbaranelli, C.; Caprara, G.V.; Pastorelli, C. Multifaceted impact of self-efficacy beliefs on academic functioning. *Child Dev.* **1996**, *67*, 1206–1222. [CrossRef]
32. Paas, F.; Tuovinen, J.E.; van Merriënboer, J.J.; Darabi, A.A. A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educ. Technol. Res. Dev.* **2005**, *53*, 25–34. [CrossRef]
33. Pintrich, P.R. Implications of psychological research on student learning and college teaching for teacher education. In *Handbook of Research on Teacher Education*; Houston, W.R., Ed.; Macmillan: New York, NY, USA, 1990; pp. 826–857.
34. Schunk, D.H.; Pintrich, P.R.; Meece, J.L. *Motivation in Education: Theory, Research and Applications*, 3rd ed.; Merrill: Upper Saddle River, NJ, USA, 1996.
35. Wigfield, A.; Eccles, J.S. Expectancy-value theory of achievement motivation. *Contemp. Educ. Psychol.* **2000**, *25*, 68–81. [CrossRef]
36. Feldon, D.F.; Franco, J.; Chao, J.; Peugh, J.; Maahs-Fladung, C. Self-efficacy change associated with a cognitive load-based intervention in an undergraduate biology course. *Learn. Instr.* **2018**, *56*, 64–72. [CrossRef]
37. Feldon, D.F.; Brockbank, R.; Litson, K. Direct effects of cognitive load on self-efficacy during instruction. *J. Educ. Psychol.* **2023**. [CrossRef]
38. Huang, X. Example-based learning: Effects of different types of examples on student performance, cognitive load and self-efficacy in a statistical learning task. *Interact. Learn. Environ.* **2017**, *25*, 283–294. [CrossRef]
39. Huang, X. Personal communication. *October* **2018**, *12*, 2018.
40. Kalyuga, S. Cognitive load theory: How many types of load does it really need? *Educ. Psychol. Rev.* **2011**, *23*, 1–19. [CrossRef]
41. Sweller, J. Element interactivity and intrinsic 2011, extraneous, and germane cognitive load. *Educ. Psychol. Rev.* **2010**, *22*, 123–138. [CrossRef]
42. van Merriënboer, J.J.; Sweller, J. Cognitive load theory and complex learning: Recent developments and future directions. *Educ. Psychol. Rev.* **2005**, *17*, 147–177. [CrossRef]
43. Chen, O.; Paas, F.; Sweller, J. A cognitive load theory approach to defining and measuring task complexity through element interactivity. *Educ. Psychol. Rev.* **2023**, *35*, 63. [CrossRef]
44. Kester, L.; Paas, F.; van Merriënboer, J.J.G. Instructional control of cognitive load in the design of complex learning environments. In *Cognitive Load Theory*; Plass, J.L., Moreno, R., Brünken, R., Eds.; Cambridge University Press: Cambridge, UK, 2010; pp. 109–130.
45. Sweller, J. Cognitive load theory, learning difficulty, and instructional design. *Learn. Instr.* **1994**, *4*, 295–312. [CrossRef]
46. Sweller, J.; Chandler, P. Why some material is difficult to learn. *Cogn. Instr.* **1994**, *12*, 185–233. [CrossRef]
47. Klepsch, M.; Seufert, T. Understanding instructional design effects by differentiated measurement of intrinsic, extraneous, and germane cognitive load. *Instr. Sci.* **2020**, *48*, 45–77. [CrossRef]
48. Pekrun, R.; Linnenbrink-Garcia, L. Academic emotions and student engagement. In *Handbook of Research on Student Engagement*; Springer: Boston, MA, USA, 2012; pp. 259–282.
49. Prinz, J. Emotions: How many are there? In *The Oxford Handbook of Philosophy of Cognitive Science*, 1st ed.; Oxford University Press: Oxford, UK, 2012; pp. 183–200. [CrossRef]
50. Scherer, K.R.; Moors, A. The emotion process: Event appraisal and component differentiation. *Annu. Rev. Psychol.* **2019**, *70*, 719–745. [CrossRef] [PubMed]
51. Zhang, L.; Liu, S.; Liu, X.; Zhang, B.; An, X.; Ming, D. Emotional arousal and valence jointly modulate the auditory response: A 40-Hz ASSR study. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2021**, *29*, 1150–1157. [CrossRef]
52. Schweizer, S.; Hampshire, A.; Dalgleish, T. Extending brain-training to the affective domain: Increasing cognitive and affective executive control through emotional working memory training. *PLoS ONE* **2011**, *6*, e24372. [CrossRef] [PubMed]
53. Pham, M.T. Emotion and rationality: A critical review and interpretation of empirical evidence. *Rev. Gen. Psychol.* **2007**, *11*, 155–178. [CrossRef]
54. Pekrun, R.; Linnenbrink-Garcia, L. Introduction to emotions in education. In *International Handbook of Emotions in Education*; Routledge: London, UK, 2014; pp. 1–10.
55. Plass, J.L.; Kaplan, U. Emotional design in digital media for learning. In *Emotions, Technology, Design, and Learning*; Academic Press: Cambridge, MA, USA, 2016; pp. 131–161.
56. Russell, J.A. Core affect and the psychological construction of emotion. *Psychol. Rev.* **2003**, *110*, 145. [CrossRef]
57. Strongman, K.T. *The Psychology of Emotion*, 3rd ed.; Wiley: Chichester, UK, 1987.
58. Lazarus, R.S. *Emotion and Adaptation*; Oxford University Press: New York, NY, USA, 1991.
59. Pekrun, R.; Perry, R.P. Control-value theory of achievement emotions. In *International Handbook of Emotions in Education*; Routledge: London, UK, 2014; pp. 120–141.
60. Huang, X.; Li, S.; Wang, T.; Lajoie, S.P. The effects of emotion regulation and students' perceived challenges on emotion synchrony in collaborative learning. *Res. Square* **2024**. [CrossRef]
61. Frijda, N.H. *The Laws of Emotion*; Psychology Press: London, UK, 2017. [CrossRef]
62. von Scheve, C.; Ismer, S. Towards a theory of collective emotions. *Emot. Rev.* **2013**, *5*, 406–413. [CrossRef]
63. Järvenoja, H.; Järvelä, S.; Törmänen, T.; Näykki, P.; Malmberg, J.; Kurki, K.; Mykkänen, A.; Isohätälä, J. Capturing motivation and emotion regulation during a learning process. *Frontline Learn. Res.* **2018**, *6*, 85–104. [CrossRef]
64. Gross, J.J. Emotion regulation: Current status and future prospects. *Psychol. Inq.* **2015**, *26*, 1–26. [CrossRef]
65. Pekrun, R. The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educ. Psychol. Rev.* **2006**, *18*, 315–341. [CrossRef]

66. Covington, M.V. *Making the Grade: A Self-Worth Perspective on Motivation and School Reform*; Cambridge University Press: Cambridge, UK, 1992.
67. Martin, A.J.; Marsh, H.W. Academic resilience and its psychological and educational correlates: A construct validity approach. *Psychol. Sch.* **2006**, *43*, 267–282. [CrossRef]
68. Sarason, I.G.; Sarason, B.R. Test anxiety. In *Handbook of Social and Evaluation Anxiety*; Leitenberg, H., Ed.; Plenum Press: New York, NY, USA, 1990; pp. 475–495.
69. Spielberger, C.D. Assessment of state and trait anxiety: Conceptual and methodological issues. *South. Psychol.* **1985**, *2*, 6–16.
70. Tobias, S. Test anxiety: Interference, defective skills, and cognitive capacity. *Educ. Psychol.* **1985**, *20*, 135–142. [CrossRef]
71. Zohar, D. An additive model of test anxiety: Role of exam-specific expectations. *J. Educ. Psychol.* **1998**, *90*, 330–340. [CrossRef]
72. Martin, A.J.; Marsh, H.W. Academic buoyancy: Towards an understanding of students' everyday academic resilience. *J. Sch. Psychol.* **2008**, *46*, 53–83. [CrossRef] [PubMed]
73. Aldao, A.; Nolen-Hoeksema, S.; Schweizer, S. Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clin. Psychol. Rev.* **2010**, *30*, 217–237. [CrossRef] [PubMed]
74. Blanchette, I.; Richards, A. The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. *Cogn. Emot.* **2010**, *24*, 561–595. [CrossRef]
75. Bodenhausen, G.V. Emotions, arousal, and stereotype-based discrimination: A heuristic model of affect and stereotyping. In *Affect, Cognition, and Stereotyping*; Mackie, D.M., Hamilton, D.L., Eds.; Academic Press: San Diego, CA, USA, 1993; pp. 13–37.
76. Cohen, J.B.; Pham, M.T.; Andrade, E.B. The nature and role of affect in consumer judgment and decision making. In *Handbook of Consumer Psychology*; Haugtvedt, C.P., Herr, P.M., Kardes, F.R., Eds.; Erlbaum: Mahwah, NJ, USA, 2007.
77. Fredrickson, B.L. The role of positive emotions in positive psychology: The broaden- and-build theory of positive emotions. *Am. Psychol.* **2001**, *56*, 218–226. [CrossRef]
78. Pekrun, R.; Goetz, T.; Frenzel, A.C.; Barchfeld, P.; Perry, R.P. Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemp. Educ. Psychol.* **2011**, *36*, 36–48. [CrossRef]
79. Pekrun, R.; Stephens, E.J. Achievement emotions: A control-value approach. *Soc. Personal. Psychol. Compass* **2010**, *4*, 238–255. [CrossRef]
80. Zeidner, M. The origins and development of test anxiety. In *Test Anxiety: The State of the Art*; Springer: Boston, MA, USA, 1998; pp. 145–170.
81. Parrott, W.G.; Spackman, M.P. Emotion and memory. *Handb. Emot.* **2000**, *2*, 476–490.
82. Gendolla, G.H.E. On the impact of mood on behavior: An integrative theory and a review. *Rev. Gen. Psychol.* **2000**, *4*, 378–408. [CrossRef]
83. Siemer, M. Mood-specific effects on appraisal and emotion judgements. *Cogn. Emot.* **2001**, *15*, 453–485. [CrossRef]
84. Harley, J.M.; Pekrun, R.; Taxer, J.L.; Gross, J.J. Emotion regulation in achievement situations: An integrated model. *Educ. Psychol.* **2019**, *54*, 106–126. [CrossRef]
85. Koole, S.L. Does emotion regulation help or hurt self-regulation. *Psychol. Self-Regul. Cogn. Affect. Motiv. Process.* **2009**, *11*, 217–231.
86. Koole, S.L. The psychology of emotion regulation: An integrative review. *Cogn. Emot.* **2010**, *23*, 4–41. [CrossRef]
87. Bargh, J.A.; Williams, L.E. The case for nonconscious emotion regulation. In *Handbook of Emotion Regulation*; Gross, J.J., Ed.; Guilford Press: New York, NY, USA, 2007; pp. 429–445.
88. Koole, S.L.; Kuhl, J. Dealing with unwanted feelings: The role of affect regulation in volitional action control. In *Handbook of Motivation Science*; Shah, J., Gardner, W., Eds.; Guilford Press: New York, NY, USA, 2007.
89. Mauss, I.B.; Bunge, S.A.; Gross, J.J. Automatic emotion regulation. *Soc. Personal. Psychol. Compass* **2007**, *1*, 146–167. [CrossRef]
90. Ochsner, K.N.; Gross, J.J. The cognitive control of emotion. *Trends Cogn. Sci.* **2005**, *9*, 242–249. [CrossRef]
91. Ochsner, K.N.; Gross, J.J. Cognitive emotion regulation: Insights from social cognitive and affective neuroscience. *Curr. Dir. Psychol. Sci.* **2008**, *17*, 153–158. [CrossRef] [PubMed]
92. Gross, J.J. Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology. *J. Personal. Soc. Psychol.* **1998**, *74*, 224–237. [CrossRef] [PubMed]
93. Porges, S.W. The polyvagal perspective. *Biol. Psychol.* **2007**, *74*, 116–143. [CrossRef] [PubMed]
94. Rothmund, K.; Voss, A.; Wentura, D. Counter-regulation in affective attentional bias: A basic mechanism that warrants flexibility in motivation and emotion. *Emotion* **2008**, *8*, 34–46. [CrossRef]
95. Gross, J.J. Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology* **2002**, *39*, 281–291. [CrossRef] [PubMed]
96. Schweizer, S.; Grahn, J.; Hampshire, A.; Mobbs, D.; Dalgleish, T. Training the emotional brain: Improving affective control through emotional working memory training. *J. Neurosci.* **2013**, *33*, 5301–5311. [CrossRef]
97. Inzlicht, M.; Schmeichel, B.J.; Macrae, C.N. Why self-control seems (but may not be) limited. *Trends Cogn. Sci.* **2014**, *18*, 127–133. [CrossRef] [PubMed]
98. Suri, G.; Whittaker, K.; Gross, J.J. Launching reappraisal: It's less common than you might think. *Emotion* **2015**, *15*, 73. [CrossRef] [PubMed]
99. Barger, S.D.; Bachen, E.A.; Marsland, A.L.; Manuck, S.B. Repressive coping and blood measures of disease risk: Lipids and endocrine and immunological responses to a laboratory stressor 1. *J. Appl. Soc. Psychol.* **2000**, *30*, 1619–1638. [CrossRef]

100. Butler, E.A.; Egloff, B.; Wilhelm, F.W.; Smith, N.C.; Erickson, E.A.; Gross, J.J. The social consequences of expressive suppression. *Emotion* **2003**, *3*, 48–67. [CrossRef]
101. John, O.P.; Gross, J.J. Healthy and unhealthy emotion regulation: Personality processes, individual differences, and life span development. *J. Personal.* **2004**, *72*, 1301–1334. [CrossRef] [PubMed]
102. Myers, L.B.; Burns, J.W.; Derakshan, N.; Elfant, E.; Eysenck, M.W.; Phipps, S. Current issues in repressive coping and health. In *Emotion Regulation: Conceptual and Clinical Issues*; Springer: Boston, MA, USA, 2008; pp. 69–86.
103. Weinberger, D.A. The construct validity of the repressive coping style. In *Repression and Dissociation*; Singer, J.L., Ed.; University of Chicago Press: Chicago, IL, USA, 1990; pp. 337–386.
104. Geraerts, E.; Merckelbach, H.; Jelicic, M.; Smeets, E. Long term consequences of suppression of intrusive thoughts and repressive coping. *Behav. Res. Ther.* **2006**, *44*, 1451–1460. [CrossRef] [PubMed]
105. Hayes, J.P.; Morey, R.A.; Petty, C.M.; Seth, S.; Smoski, M.J.; McCarthy, G.; LaBar, K.S. Staying cool when things get hot: Emotion regulation modulates neural mechanisms of memory encoding. *Front. Hum. Neurosci.* **2011**, *4*, 230. [CrossRef] [PubMed]
106. Johns, M.J.; Inzlicht, M.; Schmader, T. Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *J. Exp. Psychol. Gen.* **2008**, *137*, 691–705. [CrossRef]
107. Lane, R.D.; Sechrest, L.; Riedel, R.; Shapiro, D.E.; Kaszniak, A.W. Pervasive emotion recognition deficit common to alexithymia and the repressive coping style. *Psychosom. Med.* **2000**, *62*, 492–501. [CrossRef]
108. Richards, J.M.; Gross, J.J. Personality and emotional memory: How regulating emotion impairs memory for emotional events. *J. Res. Personal.* **2006**, *40*, 631–651. [CrossRef]
109. McRae, K. Cognitive emotion regulation: A review of theory and scientific findings. *Curr. Opin. Behav. Sci.* **2016**, *10*, 119–124. [CrossRef]
110. Folkman, S. Positive psychological states and coping with severe stress. *Soc. Sci. Med.* **1997**, *45*, 1207–1221. [CrossRef] [PubMed]
111. Folkman, S. The case for positive emotions in the stress process. *Anxiety Stress Coping* **2008**, *21*, 3–14. [CrossRef]
112. Derakshan, N.; Eysenck, M.W. Working-memory capacity in high trait-anxious and repressor groups. *Cogn. Emot.* **1998**, *12*, 697–713.
113. Richards, A.; French, C.C.; Keogh, E.; Carter, C. Test anxiety, inferential reasoning and working memory load. *Anxiety. Stress Coping* **2000**, *13*, 87–109. [CrossRef]
114. Mani, A.; Mullainathan, S.; Shafir, E.; Zhao, J. Poverty impedes cognitive function. *Science* **2013**, *341*, 976–980. [CrossRef]
115. Mullainathan, S.; Shafir, E. *Scarcity: Why Having Too Little Means So Much*; Times Books: New York, NY, USA, 2013.
116. Schulz, J.; Miyamoto, A.; Nückles, M. Through the Lens of Scarcity. How Perceived Resource Scarcity Shapes Students Mindsets and Learning Experience (Poster JURE/EARLI 2023 Thessaloniki, Greece). Available online: https://www.researchgate.net/publication/373643680_Through_the_lens_of_scarcity_How_perceived_resource_scarcity_shapes_students_mindsets_and_learning_experience_Poster_JureEarli_2023_Thessaloniki (accessed on 6 August 2024).
117. Eysenck, M.W.; Derakshan, N.; Santos, R.; Calvo, M.G. Anxiety and cognitive performance: Attentional control theory. *Emotion* **2007**, *7*, 336–353. [CrossRef]
118. Martin, A.J.; Marsh, H.W.; Debus, R.L. Self-handicapping and defensive pessimism: Exploring a model of predictors and outcomes from a self-protection perspective. *J. Educ. Psychol.* **2001**, *93*, 87. [CrossRef]
119. Sarason, I.G. Stress, anxiety, and cognitive interference: Reactions to tests. *J. Personal. Soc. Psychol.* **1984**, *46*, 929. [CrossRef]
120. Sarason, I.G. Test anxiety, cognitive interference, and performance. In *Aptitude, Learning, and Instruction*; Routledge: London, UK, 2021; pp. 131–142.
121. Ashcraft, M.H.; Kirk, E.P. The relationships among working memory, math anxiety, and performance. *J. Exp. Psychol. Gen.* **2001**, *130*, 224–237. [CrossRef]
122. Beilock, S.L.; Carr, T.H. When high-powered people fail: Working memory and “choking under pressure” in math. *Psychol. Sci.* **2005**, *16*, 101–105. [CrossRef] [PubMed]
123. Klein, K.; Boals, A. The relationship of life event stress and working memory capacity. *Appl. Cogn. Psychol.* **2001**, *15*, 565–579. [CrossRef]
124. Schmader, T.; Johns, M. Converging evidence that stereotype threat reduces working memory capacity. *J. Personal. Soc. Psychol.* **2003**, *85*, 440–452. [CrossRef]
125. Bar-Haim, Y.; Lamy, D.; Pergamin, L.; Bakermans- Kranenburg, M.J.; van IJzendoorn, M.H. Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychol. Bull.* **2007**, *133*, 1–24. [CrossRef]
126. Engle, R.W. Working memory capacity as executive attention. *Curr. Dir. Psychol. Sci.* **2002**, *11*, 19–23. [CrossRef]
127. Moran, T.P. Anxiety and working memory capacity: A meta-analysis and narrative review. *Psychol. Bull.* **2016**, *142*, 831. [CrossRef] [PubMed]
128. Derakshan, N.; Eysenck, M.W. Anxiety, processing efficiency, and cognitive performance: New developments from attentional control theory. *Eur. Psychol.* **2009**, *14*, 168–176. [CrossRef]
129. Eysenck, M.W.; Derakshan, N. New perspectives in attentional control theory. *Personal. Individ. Differ.* **2011**, *50*, 955–960. [CrossRef]
130. Stout, D.M.; Rokke, P.D. Components of working memory predict symptoms of distress. *Cogn. Emot.* **2010**, *24*, 1293–1303. [CrossRef]
131. Hembree, R. Correlates, causes, effects, and treatment of test anxiety. *Rev. Educ. Res.* **1988**, *58*, 47–77. [CrossRef]

132. Seibert, P.S.; Ellis, H.C. Irrelevant thoughts, emotional mood states, and cognitive task performance. *Mem. Cogn.* **1991**, *19*, 507–513. [CrossRef] [PubMed]
133. Brockbank, R.B.; Feldon, D.F. Conceptualizing cognitive load as motivational cost: A mixed methods perspective. *Forthcoming unpublished manuscript*.
134. Pessoa, L. How do emotion and motivation direct executive control? *Trends Cogn. Sci.* **2009**, *13*, 160–166. [CrossRef]
135. Osaka, M.; Yaoi, K.; Minamoto, T.; Osaka, N. When do negative and positive emotions modulate working memory performance? *Sci. Rep.* **2013**, *3*, 1375. [CrossRef]
136. Martin, A.J. *Using Load Reduction Instruction (LRI) to Boost Motivation and Engagement*; British Psychological Society: Leicester, UK, 2016.
137. Kim, Y.E.; Yu, S.L.; Wolters, C.A.; Anderman, E.M. Self-regulatory processes within and between diverse goals: The multiple goals regulation framework. *Educ. Psychol.* **2023**, *58*, 70–91. [CrossRef]
138. Kim, Y.E.; Zepeda, C.D.; Martin, R.S.; Butler, A.C. Situating cost perceptions: How general cost and motivational regulation predict specific momentary cost dimensions. *Educ. Psychol.* **2023**, *43*, 855–873. [CrossRef]
139. Mueller, C.J.; Classe, F.; Stürmer, B.; Kuchinke, L.; Stelzel, C. Neurocognitive Effects of Self-Determined Choice and Emotional Arousal on Time Estimation. *Adv. Cogn. Psychol.* **2021**, *17*, 161–175. [CrossRef] [PubMed]
140. Prencipe, A.; Kesek, A.; Cohen, J.; Lamm, C.; Lewis, M.D.; Zelazo, P.D. Development of hot and cool executive function during the transition to adolescence. *J. Exp. Child Psychol.* **2011**, *108*, 621–637. [CrossRef] [PubMed]
141. Chevalier, N.; Huber, K.L.; Wiebe, S.A.; Espy, K.A. Qualitative change in executive control during childhood and adulthood. *Cognition* **2013**, *128*, 1–12. [CrossRef] [PubMed]
142. Flake, J.K.; Barron, K.E.; Hulleman, C.; McCoach, B.D.; Welsh, M.E. Measuring cost: The forgotten component of expectancy-value theory. *Contemp. Educ. Psychol.* **2015**, *41*, 232–244. [CrossRef]
143. Seligman, M.E.; Ernst, R.M.; Gillham, J.; Reivich, K.; Linkins, M. Positive education: Positive psychology and classroom interventions. *Oxf. Rev. Educ.* **2009**, *35*, 293–311. [CrossRef]
144. Brand, S.; Reimer, T.; Opwis, K. How do we learn in a negative mood? Effects of a negative mood on transfer and learning. *Learn. Instr.* **2007**, *17*, 1–16. [CrossRef]
145. D'Mello, S.; Graesser, A. Dynamics of affective states during complex learning. *Learn. Instr.* **2012**, *22*, 145–157. [CrossRef]
146. Sarason, I.G.; Sarason, B.R. Teaching cognitive and social skills to high school students. *J. Consult. Clin. Psychol.* **1981**, *49*, 908. [CrossRef]
147. Kellermann, T.S.; Sternkopf, M.A.; Schneider, F.; Habel, U.; Turetsky, B.I.; Zilles, K.; Eickhoff, S.B. Modulating the processing of emotional stimuli by cognitive demand. *Soc. Cogn. Affect. Neurosci.* **2012**, *7*, 263–273. [CrossRef]
148. Harris, C.R.; Pashler, H. Enhanced memory for negatively emotionally charged pictures without selective rumination. *Emotion* **2005**, *5*, 191. [CrossRef] [PubMed]
149. Van Gog, T.; Kester, L.; Paas, F. Effects of concurrent monitoring on cognitive load and performance as a function of task complexity. *Appl. Cogn. Psychol.* **2011**, *25*, 584–587. [CrossRef]
150. Feldon, D.F. Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educ. Psychol.* **2007**, *42*, 123–137. [CrossRef]
151. Joormann, J.; Levens, S.M.; Gotlib, I.H. Sticky thoughts: Depression and rumination are associated with difficulties manipulating emotional material in working memory. *Psychol. Sci.* **2011**, *22*, 979–983. [CrossRef]
152. Schweizer, S.; Dalgleish, T. Emotional working memory capacity in posttraumatic stress disorder (PTSD). *Behav. Res. Ther.* **2011**, *49*, 498–504. [CrossRef] [PubMed]
153. Young, J.Q.; Friedman, K.A.; Sewell, J.L.; Thakker, K.; John, M.; van Merriënboer, J.J.; O'Sullivan, P.S. Influence of emotion on cognitive load experienced by trainees while performing patient handoffs. *Acad. Med.* **2021**, *96*, S221–S222. [CrossRef]
154. Lewis, M.D. Cognition-emotion feedback and the self-organization of developmental paths. *Hum. Dev.* **1995**, *38*, 71–102. [CrossRef]
155. Lewis, M.D. Interacting time scales in personality (and cognitive) development: Intentions, emotions, and emergent forms. In *Microdevelopment: Transition Processes in Development and Learning*; Granott, N., Parziale, J., Eds.; Cambridge University: Cambridge, UK, 2002; pp. 183–212.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

The Effect of Self-Monitoring on Mental Effort and Problem-Solving Performance: A Mixed-Methods Study

Madison Graham ^{1,2,*}, Marinela Ilic ^{1,3}, Martine Baars ⁴, Kim Ouwehand ¹ and Fred Paas ^{1,5}¹ Department of Psychology, Education, and Child Studies, Erasmus University Rotterdam, 3062 PA Rotterdam, The Netherlands; ouwehand@essb.eur.nl (K.O.); paas@essb.eur.nl (F.P.)² Avans University of Applied Sciences, 4817 LL Breda, The Netherlands³ CED-Groep—Educational Advice and Development, 3067 GH Rotterdam, The Netherlands⁴ Behavioural Science Institute, Radboud University, 6525 XZ Nijmegen, The Netherlands; martine.baars@ru.nl⁵ School of Education, University of New South Wales, Sydney, NSW 2033, Australia

* Correspondence: graham@essb.eur.nl

Abstract: Self-regulated learning (SRL) has become increasingly important for learners in the 21st century as they engage more frequently in student-centered education. Accurate self-monitoring of one's learning process is central to SRL. However, research has shown that learners often struggle with monitoring accuracy. One possible explanation is that the process of self-monitoring itself may add cognitive load, potentially reducing both self-monitoring accuracy and overall performance. In the current study, we investigated the effects of self-monitoring on perceived mental effort and performance during problem-solving tasks of increasing complexity. Participants in the experimental condition were instructed to think aloud while problem-solving, while those in the control condition did not receive such instructions. Our findings indicate no interaction effect between the requirement for explicit and continuous self-monitoring and task complexity on perceived mental effort and performance. However, task complexity significantly impacted the perceived mental effort and performance in both conditions, while it affected monitoring accuracy exclusively in the experimental condition. Furthermore, qualitative analyses showed that the participants in the experimental conditions engaged in self-monitoring during 21.2 % of their think-aloud protocols. Contrary to previous research, self-monitoring did not appear to increase cognitive load, although the experimental participants took significantly longer to complete all the tasks. Possible explanations for these findings are discussed.

Keywords: self-regulated learning; cognitive load; self-monitoring; problem-solving; think-aloud protocols

Citation: Graham, M.; Ilic, M.; Baars, M.; Ouwehand, K.; Paas, F. The Effect of Self-Monitoring on Mental Effort and Problem-Solving Performance: A Mixed-Methods Study. *Educ. Sci.* **2024**, *14*, 1167. <https://doi.org/10.3390/educsci14111167>

Academic Editor: Myint Swe Khine

Received: 31 August 2024

Revised: 19 October 2024

Accepted: 21 October 2024

Published: 26 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the advent of the 21st century, education has shifted paradigms to become increasingly student-centered, requiring the development of a new set of student skills that fit this educational model. One of these necessary student skills is self-regulated learning (SRL) [1]. Zimmerman [2] defined self-regulated learning as students taking charge of their learning through their (meta)cognitive processes, motivation, and behaviors. The need for SRL is evident in educational innovations such as problem-based learning [3], flipped classrooms [4], and massive open online courses (MOOCs) [5], as students in higher education are increasingly required to take charge of and critically evaluate their learning processes. This study focuses on metacognitive monitoring—a key component through which SRL functions [1,6].

However, learners often do not effectively monitor their own learning [7–11]. This may be because learning tasks that typically require good SRL skills such as self-monitoring are often high in intrinsic cognitive load, which could lead to a cognitive overload [11,12]. Cognitive load theory (CLT) provides an explanation for this since a central aim of CLT is

to optimize instructional designs to fit the limited working memory capacity of the learner in a way that encourages better performance [13,14]. In short, the working memory is the workbench of the human information processing system and plays a key role in what information will be selected for consolidation in long term memory (e.g., [15]). Tasks that impose high cognitive load are challenging for this limited capacity working memory to process [16]. Furthermore, CLT distinguishes between three types of cognitive load: intrinsic load, which is inherent to the complexity of the task [17]; extraneous load, which arises from the learning environment (instruction, physical environment); and germane load, which refers to the cognitive resources allocated by working memory to manage the intrinsic cognitive [13,14,18]. The three types of cognitive load together make up the total cognitive load a learner has to handle during learning. Hence, next to the intrinsic load already imposed by the task itself, self-monitoring learning might heighten the total amount of cognitive load, leading to an overload for learners.

Research has shown that monitoring may be seen as a concurrent secondary task during a primary (learning) task [11,12], imposing additional cognitive load to the learning process. Therefore, learning may be affected negatively in such situations that already induce high cognitive load. Consequently, reducing the ability to accurately self-monitor one's own performance may be one of these undesirable effects during high cognitive load. For example, in a study by Radović et al. [19], it was found that item difficulty hindered the ability to self-monitor accurately. Therefore, the aim of the current study is to investigate if self-monitoring indeed adds to the cognitive load experienced by learners and thereby reduces monitoring accuracy and performance. Ultimately, this study investigated the effect of self-monitoring on perceived mental effort and performance with increasing task complexity during problem-solving tasks, as well as the effect of task complexity on monitoring accuracy.

Metacognitive Monitoring within Self-Regulated Learning

Self-regulated learning (SRL) is directed by the learners themselves, and as Pintrich [20] aptly wrote, it is "an active, constructive process whereby learners set goals for their learning and then attempt to monitor, regulate, and control their cognition, motivation, and behavior, guided and constrained by their goals and the contextual features in the environment" (p. 453). SRL is a recursive process of learners defining the task at hand, planning how to achieve their task goals, utilizing various study strategies, and adapting their studying using metacognition [21]. Of note are the metacognitive aspects of SRL, which are included in most models of SRL [22]. In essence, metacognition is thinking about one's own cognitions [23] and involves learners being aware and in control of their cognitions [24]. Metacognition can be further distinguished into different regulatory skills, such as planning, organizing, controlling, monitoring, and evaluating [2]. Of interest to the present study is monitoring as it compromises effective checking of understanding of information in tasks such as problem-solving [25]. Monitoring may be seen as a necessary pre-requirement for SRL [1], as learners must identify learning issues and reflect on how these issues are related to their learning process to remedy them [26].

Importantly, for an effective and efficient SRL cycle, monitoring must be accurate, meaning it should align closely with the learner's actual performance [1,27]. One measure of monitoring accuracy is calibration, which traditionally "refers to the accuracy of learners' perceptions of their own performance" [28] (p. 4). For this experiment, calibration will be referred to as self-assessment (SA) accuracy. Various studies have noted that learners frequently struggle with accurate monitoring [7,11], which may have detrimental effects on task performance. This issue arises because inaccurate monitoring leads to metacognitive control decisions based on incorrect judgments. Metacognitive control involves learners' decisions about implementing actions that would benefit their learning [29]. Effective metacognition requires frequent monitoring and control, which function interdependently [30]. Without adequate monitoring, there are no accurate evaluations of performance and effort and therefore no proper control actions taken to improve learning.

For a successful SRL cycle, it is essential that metacognitive control is grounded in accurate monitoring judgments.

Monitoring and Cognitive Load Theory

The central question of this study is whether self-monitoring increases cognitive load, under what conditions this occurs, and how it affects the learning process. While Seufert [31] argues that all phases of self-regulation, including monitoring, increase cognitive load, Peng and Tullis [29] found that monitoring may require only minimal cognitive resources. Given this discrepancy, cognitive load theory is particularly relevant to the study of SRL, especially concerning the monitoring process.

Relevant to the measurement of cognitive load for this study, Paas [32] differentiated cognitive load into two components: mental load, defined by the structural characteristics of a task, and mental effort, which represents the number of cognitive resources a learner uses to engage with the task. Importantly, Paas [32] highlighted that subjective ratings of mental effort needed to solve a task can serve as an indicator of cognitive load. In this experiment, self-reported mental effort ratings were used to measure cognitive load for each task. Additionally, Paas and Van Merriënboer [16] explained that the complexity of a task, or its mental load, often leads to increased mental effort. As task complexity increases, the intrinsic cognitive load experienced by the learner also rises. This can inhibit performance and learning, given the limited cognitive processing available to humans [16]. When intrinsic load increases in complex tasks, additional demands such as monitoring may contribute to cognitive overload, reducing the resources available for effective task processing. Consequently, this may lead to inaccurate self-monitoring, impaired task performance, and hindered learning outcomes.

Optimizing the cognitive load experienced by the learner is crucial for achieving the most effective learning outcomes. However, since monitoring can consume cognitive resources [11], it is also important to optimize the amount of monitoring. As Winne [33] noted, monitoring that is excessive or non-directed may lead to undesirable effects. While prompts require only periodic monitoring, one study by Vangsness and Young [34] found that frequently prompting learners during a standardized exam resulted in worse performance. While Peng and Tullis's [29] study showed that monitoring may only need few cognitive resources, Brünken et al. [17] importantly noted that even if a secondary task needs few mental resources, it could still affect primary task learning. This may be because of the cost of task switching or divided attention. Frequently being prompted to monitor their learning, such as in Vangsness and Young [34], may then have required learners to switch tasks or multitask often.

Task switching has been shown to decrease working memory functioning [35], suggesting that task switching may cost cognitive resources. Additionally, high levels of multitasking during studying have been linked to lower exam performance than those who used less concurrent media technologies during studying [36]. Task switching from a primary task to a secondary task may induce a cost of additional cognitive resources that affects the performance of the primary task. In addition, divided attention may have other repercussions for monitoring accuracy and performance, as Peng and Tullis [29] found that it significantly hindered monitoring calibration accuracy and metacognitive control. Additionally, while people do see there are costs to multitasking, they are not good at predicting how much these costs affect their performance in memory tasks [37]. In this case, the secondary task of constantly monitoring learning may induce costs to the primary task by increasing mental effort, which then may cause cognitive overload as complexity of the task increases, possibly leading to decreased performance of the primary task.

Indeed, in Van Gog et al. [11], it was found that a self-monitoring instruction led to significantly higher cognitive load and lowered performance on a complex puzzle task. However, as Van Gog et al.'s [11] participants only monitored mentally, it is unclear to what extent and how often participants monitored their performance during the puzzle task. Hence, it remains unclear if and how self-monitoring adds to cognitive load and thereby could possibly hinder performance and learning.

Frequency of Monitoring

Another important question is how often do learners spontaneously monitor their learning? It is important to note that the frequency of monitoring may change from task to task, as some tasks may not require frequent monitoring to achieve proper understanding. For example, Trabasso and Magliano [38] showed that learners who read informational texts verbalized more metacognitive statements than those who read narrative stories. However, it is still an open question at what frequency learners would monitor their learning process when learning to solve problems accompanied by a prompt to continuously monitor this process.

The overall monitoring frequency when conducting a task may vary on many factors, but there have been several studies that looked at frequency of monitoring verbalizations using think-aloud protocols. In a study conducted with primary school students, it was found that they only verbalized monitoring for about 8.5 percent on average of their total verbalizations, but the participant percentage of monitoring fluctuated from zero to 27 percent [39]. The 20 teenage participants of Schellings et al.'s [40] study verbalized monitoring cumulatively for 30.6 percent of all the think-aloud protocols, which was second only to executing verbalizations at 39.9 percent. Though the participants were children in both studies, the ability to self-regulate and specifically monitor may increase as children grow older [41], indicating that adults would potentially verbalize monitoring more often than young children but potentially only slightly more than teenagers. However, it is still an open question as to what the monitoring frequency of university students during problem-solving would be.

Taken together, self-monitoring is a central aspect of SRL but could cause additional cognitive load which could harm the accuracy of self-monitoring and subsequent performance. Moreover, it is still unknown what the frequency of self-monitoring during learning to solve problems would be when students are instructed to explicitly monitor their learning aloud.

The Present Study

Think-aloud protocols are considered the most accurate representation of a participant's short-term memory [42]. Additionally, Sweller et al. [13] explained that people are only able to monitor information in the working memory. Therefore, the think-aloud sessions of the experimental group were recorded to ensure participants verbalized their thoughts consistently and to facilitate subsequent qualitative analysis. This approach also allows the experiment to explore the frequency of monitoring as captured in the think-aloud protocols.

From the literature described above, we drew the following hypotheses.

H1a. *There is a main effect of task complexity on mental effort, reflected in higher mental effort for the complex compared to the simple problems.*

H1b. *There is a main effect of the monitoring condition on mental effort, reflected in higher mental effort in the thinking-aloud condition compared to the control condition.*

H1c. *There is an interaction effect between the condition (thinking aloud and control condition) and task complexity on perceived mental effort. More specifically for complex problems, mental effort is expected to be higher in the thinking-aloud group than the control group, but no effect of the condition is expected for the simple problems.*

H2a. *There is a main effect of task complexity on performance, reflected in higher performance scores on the simple problems than on the complex problems.*

H2b. *There is a main effect of the monitoring condition on performance, reflected in higher performance scores in the thinking-aloud condition compared to the control condition.*

H2c. *There is an interaction effect between condition and task complexity on performance. More specifically, the performance scores of complex problems for the thinking-aloud group are expected to be lower than for the control group, but for the simple problems, there is no such expected effect.*

H3. *Within the thinking-aloud condition, there is a main effect of task complexity on self-assessment (SA) accuracy, reflected in lower SA accuracy for complex problems than for simple problems.*

2. Materials and Methods

Participants

Participants were recruited through the Erasmus Behavioral Lab Research Administration System (ERAS) from a population of undergraduate psychology students at Erasmus University, Rotterdam. In the information section of the study, the study was stated to be conducted in English and that all genders were being recruited. Participant recruitment started in the beginning of April 2023, and data collection lasted from the beginning of May 2023 through the beginning of June 2023.

A total of 65 participants participated in the experiment, with $n = 30$ in the experimental condition and $n = 35$ in the control condition. Originally, 70 participated in the experiment, but five were excluded due to technical difficulties or because they did not follow the instructions in the experiment. In the experiment, 51 participants identified as female, 12 identified as male, and 2 identified their gender as “Other”. The mean age was 20.09 ($SD = 1.85$) years old, with participants all between the ages of 18 to 25. Only 16.9 percent of the participants reported that their (primary) native language was English.

Participants were randomized to each of the two conditions by way of alternating participants to each group. Through the computer command script used in the lab, subjects with an odd participant number were assigned to the experimental condition and subjects with an even participant number were assigned to the control condition. Participant numbering started at one and went up in numerical order until all participants had been tested.

All participants gave informed consent about the nature of the study and their rights beforehand and were instructed they could rescind their participation at any time during the experiment. Compensation for successful completion of the study was provided as undergraduate research credits.

Materials

Apparatus and Software

This study was conducted in the Behavioral Laboratory spaces of the hosting institute. Each participant completed the experiment in a closed and quiet individual cubicle with a computer and keyboard with a webcam with built in audio recording. The survey part of the experiment utilized Qualtrics [43], with one survey for the control and a separate survey for the experimental condition.

Demographics

The demographic section of the study asked three simple questions of a participant’s age, gender, and native language. Gender options were listed as female, male, or other. The native language question could be answered by filling out a text box. If a participant put more than one native language in the text box, only the first one written was included in determining the amount of native English speakers.

Pretest

The pretest consisted of nine open-ended questions, adapted from Baars, Vink et al. [10], to assess each participant’s theoretical knowledge of heredity. Participants had 10 minutes to answer all nine questions until the pretest was automatically submitted. They were instructed to answer each question in a sentence or two in English, and if they absolutely could not answer a question, to put a question mark as the answer. An example pretest question is: “What is the genotype of a hereditary quality such as hair color?” The full list of pretest questions can be found in Appendix A.

The pretest performance was assessed by the researchers through a keyword grading scheme that was made by the researchers before the data collection started. The keyword grading scheme used can be found in Appendix B. If the researcher saw the required keywords were present in the answers, the participant received a point per question for nine possible points. If participants did not include the proper keywords but could still describe the answer to the question well enough, the answer was marked as correct. An example of this can also be seen in Appendix B.

Learning Phase

Brief information on genotypes, homozygous and heterozygous alleles, cross tables, and the types of reasoning needed to solve the problems was presented to all participants before the main problem-solving tasks. The cross-table information was not originally included in this learning phase material from a previous master thesis who used it, but it was added after testing the procedure with some undergraduate research assistants who gave feedback that it would be beneficial to include. The cross-table information was added by the author, agreed upon by both primary researchers, and tested again with another undergraduate research assistant. The exact information provided in the whole learning phase can be found in Appendix C.

Instructions for Problem-Solving Tasks

Participants in the control group were instructed they would solve 10 problems. They were instructed to solve the problems as quickly and accurately as they could and if they could not answer a question to fill in a question mark. They were not allowed to use their hands, write things down on paper, or speak when solving the problems. Participants in the experimental group received similar instructions as the control group but were additionally provided with written instructions to think aloud (based on Moos and Bonde [44]) in English: "Please say what you're doing or thinking while you are doing or thinking it". While their audio was recorded, participants were also instructed not to give any personally identifiable information during the recording, not to write anything down on paper, and not to use their hands. It was instructed that the audio recording would start automatically, so they were to start speaking as soon they began to solve the problems. The exact instructions provided to the experimental group can also be found in Appendix G.

Heredity Problem-Solving Task

In this phase, participants were asked to quickly work through ten problem-solving tasks about Mendelian inheritance. The participants solved the tasks in the learning environment where they could write down their answers per step. Both conditions received the same set of tasks. There was no time limit imposed per problem-solving task. The tasks were adapted from material from Baars, Vink et al. [10]. The basic steps of each question, as described by Baars, Vink et al. [10], included determining possible genotypes from phenotypes, completing a heredity diagram, establishing a direction of logical reasoning, calculating how many cross tables were needed, filling in the cross tables, and deriving the final answer from them. There were five complexity levels in total with two questions per complexity level, presented in ascending order of complexity. Three examples of tasks from this experiment can be found in Appendix D. The higher complexity levels had more steps per question than the lower complexity levels. Level four had an extra direction of reasoning and cross table section than previous levels, and level five had an additional final answer multiple choice step on top of that. For the analyses, simple tasks were defined as complexity levels one and two, whereas complex tasks were defined as complexity levels three through five (see Appendix J for the rationale behind this decision).

Performance on the problem-solving tasks was graded by the researchers with an answer sheet that was made by the researchers before the data collection started. In each task, only the steps of determining genotypes, the heredity diagram, the direction of reasoning, and the steps with the final answers were graded. Participants received one point per correct step. For the first three complexity levels, participants could receive a maximum of four points per problem-solving task, for complexity level four they could receive a maximum of five points, and for complexity level five they could receive a total of

six points maximum. Mean percentage scores were calculated for the simple (level 1 and 2) and difficult problems (level 3, 4 and 5) by dividing the score of the participant by the perfect score and averaging them across the simple and complex levels, respectively.

Invested Mental Effort Rating Scale

The participant's subjective mental effort was measured with a single-item nine-point scale, in which participants rated how much mental effort they invested to solve the just previously shown problem-solving task, with a (1) *very, very low mental effort* rating ranging to a (9) *very, very high mental effort* rating [32]. The reliability of the scale was estimated with Cronbach's coefficient alpha. A coefficient of reliability (α) of 0.90 was obtained in the study of Paas [32]. An example of the scale used in the survey can be found in Appendix E. Mean mental effort ratings were calculated for the simple and complex problems.

Absolute Self-Assessment (SA) Accuracy Measures

A self-assessment rating of their own performance was only given to participants in the experimental condition, so as to not have the control condition unintentionally monitor their learning. For the first three complexity levels, participants rated how many points they thought they scored from 0 points to 4 points. For complexity level four, participants rated how many points they thought they scored from 0 points to 5 points. Finally, for complexity level five, they rated how many points they thought they scored from 0 points to 6 points. Examples of the scales can be found in Appendix F. After converting the task performance points and self-assessed points to percentages, absolute SA accuracy was determined by subtracting the self-assessed percentage from the actual performance percentage per task. Then, the absolute value of this accuracy percentage was taken to obtain the absolute SA accuracy, showing the deviation from a perfectly assessed score of zero. Additionally, for each complexity level a mean SA score was calculated. To obtain the mean SA score for the simple tasks, the absolute SA accuracies from complexity levels one and two were all averaged together, and the same was done for levels three through five to obtain the mean SA score for the complex tasks.

Procedure

At the beginning, participants were instructed to wait until the experiment appeared on the screen and to read everything very carefully and follow written instructions. If they had questions they could press a help button, and a researcher would come to answer their question during the experiment. Once finished, they could leave the cubicle, and the researcher would shut down the experiment for them. They completed the experiment alone in a sound-isolated cubicle.

Upon starting the experiment, participants gave informed consent and then reported some demographic information (see the participants and demographics sections for a detailed description). Next, participants completed the heredity pretest section to obtain a baseline assessment of their theoretical knowledge of heredity. Then, in the following learning phase, basic information on genotypes and types of reasoning needed to answer the heredity problem-solving tasks was provided. Afterwards, they moved onto the instructions of the heredity problem-solving task for their respective groups. Control participants did not receive instructions to think aloud, while the experimental participants did, as can be seen in Appendix G. Participants were then asked to answer each of the ten problem-solving tasks. Reminders to think aloud in bold red text were written on every task in the experimental survey version.

In between each of the ten problems, all participants rated their subjective mental effort spent on the previous problem with the mental effort rating scale. Additionally, after each question, those in the experimental condition completed the self-assessment rating. After a participant had completed all 10 problems with accompanying rating scales, the experiment ended, and the recording stopped for the experimental group. Because there were no time limits on the task, the total duration of the experiment ranged between 26 to 78 minutes.

Qualitative Analysis Approach

Audios were transcribed non-verbatim by two research assistants and the first two authors. Transcripts were then segmented by each step in the problem-solving tasks. An adapted version of the base coding scheme by Azevedo et al. [45] (see Appendix H) was used to code the transcripts. The first adaptation we applied was splitting the Judgement of Learning (JOL) code into positive and negative JOLs for more in-depth analysis of the kinds of JOLs the participants verbalized. This would lead also to a better understanding of how participants self-monitored. Secondly, due to the presence of many responses that could not be categorized with the base scheme, we added codes for verbalizations on the reading process (named “Reading the Question”), cognitive load (“Mental Effort”), and coming to an answer (“Stating Answer”). Third, we added a Self-Assessment code for monitoring verbalizations when we prompted participants to self-assess in the task. Thirdly, we removed the Self-Questioning code of Azevedo et al. [45] due to redundancy with the Rereading code, as well as the code Selecting a New Informational Source and Control of Context as participants were unable to execute the definition of these codes in this study.

All codes were given new anchor examples from the transcripts if possible or given a researcher-derived example. Ten transcripts were selected out of the 30 via a random number generator (cf. Johnstone et al. [46]). Because in two transcripts most verbalizations were too unclear, we decided to exclude these from further analysis, leaving eight total transcripts included in the final qualitative analysis. To prepare for the intercoder reliability (ICR) analysis, about 30 percent of the total number of segments in the transcripts used in the analysis were selected (cf., Azevedo et al. [45]), which gave a selection of 17 segments per transcript that were chosen via a random number generator. The segments were compiled into one document and imported into Atlas.ti Version 24 [47] for the ICR analysis. Coder one freely coded the ICR document with the adapted coding scheme, then removed their codings but kept the quotations in the document. Coder two then independently coded the pre-chosen quotations. Cohen’s kappa measures were calculated per code group in SPSS Version 29 [48]. Cohen’s kappa was computed to determine if there was agreement between coder one and two on which code group each verbalization in the ICR analysis belonged to. There was fair agreement for code group “A — Forethought, Planning, and Activation”, $\kappa = 0.401$, $p < 0.001$, 95% CI [0.187, 0.615]. There was substantial agreement for code group “B — Monitoring”, $\kappa = 0.401$, $p < 0.001$, 95% CI [0.187, 0.615]. There was substantial agreement for code group “B—Monitoring”, $\kappa = 0.685$, $p < 0.001$, 95% CI [0.587, 0.783]. For the code group “C — Strategy Use”, there was moderate agreement, $\kappa = 0.498$, $p < 0.001$, 95% CI [0.408, 0.588]. For the code group “D — Task Difficulty, Complexity, and Demands”, there was perfect agreement, $\kappa = 1.000$, $p < 0.001$, 95% CI [1.000, 1.000]. Lastly, for the code group “E — Miscellaneous Actions”, there was moderate agreement, $\kappa = 0.543$, $p < 0.001$, 95% CI [0.449, 0.637]. Disagreements between coders were resolved via discussion to reach full agreement. Both coders gave their reasoning for a chosen code and changed their answer if they agreed with the other coder. Full agreement was important in this case because one coder would code all eight transcripts. Coder one then coded all eight transcripts using Atlas.ti [47]. Additionally, to obtain a frequency and percentage of Monitoring, the Self-Assessment codes were removed from the Monitoring code group.

3. Results

Pretest

A two-tailed independent sample *t*-test revealed no significant differences between the control and experimental conditions on total pretest scores: $t(63) = 1.76$, $p = 0.084$, $d = 0.44$. This indicates that both groups had similar performance on the pretest, suggesting no significant differences in their prior knowledge of Mendelian inheritance.

Quantitative Analysis Approach

Separate one-way mixed ANOVAs were conducted for mental effort and performance, with two factor levels with the within-subject variable for task complexity (simple vs. complex) and the between-subjects variable for condition (thinking aloud vs. no prompted

monitoring), using a Bonferroni confidence interval adjustment. A one-way ANOVA was conducted on self-assessment (SA) accuracy (in the experimental condition only) with a within-subjects factor for task complexity (simple vs. complex), with a Bonferroni confidence interval adjustment.

All statistical analyses were conducted with a significance level of $\alpha \leq 0.05$ unless otherwise stated. Effect sizes were noted where appropriate with Cohen's [49] guidelines on the partial eta-squared effect size being small at 0.01, a medium effect at 0.06, and a large effect at 0.14. The quantitative data were analyzed with SPSS Version 29 [48].

Mental Effort Ratings

There was a main effect for task complexity, $F(1, 63) = 109.62$, at 0.01, a medium effect at 0.06, and a large effect at 0.14. The main effect of task complexity showed that participants rated higher mental effort for the complex problems compared to the simple problems (see Figure A1 of Appendix I). See Table 1 for the means and standard deviations for both simple and complex across conditions for all variables.

Table 1. Means and standard deviations of performance (range 0–100), mental effort ratings (range 1–9), and absolute value of self-assessment accuracy percentage (0 = perfect, 100 = largest deviation in monitoring) in the experimental and control conditions in simple and complex tasks.

		Control ($n = 35$)		Experimental ($n = 30$)	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Simple	Mental Effort	3.96	2.08	3.83	1.57
	Performance	91.79	14.60	86.25	19.03
	SA Accuracy			13.33	16.39
Complex	Mental Effort	5.51	1.68	5.55	1.39
	Performance	86.29	19.79	76.13	24.05
	SA Accuracy			18.30	12.08

Note: There is no absolute self-assessment (SA) accuracy data for the control group as they did not self-assess after any of the problem-solving tasks. An absolute SA accuracy score of zero is a perfectly assessed score, meaning that there was no difference between the actual performance and the self-assessed score.

Performance Scores

There was a main effect for task complexity, $F(1, 63) = 20.77$, $p < 0.001$, $\eta_p^2 = 0.25$, but not for condition $F(1, 63) = 3.00$, $p = 0.088$, $\eta_p^2 = 0.045$. The interaction effect of task complexity and condition on performance was not significant, $F(1, 63) = 1.83$, $p = 0.182$, $\eta_p^2 = 0.028$.

The main effect of task complexity reflects the lower performance on the complex versus the simple problems (see Figure A2 of Appendix I).

Absolute Self-Assessment Accuracy in Experimental Group

There was a main effect of task complexity, $F(1, 29) = 5.62$, $p = 0.025$, $\eta_p^2 = 0.16$, reflecting higher SA accuracy for the complex problems than the simple problems (see Figure A3 of Appendix I). This suggests that participants in the experimental condition became significantly less accurate at evaluating their performance on complex tasks than on simple tasks.

Time to Complete Study

The experimental group on average ($M = 45.40$, $SD = 11.24$) took more time to complete the entire study than the control group ($M = 38.79$, $SD = 8.96$). According to a two-sided independent sample *t*-test, there was a significant difference between the two groups for the time to complete the study, $t(63) = -2.64$, $p = 0.011$, $d = -0.66$. This finding reflects that the participants in the experimental condition, in which they were instructed to explicitly monitor their learning process, invested more effort, in this case time, to learn how to solve the problem-solving tasks.

Qualitative Results on the Code Frequencies

Across all eight transcripts analyzed, there were 1536 coded verbalizations. For the Monitoring code group, there were 351 coded verbalizations, or about 22.9 percent of

all codes as (see Table 2). For the frequency of Spontaneous Monitoring, which required removing the prompted Self-Assessment code frequencies within the Monitoring code group, there were then 325 coded verbalizations, or about 21.2 percent of all coded verbalizations. Additionally, because not every participant verbalized monitoring the same amount, an average frequency and percentage per transcript was calculated. For the Spontaneous Monitoring group specifically, this gave a mean frequency of 40.63 coded verbalizations per transcript, or about on average 21.2 percent of coded verbalizations per transcript. While more examples can be found in Appendix H, one example of a Negative JOL verbalization was “I don’t know if that’s good”. Another example of a Positive JOL verbalization was “Which makes sense, because their other parent has blue eyes”. Finally, an example of Content Evaluation verbalization from a transcript was “You can’t guess from the information”.

Table 2. Self-regulated learning code frequencies across all eight transcripts.

<i>Codes</i>	<i>Raw Frequency</i>	<i>Percentage</i>
<i>A—Forethought, Planning, and Activation</i>	<u>51</u>	<u>3.32</u>
A1—Planning	7	0.46
A2—Sub-Goals	19	1.24
A3—Prior Knowledge Activation	12	0.78
A4—Recycle Goal in Working Memory	13	0.85
<i>B—Monitoring</i>	<u>351</u>	<u>22.85</u>
B1—Negative Judgment of Learning (NJOL)	41	2.67
B2—Positive Judgment of Learning (PJOL)	71	4.62
B3—Self-assessment	26	1.69
B4—Feeling of Knowing (FOK)	38	2.47
B5—Content Evaluation	115	7.49
B6—Identify Adequacy of Information and Task Instructions	60	3.91
<i>C—Strategy Use</i>	<u>480</u>	<u>31.25</u>
C1—Goal-Free Search	0	0
C2—Summarization	9	0.59
C3—Rereading	18	1.17
C4—Inferences	311	20.25
C5—Hypothesizing	25	1.63
C6—Knowledge Elaboration	117	7.62
<i>D—Task Difficulty, Complexity, and Demands</i>	<u>43</u>	<u>2.80</u>
D1—Time and Effort Planning	9	0.59
D2—Help-Seeking Behavior	1	0.07
D3—Task Difficulty	9	0.59
D4—Mental Effort	24	1.56
<i>E—Miscellaneous Actions</i>	<u>611</u>	<u>39.78</u>
E1—Stating Answer	72	4.69
E2—Copying Information	233	15.17
E3—Reading Question	306	19.92

Note: Due to rounding, percentages within the code groups may not add up exactly to the total percentage of the code group.

4. Discussion

This study has contributed to the scientific understanding of cognitive load theory and SRL by investigating the effect of self-monitoring on the perceived mental effort and performance during problem-solving tasks of increasing complexity. In line with H1a and H2a, participants rated higher perceived mental effort when working on the complex compared to the simple problems and performed better on the simple than complex problems. Contrary to the expectations outlined in H1b and H2b, there was no significant effect of the monitoring condition on either mental effort or performance. In addition, we found no support for H1c and H2c, as task complexity had a similar effect on mental effort ratings and performance in both the experimental (monitoring via thinking aloud) and the control condition. In line with H3, participants showed more accurate self-assessments for the simple than the complex problem-solving tasks.

In sum, our findings indicate that there was no interaction effect between having to explicitly and continuously monitor and the task complexity on perceived mental effort or performance. However, participants in the experimental condition did spend more time on learning how to solve the problem-solving tasks compared to the participants in the control condition. In addition, the task complexity did affect the perceived mental effort and performance as well as the monitoring accuracy (only in the experimental condition). Furthermore, the qualitative analyses showed that the participants in the experimental conditions on average were engaged in self-monitoring in 21.2 % of their think-aloud protocols. While this percentage of monitoring in the think-aloud protocols is considerably lower than that which previous studies found for the self-monitoring of children [39,40], this type of heredity problem-solving task was not new to the participants as they had already encountered it in their undergraduate coursework. It is possible then that not as much self-monitoring for this specific task was needed, as they may have already been fluent at the task. However, this does indicate that the instruction that was given to elicit explicit monitoring worked. Nonetheless, monitoring did not seem to increase the cognitive load, in contrast to earlier research by Van Gog et al. [11]. There are some possible explanations for this difference that will be discussed below.

Importantly, the manipulation of the intrinsic cognitive load of the task was successful, as can be seen by the significant positive effect of task complexity on mental effort ratings. However, Paas [32] divided cognitive load into mental load and mental effort, stating that mental load was the instructional demands of a task versus mental effort being the invested mental effort to the task. Paas [32] denoted that this invested mental effort was reflective of the cognitive load of a task. Additionally, Leppink et al. [50] wrote that task complexity helps to determine intrinsic load but also the prior knowledge that participants have. Recent literature has even questioned the use of certain cognitive load-related appraisals like mental effort as they can be biased by other cues like response time [51]. It is not entirely clear, when using invested mental effort measures, if it is possible to obtain a pure measure of intrinsic cognitive load. However, because the format and wording of the steps of the ten increasingly complex problems were kept as consistent as possible with limited distractions, it is thought that extraneous cognitive load was kept at a constant. However, that leaves the possibility that the invested mental effort could also have been affected by intrinsic and germane load. It is possible that as intrinsic cognitive load increases, germane load increases as well for participants to handle the intrinsic cognitive load. Notably, Paas et al. [18] described that germane load can be considered as the resources needed by the working memory to manage the intrinsic load. Therefore, though the invested mental effort ratings did increase significantly as intended as the task complexity increased, it could have been caused by other factors next to intrinsic load, such as germane load.

There was also a significant negative effect of task complexity on performance, which aligns with H2a. As Paas and Van Merriënboer [16] noted, performance and learning during complex tasks can be hindered by the limited cognitive processing capacity of humans. As the intrinsic load increased, it likely placed greater demands on the participants' available cognitive resources while they engaged in the problem-solving tasks. On simple tasks, this may not have been an issue as the general cognitive load of the task was low, leaving plenty of resources to process the material. However, once the intrinsic cognitive load of a task became too high on the highest complexities, there may have been not enough cognitive resources for the germane load to adequately help the processing of the highest complexity level tasks. This overload may then have the consequences of incorrect and inefficient processing of the materials, possibly leading then to decreased performance among both groups due to the increasing complexity of the tasks. These results indicate that our working memory plays an important role in problem-solving, especially when a task becomes more complex. Teachers should be aware of this when checking the complexity of a task. If the task becomes too complex it may negatively impact the learning process and may lead to worse performance.

Notably, no significant interaction effects between self-monitoring and task complexity on mental effort or performance were observed, not giving evidence towards H1c nor H2c. While it was expected to find similar results to Van Gog et al. [11], this was not the case. However, a significant difference between the experimental and control condition in the time the participants spent on learning how to solve the problem-solving tasks was found. This indicates that the participants in the experimental condition, in which they explicitly monitored their learning process, did invest more effort (i.e., time) but did not reach a higher performance compared to the control condition. This possibly also explains the difference between the findings in the current study and the study by Van Gog et al. [11]. While Van Gog et al. [11] only used two Sudoku puzzle tasks, one simple and one complex, they gave their participants a time limit of two minutes per puzzle. Importantly, Paas and Van Merriënboer [16] noted that tasks with time limits often have a higher cognitive load. If there had been a time limit imposed per question in this study, it is hypothesized that there would have been an interaction effect of task complexity and monitoring condition. Setting a time limit per problem-solving task potentially would have caused higher cognitive load within the think-aloud condition, as then the think-aloud condition would not have the time and subsequent cognitive resources available to task switch between solving the task and constantly monitoring their learning by thinking aloud. This could possibly create even higher demands on participants' cognitive resources, which could be reflected in their perceived mental effort and performance. These results indicate that time is an important factor and should be taken into account when giving students a complex task. By giving students enough time to finish their work, students will not experience more mental effort while monitoring their learning.

Though there were frequent reminders to think aloud within the experimental version of the Qualtrics survey and audio was checked to make sure participants spoke consistently, participants did take pauses in their thinking aloud, often when executing an action such as typing. These pauses may have served as a miniature mental reprieve from monitoring, or they can be seen as microbreaks [52]. Previous studies like Tyler and Burns [53] have shown that 10-minute breaks between regulatory tasks can replenish self-regulatory resources, resulting in similar performance by experimental and control participants. Though there is no ideal micro-break duration established yet [52], another study has shown even just 3-minute breaks were beneficial for performance [54]. Without time pressure added, participants in the experimental condition in the present study could take brief pauses at their leisure, which may have added up, allowing them to potentially gain cognitive resources back over time. This may be a possible explanation for the similar mental effort and performance found in both conditions across the complexity levels, and for why the experimental condition took significantly longer to complete the whole experiment. Therefore, a lack of time limit per question may be one explanation for the results found. Future research could investigate the effect of self-monitoring on mental effort and performance while using a time limit per problem-solving task to investigate this further.

In addition, there was a significant effect of task complexity on absolute SA accuracy, meaning in the context of the data the experimental group participants were significantly worse at judging their performance in the complex tasks than in the simple ones. They had a greater absolute deviation between actual and self-assessed scores on the complex tasks. This is in line with results found in earlier research by Baars et al. [8,9] and Radović et al. [19], who found that SA accuracy decreased with increasing task complexity. Additionally, this is in line with previous research by Nietfeld et al. [55] which found that students showed more accurate judgements of performance on easier items than more difficult ones. Just as the increasing task complexity affected both mental effort and performance, it may be that, due to the cognitive overload on the highest complexity levels caused by the higher intrinsic and germane load, there were fewer cognitive resources available to accurately monitor their learning [12]. When complexity becomes too high, then it may have negative effects on learners' ability to judge their own performance.

There are some limitations to the current study. First, the lack of a time limit per problem-solving task made it harder to find an effect of self-monitoring on mental effort and performance. However, participants evidently needed more time to explicitly monitor their learning process, which could also be seen as an indication of an effect of self-monitoring on the effort participants invested in the learning process. Second, there was an inconsistent point system between the complexity levels in the problem-solving tasks. For example, in level one participants could score up to four points but for level five participants could score up to six. While the performance scores and SA accuracy were converted to percentages for the analysis, if this study were to be redone, a consistent point system could be arranged. Additionally, due to the open-ended nature of questions, all grading had to be done by the first two authors, potentially introducing bias when answers were partially correct. This could have been corrected by making the entire experiment able to be graded by computer. Finally, the small sample size may have impacted the statistical power and the reliability of our findings. While our limitations could have influenced the results, the observed differences in performance between groups, though not statistically significant, might have reached significance with a larger sample size or if a time limit had been imposed.

5. Conclusions

In the current study, the effect of self-monitoring using a think-aloud method on mental effort and performance was investigated for problems at increasing complexity levels. Although no interaction effect between self-monitoring and task complexity was found for mental effort and performance, participants in the experimental condition did spend more time on learning how to solve the problems. In addition, task complexity significantly increased mental effort and lowered performance in both conditions, and it lowered self-assessment accuracy (in the experimental condition only). These findings indicate that both monitoring and task complexity can take up cognitive resources from participants, which consequently affects their mental effort, self-assessment accuracy, performance, and the time they need to work on problem-solving tasks. As participants in the experimental condition used more time to work on the problem-solving tasks, it is possible that they were able to replenish their cognitive resources over time to be able to perform similarly to the control condition in both simple and complex tasks. Therefore, the time given to solve a task seems to be an important factor when considering self-monitoring. Additionally, it was found that participants in the experimental group verbalized self-monitoring for on average 21.2% of the think-aloud protocols, meaning that the experimental instructions elicited self-monitoring. Future research could investigate the effect of self-monitoring on mental effort and performance while using a time limit per problem-solving task to investigate this further. Implication-wise, because task complexity had an effect on self-assessment accuracy, teachers should be mindful of the complexity of a task when asking to students to self-assess.

Author Contributions: Conceptualization, M.G., M.I. and M.B.; methodology, M.G., M.I. and M.B.; data collection and analysis, M.G. and M.I.; writing—original draft preparation, M.G., M.I., M.B., K.O. and F.P.; writing—review and editing, M.G., M.I., M.B., K.O. and F.P.; supervision, M.B., K.O. and F.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of Erasmus University Rotterdam (protocol code ETH2223-0393 and 8 March 2023).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: The authors express gratitude to the Erasmus Behavioral Lab for their invaluable contributions to this study. The authors would also like to thank the three bachelor-level research assistants, Ivan Orekhov, Yu Kong, and David Hahn, who helped with testing the experimental procedures, data collection, and transcribing

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Hereditary Pretest

1. What is the genotype of a hereditary quality such as hair color?
2. What is the phenotype of a hereditary quality such as eye color?
3. What does it mean if a person is homozygous for a hereditary quality such as hair color?
4. What does it mean if a person is heterozygous for a hereditary quality such as eye color?
5. What does it mean if an allele in a genotype is dominant?
6. What does it mean if an allele in a genotype is recessive?
7. Which feature will be visible if the same recessive alleles are present in the genotype?
8. Which feature will be visible if two different alleles are present in the genotype?
9. What do you have to do to find the genotypes of children if you know the genotypes of the two parents?

Appendix B

Keyword Grading Scheme for Pretest

Question	Answer	Keywords
1	Genotype refers to the two alleles a person/organism has inherited for a particular gene	1. Alleles 2. Gene(s)
2	Phenotype is how this genetic information (genotype) is reflected in physical characteristics	1. Genetic information/genotype 2. Physical
3	Homozygous means that both alleles are the same	1. Alleles 2. both are the same/other words that describe this
4	Heterozygous means that there is dominant and a recessive allele	1. Alleles 2. Different
5	A dominant allele produces the dominant phenotype. Even with one dominant allele they will have the dominant phenotype	1. Dominant phenotype 2. one copy (3. Uppercase)
6	A recessive allele produces the recessive phenotype. To produce this phenotype the genotype must have two copies of the recessive allele.	1. Recessive phenotype 2. Two copies needed (3. Lowercase)
7	The recessive phenotype	1. Recessive
8	The dominant phenotype	2. Dominant
9	Put the genotypes of the parents in a cross table so you can form the possible combinations of genotypes for the child.	1. Cross table or Punnett Square

An example of not using the keywords but being given a point on the pretest:

Question one was “What is the genotype of a hereditary quality such as hair color?” One participant answered, “its the genetics passed on so like HH or sm”. This was given a point for correctness because while they did not use either of the keywords of gene(s) or allele(s) for their answer, they did use the word genetics, and it was clear to the researchers that they understood that from Mendelian inheritance that a genotype was made up of two alleles which are represented with two letters, either upper- or lower-case.

Appendix C

Learning Phase Material

In the upcoming section, you will be asked to solve 10 problems on heredity and genetics. Here is some information you will need to be able to solve the problems. Please read the information carefully before you begin.

Genotype

1. A genotype (such as for hair shape) is the information that lies in the genes and consists of two alleles (e.g., BB, Bb, or bb).

Phenotype

2. A phenotype is how this information is reflected in physical characteristics (e.g., curly or straight hair).

Dominant and Recessive Alleles

3. A dominant allele (represented by a capital/uppercase letter) produces the dominant phenotype. If individuals have even only one dominant allele in their genotype, they will have the dominant phenotype. This dominant allele can come from one parent or both.
4. For a recessive allele (represented by a lowercase letter), to produce a recessive phenotype, the individual's genotype must have two copies of the recessive allele, which means they received one recessive allele from each parent.
5. For hair shape, the dominant allele for curly hair is (H) and the recessive allele for straight hair is (h).
6. Each genotype is composed of two alleles (i.e., HH, Hh, or hh). A dominant phenotype (i.e., curly hair) appears when an individual's genotype has at least one copy of the dominant allele (i.e., HH or Hh).
7. A recessive phenotype (i.e., straight hair) only appears when an individual's genotype has two recessive alleles (i.e., hh).

Homozygous or heterozygous?

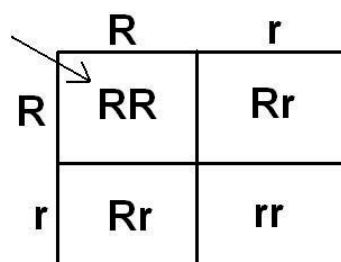
8. Homozygous means that both alleles are the same. This could mean there are two uppercase letters or two lowercase letters (i.e., HH or hh).
9. Heterozygous means that there is a dominant and a recessive allele in the genotype (i.e., Hh).

Type of Reasoning

10. Deductive reasoning means you reason from the parents down to the child.
11. Inductive reasoning means you reason from a child and one parent up to the other parent.

Cross Tables/Punnett Squares

12. To use a cross table/Punnett square, put the two alleles from the genotype of the first parent on the left-hand side of the cross table so that one allele is in each row (e.g., see image below). For the second parent, put one allele on top of the first column and the second allele in the next column. In this example, each parent is heterozygous for a trait. Combine each of the alleles, one from the row and one from the column, in the appropriate boxes to form all the possible genotypes of their child. It is then possible that their child could have the genotypes of RR, Rr, OR rr for this trait.



	R	r
R	RR	Rr
r	Rr	rr

Appendix D

Appendix D.1. Level 1 Complexity Hereditary Example Problem

Eye color

Eye color in humans is partly determined by a gene, with brown (B) being dominant and blue (b) recessive. Two parents are having a child. Both parents have brown eyes and are homozygous for this trait.

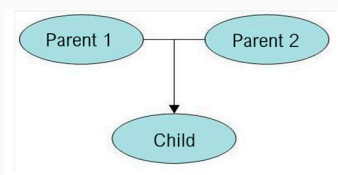
What would be the possible genotype(s) of the child?

Step 1: Translate into genotypes (1 point)

Parent 1

Parent 2

Step 2: Enter the genotypes from step 1 in a family chart. Please use the boxes below the example. For the unknown family member(s) you can use a question mark. (1 point)



Step 3: Determine direction of your reasoning. (1 point)

☐ Deductive

☐ Inductive

Step 4: Determine the number of cross tables you need and fill them in.

☐ 1 cross table

☐ 2 cross tables

☐ 3 cross tables

Step 5: Answer (1 point)

Compare the outcomes you received from the cross table(s) and use them to determine your final answer. Mark the right answer below.

What would be the possible genotype(s) of the child?

☐ BB

☐ Bb

☐ bb

☐ BB and Bb

☐ BB and bb

☐ Bb and bb

☐ BB, Bb, and bb

Appendix D.2. Level 3 Complexity Hereditary Example Problem

Eye Color

Eye colour in humans is partly determined by a gene, with brown (B) being dominant and blue (b) recessive. Two parents are having a child. One parent has brown eyes and is heterozygous for this trait. No information is known about the other parent. The child they are having has blue eyes and is homozygous for this trait.

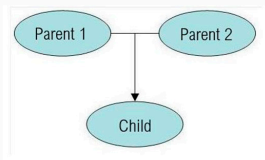
What could be the possible genotype(s) of the parent?

Step 1: Translate to genotypes (1 point)

Parent 1

Child

Step 2: Enter the genotypes from step 1 in a family chart. Please use the boxes below the example. For the unknown family member(s) you can use a question mark. (1 point)



Step 3: Determine direction of your reasoning. (1 point)

- ☐ Deductive
- ☐ Inductive

Step 4: Determine the number of cross tables you need and fill them in.

- ☐ 1 cross table
- ☐ 2 cross tables
- ☐ 3 cross tables

Cross table 1

	Information parent 1	Information parent 2	Information parent 2
Information parent 1			
Information parent 1			

Step 5: Answer (1 point)

Compare the outcomes you received from the cross table(s) and use them to determine your final answer. Mark the right answer below.

What could be the possible genotype(s) of the parent?

- ☐ BB
- ☐ Bb
- ☐ bb
- ☐ BB and Bb
- ☐ BB and bb
- ☐ Bb and bb
- ☐ BB, Bb and bb

Appendix D.3. Level 5 Complexity Hereditary Example Problem

Hair type

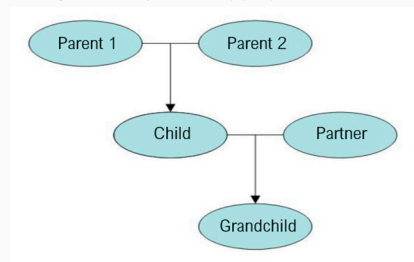
Hair type in humans is partly determined by a gene, with curly hair (H) being dominant and straight hair (h) being recessive. Two parents (generation P) have a child (F1). One parent has curly hair and is homozygous for this trait. The other parent has straight hair and is homozygous for this trait as well. No information is known about their child. This child (F1) and their partner have a (grand)child (F2). The partner has curly hair and is heterozygous for this trait. No information is known about the (grand)child (F2).

What could be the possible genotype(s) of the child (F1) and the (grand)child (F2)?

Step 1: Translate to genotypes (1 point)

Parent 1	
Parent 2	
Child	
Partner	
Grandchild	

Step 2: Enter the genotypes from step 1 in to the family chart. Please use the boxes below the example. For unknown family members you can use a question mark. (1 point)



Step 3: You will solve this question in two steps. For the first set of cross tables, determine the direction of your reasoning. (1 point)

- ☐ Deductive
☐ Inductive

Step 4: Determine the number of cross tables you need and fill them in.

- ☐ 1 cross table
☐ 2 cross tables
☐ 3 cross tables

Step 5: Determine the direction of your reasoning for the second set of cross tables. (1 point)

- ☐ Deductive
☐ Inductive

Step 6: Determine the number of cross tables you need and fill them in.

- ☐ 1 cross table
- ☐ 2 cross tables
- ☐ 3 cross tables
- ☐ 4 cross tables
- ☐ 5 cross tables
- ☐ 6 cross tables

Step 7: Unknown genotype of the child (F1) (1 point)

Compare the outcomes you received from the cross tables and use them to determine your final answer. Mark the right answer below.

- ☐ HH
- ☐ Hh
- ☐ hh
- ☐ HH and Hh
- ☐ HH and hh
- ☐ Hh and hh
- ☐ HH, Hh and hh

Step 8: Unknown genotype of the grandchild (F1) (1 point)

Compare the outcomes you received from the cross tables and use them to determine your final answer. Mark the right answer below.

- ☐ HH
- ☐ Hh
- ☐ hh
- ☐ HH and Hh
- ☐ HH and hh
- ☐ Hh and hh
- ☐ HH, Hh and hh

Appendix E.

Mental Effort Rating Scale Adapted from Paas [32]

How much mental effort did you invest to solve this problem?

	very, very low mental effort			low mental effort			not high and not low mental effort		high mental effort		very, very high mental effort
	1	2	3	4	5	6	7	8	9		
Mental effort	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Appendix F

Appendix F.1. Complexity Levels 1–3 Self-Assessment Measure

You can get a maximum of 4 points for solving the previous problem. How many points do you think you scored?

0 points

1 point

2 points

3 points

4 points

Points Scored

Appendix F.2. Complexity Level 4 Self-Assessment Measure

You can get a maximum of 5 points for solving the previous problem. How many points do you think you scored?

0 points

1 point

2 points

3 points

4 points

5 points

Points Scored

Appendix F.3. Complexity Level 5 Self-Assessment Measure

You can get a maximum of 6 points for solving the previous problem. How many points do you think you scored?

0 points

1 point

2 points

3 points

4 points

5 points

6 points

Points Scored

Appendix G.

Appendix G.1. Instructions for the Problem-Solving Tasks for the Control Group

In the next section, you will be solving 10 problems. When working through each of these problems, please **DO** the following:

1. Try to solve the questions as quickly and accurately as you can.

2. If you absolutely cannot answer a question, please put a question mark (?) as the answer.

Do **NOT**:

1. Do **NOT** use your hands, write things down, or speak when solving the problems.

Please click **yes** if you understand and accept the instructions.

Appendix G.2. Instructions for the Problem-Solving Tasks for the Experimental Group

In the next section, you will be solving 10 problems. When working through each of these problems, please **DO** the following:

1. **Think aloud** (say what you are doing and thinking as you are doing or thinking it) as you work through each of the problems. Please do this **in English**.
2. Try to solve the questions as quickly and accurately as you can.
3. If you absolutely cannot answer a question, please put a question mark (?) as the answer.

Do **NOT**:

1. Do **NOT** write anything down or use your hands when solving the problems.
2. Do **NOT** say any personally identifiable information such as your name or age during the recordings.

If you do not want your voice recorded, you may stop your participation now or later without penalty or justification. Also, a reminder you have the right to ask for your data not to be used.

The audio recording will start automatically, so just follow the think aloud instructions as soon as you start the problems!

Please click yes if you understand and accept the instructions.

☐ Yes



Appendix H.

Coding Scheme for Qualitative Analysis (Based on Azevedo et al. [45])		
Variable	Description	Example
A—Forethought, Planning, and Activation		
A1—Planning	Includes listing the actions needed to solve the problem before executing them.	<i>“I need to do the table first”</i>
A2—Sub-Goals	Participant states actions that are intended to help solve the problem while solving the problem.	<i>“And then so we know that so we can do deductive again”</i>
A3—Prior Knowledge Activation	Participant recalls relevant prior knowledge before a task or during.	<i>“Homozygous means they are both the same gene”</i>
A4—Recycle Goal in Working Memory	Participant repeats the goal verbally to refresh it in their working memory.	<i>“Both parents have brown eyes. . .both parents have brown eyes”</i>

B—Monitoring		
B1—Negative Judgment of Learning (NJOL)	Participant becomes aware that they do not understand or know the material during problem-solving.	<i>"Still not sure"</i> <i>"The child, wait hold on, this is so confusing, what am I doing?"</i>
B2—Positive Judgment of Learning (PJOL)	Participant becomes aware that they do understand or know the material during problem solving.	<i>"Which means yeah that makes sense"</i> <i>"So parent one is big B, big B. Right? Yeah"</i>
B3—Self-assessment	Participant gives a positive or negative self-assessment of their performance (when prompted to during self-assessment rating question)	<i>"I think I scored all of them, probably"</i> <i>"Guessing four points hopefully"</i>
B4—Feeling of Knowing (FOK)	Participant states that they remember the material vaguely and understand some of it, but not well.	<i>"There was another way to do this, but I forgot"</i> <i>"And it can either be both recessive or large, I think that's possible? Yeah"</i>
B5—Content Evaluation	Participant monitors content relative to goals.	<i>"I don't think the child's eye colour was mentioned even"</i> <i>"So the parent has a big B, otherwise they would not had brown eyes"</i>
B6—Identify Adequacy of Information and Task Instructions	Participant determines the usefulness or adequacy of the material and task instructions.	<i>"So I am just going to put information parent two as. . . I don't know how to put that in there, honestly"</i> <i>"I think we are talking about the child here, because that is the only one that is kinda missing and I would technically say it is both. . . oh for the first set of the cross tables, okay okay that makes a lot of sense now."</i>
C—Strategy Use		
C1—Goal-Free Search	Participant states that they will look around the information without stating a particular goal or plan.	[Did not appear in any the transcripts]
C2—Summarization	Participant summarizes what they read.	<i>"The partner has curly hair and is heterozygotic, so they have the same traits as the parents of the child"</i> <i>"Okay, so one parent is homozygous and dominant and the other one is heterozygous"</i>
C3—Rereading	Participants reread or revisit earlier material.	<i>"And then the partner had heterozygotic set, right? Curly hair, yes. And is heterozygotic."</i>
C4—Inferences	Participant makes inferences based on what they read or saw in the environment.	<i>"So both have brown eyes and are homozygous so big B big B."</i> <i>"Deductive reasoning"</i> <i>"Which means the parent can be big little and then little. So then this one"</i>
C5—Hypothesizing	Participant asks questions that go beyond what they read in the material.	<i>"But these are the options that the child definitely does not have, if it is two small h's it would not work. But it is probabilities so who knows."</i> <i>"So we have a 50/50 chance that the child will have homozygous brown eyes or homozygous. . . heterozygous brown eyes."</i>
C6—Knowledge Elaboration	Participant elaborates what they read in the material with prior knowledge.	<i>"Ok so h h. And to have a child that has one big H that means that they have to have at least one big h."</i> <i>"Again this is deductive, because we are looking at the parents and then trying to guess the child one"</i>
D—Task Difficulty, Complexity, and Demands		
D1—Time and Effort Planning	Participant deliberately controls their behavior due to limited energy or other physical demands.	<i>"Just going to say small instead of lowercase or uppercase"</i> <i>"Since question is going to ask I'm just going to fill the first one because I can't"</i>

D2—Help-Seeking Behavior	Participant seeks assistance from the experimenter regarding the task.	[Any dialogues with the experimenters about the task]
D3—Task Difficulty	Participant indicates that the task is easy or difficult.	<i>"It's a brawl of kk, this is easy"</i> <i>"Okay, this is a bit more complex"</i>
D4—Mental Effort	Participant gives an indication of their mental effort invested in the task.	<i>"I guess same mental effort or a bit more maybe"</i> <i>"Say low, probably. But I did think about it"</i>
E—Miscellaneous Actions		
E1—Stating Answer	Participant states an answer to a final step in the problem-solving task.	<i>"All the possibilities"</i>
E2—Copying information	Participant copies information they wrote again or found in the environment or verbalizes information they input into a blank or cross table.	<i>"Parent one is big big, parent two is little little"</i> <i>"So big big little big big and big little"</i>
E3—Reading Question	Participant reads questions or information (close to) verbatim on screen out loud.	<i>"Two parents are having a child."</i>

Note. All italicized examples are direct quotes from the transcripts. All unitalicized examples in brackets are explanations

List of Adjustments from Base Coding Scheme

- Simplified or adjusted definition for this context and changed example to relevant example from the transcripts
 - A1—Planning
 - A3—Prior Knowledge Activation
 - A4—Recycle Goal in Working Memory
 - B4—Feeling of Knowing (FOK)
 - C2—Summarization
 - C3—Re-reading
 - C4—Inferences
 - C5—Hypothesizing
 - C6—Knowledge Elaboration
 - D1—Time and Effort Planning
 - D2—Help-Seeking Behavior
 - D3—Task Difficulty
- Heavily adjusted definition for this context, moved it to a completely new code group, and put relevant example taken from the transcripts
 - E2—Copying information
- Simplified definition for this context and gave a hypothetical example not from the transcripts
 - A2—Sub-Goals
 - C1—Goal-Free Search
- Definition kept the same but relevant example from the transcripts replaced original example
 - B5—Content Evaluation
- Added this code due to frequent occurrences in the text and/or relevancy to research
 - E3—Reading Question
 - D4—Mental Effort
 - E1—Stating Answer
- Originally one code of called judgments of learning, this code was split up into two different codes, opposites of each other, for more in depth analysis of kinds of judgments of learning the participants verbalized. Definitions were adjusted and relevant examples from the transcripts were given.
 - B1—Negative Judgment of Learning (NJOL)
 - B2—Positive Judgment of Learning (PJOL)

- This code was added to the monitoring code group to be able to separate out when participants monitored spontaneously (such as a JOL) and when participants monitored their performance when prompted in the task environment when doing the self-assessment rating scale.
 - B3—Self-assessment
- Originally the two codes of “Identify Adequacy of Information” and “Expectation of Adequacy of Information”, these codes were combined together to form this code, which included judgments based on the usefulness and adequacy of task instructions and material presented, which fit better with this context. Definition was adjusted to fit the new code and relevant examples from the transcripts were added.
 - B6—Identify Adequacy of Information and Task Instructions
- These codes were removed due to irrelevance or redundancy.
 - Self-Questioning
 - Selecting a New Informational Source
 - Control of Context
- Code group called Task Difficulty and Demands was renamed to Task Difficulty, Complexity, and Demands
- Added code group called Miscellaneous Actions due to these codes not fitting in with other code groups well but being frequent types of verbalizations

Appendix I.

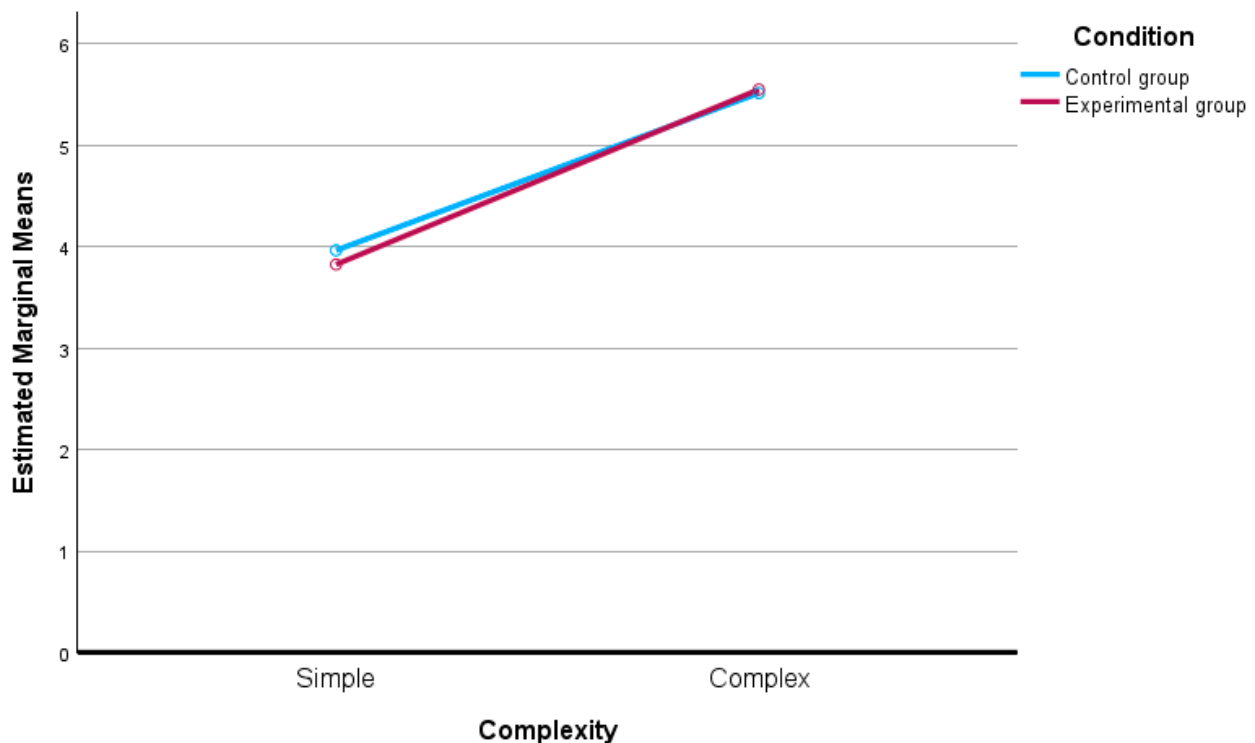


Figure A1. Average mental effort rating per condition for both complexity levels.

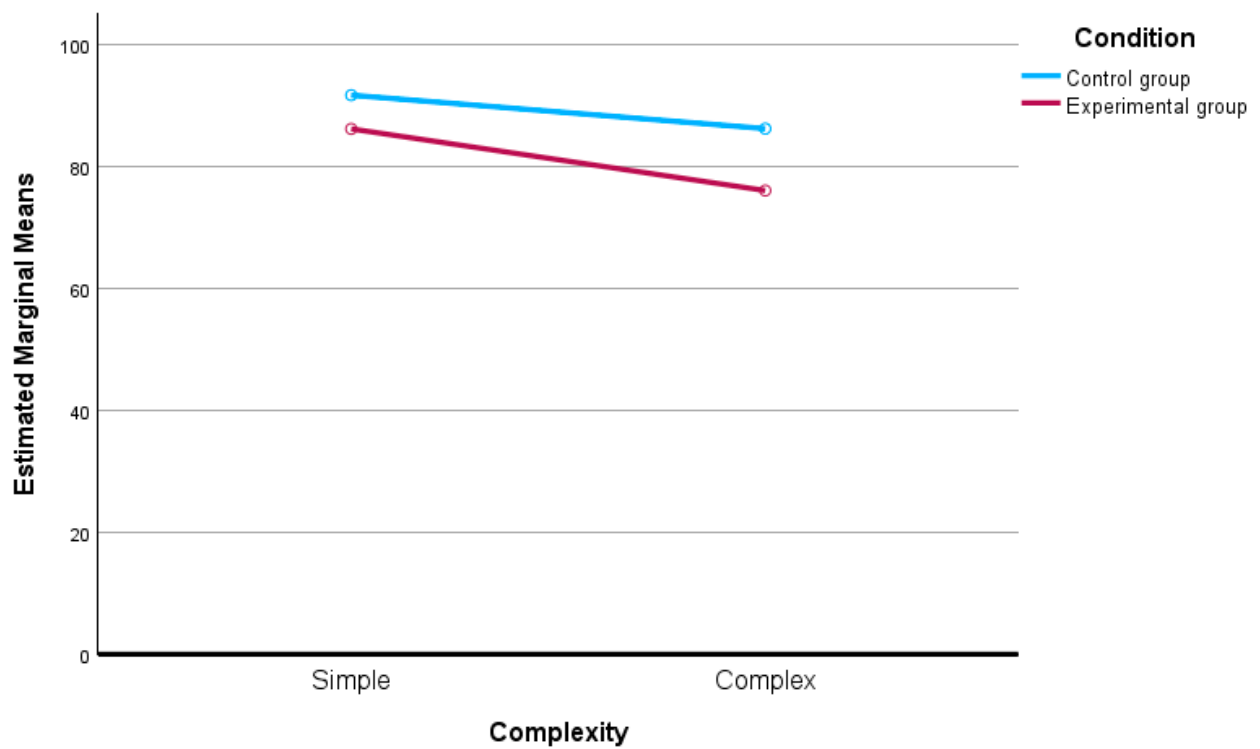


Figure A2. Average performance score per condition for both complexity levels.

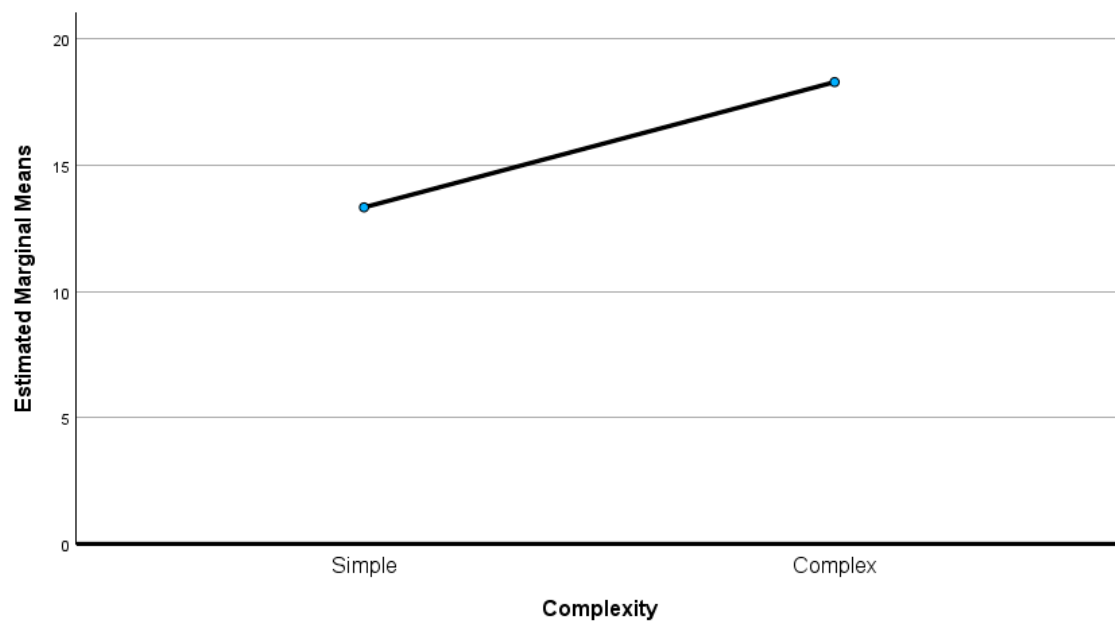


Figure A3. Average absolute self-assessment accuracy score of the experimental group for both complexity levels.

Appendix J.*Pairwise Comparisons for Average Performance per Complexity Level*

Complexity (I)	Complexity (J)	Mean Difference (I–J)	Standard Error	Sig.	95% CI for Difference	
					Lower Bound	Upper Bound
1	2	1.310	2.299	1.000	–5.380	7.999
	3	7.530	2.572	0.047	0.047	15.012
	4	9.292	2.757	0.013	1.272	17.311
	5	8.562	2.520	0.012	1.230	15.893
2	1	–1.310	2.299	1.000	–7.999	5.380
	3	6.220	1.919	0.019	0.637	11.803
	4	7.982	2.264	0.008	1.396	14.568
	5	7.252	2.106	0.010	1.126	13.378
3	1	–7.530	2.572	0.047	–15.012	–0.047
	2	–6.220	1.919	0.019	–11.803	–0.637
	4	1.762	2.120	1.000	–4.407	7.930
	5	1.032	1.983	1.000	–4.739	6.802
4	1	–9.292	2.757	0.013	–17.311	–1.272
	2	–7.982	2.264	0.008	–14.568	–1.396
	3	–1.762	2.120	1.000	–7.930	4.407
	5	–0.730	1.994	1.000	–6.531	5.071
5	1	–8.562	2.520	0.012	–15.893	–1.230
	2	–7.252	2.106	0.010	–13.378	–1.126
	3	–1.032	1.983	1.000	–6.802	4.739
	4	0.730	1.994	1.000	–5.071	6.531

Note: For the performance measure, complexity levels one and two are significantly different from levels three through five, but they are similar to each other. Meanwhile, complexity levels three through five also are statistically similar to each other. Therefore, for the greater analysis, we decided to average together the performances on levels one and two for a simple performance mean and then averaged levels three through five together to obtain a complex performance mean. This was similarly done with mental effort to obtain a simple and complex mental effort mean and absolute SA accuracy to obtain a simple and complex absolute SA accuracy mean.

References

1. Bjork, R.A.; Dunlosky, J.; Kornell, N. Self-regulated learning: Beliefs, techniques, and illusions. *Annu. Rev. Psychol.* **2013**, *64*, 417–444. [CrossRef] [PubMed]
2. Zimmerman, B.J. Becoming a self-regulated learner: Which are the key subprocesses? *Contemp. Educ. Psychol.* **1986**, *11*, 307–313. [CrossRef]
3. Stefanou, C.; Stolk, J.D.; Prince, M.; Chen, J.C.; Lord, S.M. Self-regulation and autonomy in problem- and project-based learning environments. *Act. Learn. High. Educ.* **2013**, *14*, 109–122. [CrossRef]
4. Lai, C.; Hwang, G. A self-regulated flipped classroom approach to improving students' learning performance in a mathematics course. *Comput. Educ.* **2016**, *100*, 126–140. [CrossRef]
5. Littlejohn, A.; Hood, N.; Milligan, C.; Mustain, P. Learning in MOOCs: Motivations and self-regulated learning in MOOCs. *Internet High. Educ.* **2016**, *29*, 40–48. [CrossRef]
6. Winne, P.H. A metacognitive view of individual differences in self-regulated learning. *Learn. Individ. Differ.* **1996**, *8*, 327–353. [CrossRef]
7. Azevedo, R.; Cromley, J. Does training on self-regulated learning facilitate students' learning with hypermedia? *J. Educ. Psychol.* **2004**, *96*, 523–535. [CrossRef]
8. Baars, M.; Visser, S.; Van Gog, T.; de Bruin, A.; Paas, F. Completion of partially worked-out examples as a generation strategy for improving monitoring accuracy. *Contemp. Educ. Psychol.* **2013**, *38*, 395–406. [CrossRef]
9. Baars, M.; Van Gog, T.; de Bruin, A.; Paas, F. Effects of problem solving after worked example study on primary school children's monitoring accuracy. *Appl. Cogn. Psychol.* **2014**, *28*, 382–391. [CrossRef]
10. Baars, M.; Vink, S.; Van Gog, T.; De Bruin, A.; Paas, F. Effects of training self-assessment and using assessment standards on retrospective and prospective monitoring of problem solving. *Learn. Instr.* **2014**, *33*, 92–107. [CrossRef]

11. Van Gog, T.; Kester, L.; Paas, F. Effects of concurrent monitoring on cognitive load and performance as a function of task complexity. *Appl. Cogn. Psychol.* **2011**, *25*, 584–587. [CrossRef]
12. De Bruin, A.B.; Roelle, J.; Carpenter, S.K.; Baars, M.; EFG-MRE. Synthesizing cognitive load and self-regulation theory: A theoretical framework and research agenda. *Educ. Psychol. Rev.* **2020**, *32*, 903–915. [CrossRef]
13. Sweller, J.; Van Merriënboer, J.J.G.; Paas, F.G.W.C. Cognitive architecture and instructional design. *Educ. Psychol. Rev.* **1998**, *10*, 251–296. [CrossRef]
14. Sweller, J.; Van Merriënboer, J.J.G.; Paas, F.G.W.C. Cognitive architecture and instructional design: 20 years later. *Educ. Psychol. Rev.* **2019**, *31*, 261–292. [CrossRef]
15. Baddeley, A. Working memory. *Science* **1992**, *255*, 556–559. [CrossRef]
16. Paas, F.G.W.C.; Van Merriënboer, J.J.G. Instructional-control of cognitive load in the training of complex cognitive tasks. *Educ. Psychol. Rev.* **1994**, *6*, 351–371. [CrossRef]
17. Brünken, R.; Plass, J.L.; Leutner, D. Direct measurement of cognitive load in multimedia learning. *Educ. Psychol.* **2003**, *38*, 53–61. [CrossRef]
18. Paas, F.; Van Gog, T.; Sweller, J. Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. *Educ. Psychol. Rev.* **2010**, *22*, 115–121. [CrossRef]
19. Radović, S.; Seidel, N.; Haake, J.M.; Kasakowskij, R. Analyzing students' self-assessment practice in a distance education environment: Student behavior, accuracy, and task-related characteristics. *J. Comput. Assist. Learn.* **2023**, *40*, 654–666. [CrossRef]
20. Pintrich, P.R. The role of goal orientation in self-regulated learning. In *Handbook of Self-Regulation*; Boekaerts, M., Pintrich, P.R., Zeidner, M., Eds.; Academic Press: Cambridge, MA, USA, 2000; pp. 451–502. [CrossRef]
21. Winne, P.H.; Hadwin, A.F. Studying as self-regulated learning. In *Metacognition in Educational Theory and Practice*, 2nd ed.; Hacker, D., Dunlosky, J., Graesser, A., Eds.; Lawrence Erlbaum: Mahwah, NJ, USA, 1998; pp. 277–304. [CrossRef]
22. Panadero, E. A review of self-regulated learning: Six models and four directions for research. *Front. Psychol.* **2017**, *8*, 422. [CrossRef]
23. Winne, P.H. Cognition and metacognition within self-regulated learning. In *Handbook of Self-Regulation of Learning and Performance*, 2nd ed.; Schunk, D.H., Greene, J.A., Eds.; Routledge: London, UK, 2017; pp. 36–48. [CrossRef]
24. Ku, K.Y.L.; Ho, I.T. Metacognitive strategies that enhance critical thinking. *Metacogn. Learn.* **2010**, *5*, 251–267. [CrossRef]
25. Mirandola, C.; Ciriello, A.; Gigli, M.; Cornoldi, C. Metacognitive monitoring of text comprehension: An investigation on postdictive judgments in typically developing children and children with reading comprehension difficulties. *Front. Psychol.* **2018**, *9*, 2253. [CrossRef] [PubMed]
26. Barenberg, J.; Dutke, S. Testing and metacognition: Retrieval practise effects on metacognitive monitoring in learning from text. *Memory* **2019**, *27*, 269–279. [CrossRef] [PubMed]
27. Dunlosky, J.; Rawson, K.A. Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learn. Instr.* **2012**, *22*, 271–280. [CrossRef]
28. Pieschl, S. Metacognitive calibration—An extended conceptualization and potential applications. *Metacogn. Learn.* **2009**, *4*, 3–31. [CrossRef]
29. Peng, Y.; Tullis, J.G. Dividing attention impairs metacognitive control more than monitoring. *Psychon. Bull. Rev.* **2021**, *28*, 2064–2074. [CrossRef]
30. Tuysuzoglu, B.B.; Greene, J.A. An investigation of the role of contingent metacognitive behavior in self-regulated learning. *Metacogn. Learn.* **2015**, *10*, 77–98. [CrossRef]
31. Seufert, T. The interplay between self-regulation in learning and cognitive load. *Educ. Res. Rev.* **2018**, *24*, 116–129. [CrossRef]
32. Paas, F.G.W.C. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *J. Educ. Psychol.* **1992**, *84*, 429–434. [CrossRef]
33. Winne, P.H. Inherent details in self-regulated learning. *Educ. Psychol.* **1995**, *30*, 173–187. [CrossRef]
34. Vangsnæs, L.; Young, M.E. More isn't always better: When metacognitive prompts are misleading. *Metacogn. Learn.* **2021**, *16*, 135–156. [CrossRef]
35. Liefvooghe, B.; Barrouillet, P.; Vandierendonck, A.; Camos, V. Working memory costs of task switching. *J. Exp. Psychol. Learn. Mem. Cogn.* **2008**, *34*, 478–494. [CrossRef] [PubMed]
36. Patterson, M.C. A naturalistic investigation of media multitasking while studying and the effects on exam performance. *Teach. Psychol.* **2017**, *44*, 51–57. [CrossRef]
37. Finley, J.R.; Benjamin, A.S.; McCarley, J.S. Metacognition of multitasking: How well do we predict the costs of divided attention. *J. Exp. Psychol. Appl.* **2014**, *20*, 158–165. [CrossRef]
38. Trabasso, T.; Magliano, J.P. Conscious understanding during comprehension. *Discourse Process.* **1996**, *21*, 255–287. [CrossRef]
39. Meyers, J.; Lytle, S.; Palladino, D.; Devenpeck, G.; Green, M. Think-aloud protocol analysis: An investigation of reading comprehension strategies in fourth- and fifth-grade students. *J. Psychoeduc. Assess.* **1990**, *8*, 112–127. [CrossRef]
40. Schellings, G.L.M.; Van Hout-Wolters, B.H.A.M.; Veenman, M.V.J.; Meijer, J. Assessing metacognitive activities: The in-depth comparison of a task-specific questionnaire with think-aloud protocols. *Eur. J. Psychol. Educ.* **2013**, *28*, 963–990. [CrossRef]
41. Zachariou, A.; Whitebread, D. Developmental differences in young children's self-regulation. *J. Appl. Dev. Psychol.* **2019**, *62*, 282–293. [CrossRef]

42. Jordano, M.L.; Touron, D.R. How often are thoughts metacognitive? Findings from research on self-regulated learning, think-aloud protocols, and mind-wandering. *Psychon. Bull. Rev.* **2018**, *25*, 1269–1286. [CrossRef]
43. Qualtrics XM—Experience Management Software. Available online: <https://www.qualtrics.com/> (accessed on 5 May 2023).
44. Moos, D.C.; Bonde, C. Flipping the classroom: Embedding self-regulated learning prompts in videos. *Technol. Knowl. Learn.* **2016**, *21*, 225–242. [CrossRef]
45. Azevedo, R.; Guthrie, J.T.; Seibert, D. The role of self-regulated learning in fostering students' conceptual understanding of complex systems with hypermedia. *J. Educ. Comput. Res.* **2004**, *30*, 87–111. [CrossRef]
46. Johnstone, C.J.; Bottsford-Miller, N.A.; Thompson, S.J.; Council of Chief State School Officers (CCSSO); National Association of State Directors of Special Education (NASDSE). *Using the Think Aloud Method (Cognitive Labs) to Evaluate Test Design for Students with Disabilities and English Language Learners*; University of Minnesota, National Center on Educational Outcomes: Minneapolis, MN, USA, 2006.
47. ATLAS.ti Scientific Software Development GmbH. ATLAS.ti Windows (Version 24.0.0) [Qualitative Data Analysis Software]. 2024. Available online: <https://atlasti.com> (accessed on 1 January 2024).
48. IBM Corp. *IBM SPSS Statistics for Windows, version 29.0*; Computer software; IBM Corp.: Armonk, NY, USA, 2023.
49. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1988.
50. Leppink, J.; Paas, F.; Van der Vleuten, C.P.M.; Van Gog, T.; Van Merriënboer, J.J.G. Development of an instrument for measuring different types of cognitive load. *Behav. Res. Methods* **2013**, *45*, 1058–1072. [CrossRef] [PubMed]
51. Hoch, E.; Sidi, Y.; Ackerman, R.; Hoogerheide, V.; Scheiter, K. Comparing mental effort, difficulty, and confidence appraisals in problem-solving: A metacognitive perspective. *Educ. Psychol. Rev.* **2023**, *35*, 61. [CrossRef]
52. Albulescu, P.; Macsinga, I.; Rusu, A.; Sulea, C.; Bodnaru, A.; Tulbure, B.T. “Give me a break!” A systematic review and meta-analysis on the efficacy of micro-breaks for increasing well-being and performance. *PLoS ONE* **2022**, *17*, e0272460. [CrossRef] [PubMed]
53. Tyler, J.M.; Burns, K.C. After depletion: The replenishment of the self's regulatory resources. *Self Identity* **2008**, *7*, 305–321. [CrossRef]
54. Steinborn, M.B.; Huestegge, L. A walk down the lane gives wings to your brain. Restorative benefits of rest breaks on cognition and self-control. *Appl. Cogn. Psychol.* **2016**, *30*, 795–805. [CrossRef]
55. Nietfeld, J.L.; Cao, L.; Osborne, J.W. Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *J. Exp. Educ.* **2005**, *74*, 7–28.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

The Interplay of Self-Regulated Learning, Cognitive Load, and Performance in Learner-Controlled Environments

Anna Gorbunova¹, Christopher Lange², Alexander Savelyev³, Kseniia Adamovich¹ and Jamie Costley^{4,*}

¹ Institute of Education, HSE University, 101000 Moscow, Russia; agorbunova@hse.ru (A.G.); kadamovich@hse.ru (K.A.)

² Department of British and American Humanities, Dankook University, Yongin City 16890, Republic of Korea; christopherhlange@dankook.ac.kr

³ Faculty of Law, HSE University, 101000 Moscow, Russia; asavelev@hse.ru

⁴ College of Education, United Arab Emirates University, Al Ain 15551, United Arab Emirates

* Correspondence: jcostley@uaeu.ac.ae

Abstract: Learner control allows for greater autonomy and is supposed to benefit learning motivation, but it might be more advantageous for students with specific learner characteristics. The current study looks into the relationships between self-regulated learning, cognitive load, and performance within learner-controlled environments. The research was conducted in an asynchronous online setting, allowing for learner control. Cognitive load and self-regulated learning were measured using self-report questionnaires. Performance was assessed through case solutions. The participants were 97 graduate law students studying the civil code. Analysis based on structural equation modeling showed that both prior knowledge and self-regulated learning skills significantly contribute to the increase in germane cognitive load and are positively correlated with performance. The implications of these findings underscore the critical role of prior knowledge and self-regulated learning skills in shaping the cognitive processes involved in learning, ultimately impacting academic achievement. These results emphasize the need for careful consideration of learner-control options in asynchronous online environments.

Keywords: cognitive load theory; self-regulated learning; learner control; academic performance; prior knowledge

Citation: Gorbunova, A.; Lange, C.; Savelyev, A.; Adamovich, K.; Costley, J. The Interplay of Self-Regulated Learning, Cognitive Load, and Performance in Learner-Controlled Environments. *Educ. Sci.* **2024**, *14*, 860. <https://doi.org/10.3390/educsci14080860>

Academic Editor: Mido Chang

Received: 2 May 2024

Revised: 29 July 2024

Accepted: 31 July 2024

Published: 8 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recent advancements in computer-supported and digital learning, especially in asynchronous formats, have led to the increasing popularity of learner-controlled settings in which students are provided with opportunities to select the activities, the order, and the pace in which they prefer to interact with the content [1]. It is assumed that learner control will lead to several beneficial educational results, such as an increase in motivation [2], customized instruction that caters to individual needs and interests, the creation of effective teaching methods, and a sense of personal control [3]. However, learner control does not always result in enhanced learning outcomes [4]. Moreover, previous research shows that it may only be advantageous for learners with certain individual characteristics or when used under specific circumstances, for example, learners with high prior knowledge [3] or conditions in which it does not generate excessive cognitive load [5].

The autonomous nature of asynchronous online learning environments requires a certain level of self-regulated learning skills, such as setting goals, planning, and monitoring [6,7]. Self-regulated learning is particularly significant in learner-controlled environments, as these environments require that students realize their limitations and have the ability to organize their learning in such a way as to meet them [8]. It is, therefore, possible that this self-regulatory challenge causes problems for students [9]. If self-regulated learning consumes all of the students' mental resources, learning might be hampered, as

no resources would be available for processing new information in their working memory and transferring it to long-term memory.

One way of understanding how students use their mental resources in a learner-controlled environment is with cognitive load theory. This theory has been used as a framework to understand what is happening in learner-controlled environments [2,4,9–11]. Cognitive load represents information to be processed given the limited amount of available cognitive resources [12]. The objective of cognitive load theory is to explain how the cognitive load imposed by learning can impact students' capacity to process new information and build knowledge by transferring information from working memory to long-term memory [13]. The theory posits that learning is optimized when the instructional design takes into account the limitations of working memory and reduces the need to process information not relevant to learning, thereby promoting schema acquisition [14]. Originally, researchers classified cognitive load into three types: intrinsic load, which can be broadly defined as the complexity of the content due to the number of interacting elements combined with learners' prior knowledge; extraneous load, which can be broadly defined as poorly designed instruction that induces learners to process information not relevant to learning the task; and germane load, which can be broadly defined as the availability of cognitive resources directed at dealing with intrinsic load, with instructional interventions aimed at minimizing extraneous load and enhancing germane [12]. Cognitive load theory aims to provide recommendations to researchers and practitioners on how to design efficient instruction and facilitate learning [15].

There is evidence that the use of learner-control options can require additional cognitive resources [2], and specific cognitive load elements (intrinsic load, extraneous load, and germane load) are affected by the learning task and the type of learner-control options used to complete the learning task [2,11]. The goal from a cognitive load perspective is to help learners deal with intrinsic load and mitigate excessive extraneous load so that germane load can be promoted [13], which previous research suggests is possible in learner-controlled environments if the learner-control options allow learners to more efficiently process information required to complete a task [11]. The investigation of effects on cognitive load in learner-controlled settings ultimately contributes to the better instructional design of these environments [4].

2. Literature Review

2.1. Self-Regulated Learning in Learner-Controlled Environments

Self-regulated learning refers to the ability of students to plan their learning activities and monitor, control, and assess their progress in order to apply the most efficient learning strategies and modify them as needed [7]. Pintrich and de Groot (1990) [16] consider that self-regulated learning has at least three major components: metacognitive strategies, students' control of the effort to deal with learning tasks, and actual cognitive strategies to learn the material. Self-regulation, in this specific context, refers to the deliberate and strategic efforts undertaken by students to manage their learning; however, it is only one of the components of self-regulated learning.

Research indicates that students who exhibit strong self-regulatory skills tend to resort to more sophisticated cognitive strategies, allowing them to demonstrate better academic performance [16]. As a result, understanding and promoting self-regulated learning has become increasingly important in educational settings to equip students with the necessary skills to effectively navigate their learning experiences. Self-regulated learning is assumed to be crucial for online environments due to the level of autonomy associated with online learning [6], particularly for learner-controlled environments [8], as it enables learners to navigate the learning materials effectively and manage their cognitive load. Moreover, it is suggested that learner control can have a significant impact on the quality of online learning [17].

It is also important to note that the self-regulated learning strategies themselves can, at times, help students avoid extraneous processing, as such strategies are used to select content in a way that prevents them from processing extraneous information [18]. Another

characteristic is prior knowledge, which plays a significant role, as novice learners cannot process new information and benefit from learning strategies at the same time [10]. Nevertheless, distributing limited mental resources between learning and using self-regulated learning skills might create a challenge [19]. Previous research suggests that self-regulated learning predominantly hinges on working memory processes when learners lack sufficient prior knowledge and have low levels of self-regulated learning skills [20], creating cognitive overload.

2.2. Cognitive Load in Learner-Controlled Environments

One of the factors that has a positive impact on cognitive load while self-regulating in learner-controlled environments is prior knowledge, as learners' familiarity with the material makes these complex environments more comfortable for them. Novice learners cannot process new information and benefit from learning strategies at the same time [10]. Previous research has shown that learners with high domain knowledge will benefit more in a learner-controlled environment than learners with low prior knowledge [3], which is most likely due to the fact that learners with low prior knowledge have to learn new information and make choices about how to proceed with their instruction at the same time [4]. Moreover, it is suggested that compared to learner-controlled environments, instructor-controlled environments may be more beneficial for novice learners due to the instructional guidance received when implementing solutions [10].

The nature of intrinsic cognitive load lies in the difficulty of particular content in terms of the number of interacting elements for learners, taking into account their domain prior knowledge [13]. In general, it is assumed that intrinsic load is lower for learners with higher levels of prior knowledge as, for an expert learner, this material would be easier to interact with than for a novice learner [21]. In the case of learner-controlled environments, this effect could be exaggerated: intrinsic load might be lower for learners with more prior knowledge due to the way in which they can effectively manipulate the content to be learned [2]. Learner-controlled environments allow learners to control the amount of content processed at a single time through the selection of content and the way in which they process it. For example, learners can select the content to be learned and then pause or segment that content, effectively breaking down element interactivity so that they can process isolated elements of the content before moving on [18,22]. While the breakdown of element interactivity does not directly reduce intrinsic load when confronted with the entire task [23], the intrinsic load is lower during the learning of the task due to the breakdown of element interactivity [24]. From this perspective, selecting the way in which to process the content may allow learners to control the number of interacting elements processed at a single time, thus processing content with lower levels of intrinsic load while learning a specific task. This is particularly advantageous for learners with higher prior knowledge in that using learner control to isolate elements of content so that it matches one's domain knowledge takes a higher level of expertise in the content itself [5]. Furthermore, such an approach to information processing requires the use of self-regulated learning skills as highly self-regulated learners are able to self-manage their load, optimally processing suboptimal instruction as well as managing the number of elements interacting at a specific time [25]. Due to the nature of intrinsic load, a negative relationship between high levels of intrinsic cognitive load and performance might be expected; however, high levels of prior knowledge might mitigate this effect.

According to cognitive load theory, extraneous load occurs because of poorly designed instruction, causing learners to spend more of their mental resources on processing information that is irrelevant to learning [26]. It is expected that expert learners having constructed schemata stored in their long-term memory are less constrained in working memory resources and are thus better able to overcome extraneous load or self-manage it [27]. With that being said, an expertise reversal effect can occur if expert learners waste an exorbitant amount of cognitive resources on information they already know, creating higher levels of extraneous load [28]. However, those with expertise in the subject matter should

be more likely to avoid such a situation in learner-controlled environments, as it is the expertise of learners that allows them to alter the element interactivity by controlling how they process content to match their domain knowledge [5], thus avoiding the processing of redundant information and ensuring that their cognitive resources are directed toward processing information relevant to the task [28]. Additionally, one noteworthy point is that according to the element interactivity effect [29], when instructional materials consist of multiple elements that interact with one another, extraneous load can increase due to poor instructional presentation, creating the inability to manage high levels of intrinsic load, which has implications for both working memory and long-term memory regardless of learners' prior knowledge [21]. There is an assumption that extraneous cognitive load leads to poor performance, as learners' working memory resources are distracted and allocated to overcoming badly designed instruction. Learners have to process unnecessary information, increasing the element interactivity, which impedes learning and results in unsatisfactory learning outcomes [15].

The conceptualization of germane load has changed over time. At one point, it was understood as mental effort directed to the processes that contribute to learning [30]. However, it has since been reconceptualized as the availability of mental resources allocated to manage intrinsic load [12]. This is due to studies that have shown that mental effort does not add to the internal consistency of the load [12]. Furthermore, while it has been described as no longer being an independent source of load due to its association with intrinsic load, it has further been postulated that it may be more accurately described as a representation of understanding [12].

Previous studies [11] suggest that in a learner-controlled environment, students' understanding is enhanced as they engage their mental resources while deliberating on their choices, ultimately contributing to the learning experience. Learner control also contributes to the deeper processing of information, given the availability of cognitive resources, as it requires students to mentally reorganize information [2]. Additionally, learners' characteristics ultimately determine their ability to understand the content, as highly motivated learners are more likely to invest available resources into learning [31]. For example, prior knowledge is viewed as an important individual characteristic of learners; it changes the way learners process information, and it is assumed that higher levels of knowledge provide them with abilities to invest more resources in learning [21]. Thus, expert learners are more likely to change the way in which they process information by investing more cognitive resources directed to selecting appropriate content to match their domain knowledge within learner-controlled environments [5]. Moreover, prior knowledge can assist learners in making strategic choices and thus facilitate their learning in learner-controlled environments [1]. Learner-controlled environments promote increased mental processing of the content [32], which is germane to learning and thus results in better performance. Previous research has shown that in a learner-controlled environment, students perform better in high element interactivity material [11], as germane load is fostered, ensuring better results.

2.3. Self-Regulated Learning and Cognitive Load

Learners' level of self-regulated learning allows them to overcome challenges, as by employing self-regulated learning skills, learners are empowered to identify and address the intrinsic complexity of the material [33]. Furthermore, there is a relationship between self-regulated learning and the use of learner-control options available in asynchronous online environments [34]. It may be that self-regulated learners experience less intrinsic load because they are more likely to use learner-control tools, allowing them to break up element interactivity [24]. It has been previously suggested that learners self-manage their cognitive load, guiding their cognitive resources to the most effective use, enabling them to extract maximum learning outcomes [25]. Additionally, it has been found that self-regulation used in learner-controlled environments allowed learners to overcome high levels of intrinsic load and produce higher levels of germane load, thus indicating that

even if the learners do not necessarily reduce the element interactivity, other self-regulated learning strategies may help them overcome the high levels of intrinsic load without reducing the load [35]. Furthermore, it has been suggested that when learners make judgments of their effort based on intrinsic cognitive load, they can be more motivated to use self-regulated learning strategies [36].

Extraneous load should be a primary concern of instructional design [21]; it is the aspect of cognitive load that researchers and practitioners need to focus on in order to improve the learning experience and learning outcomes. On the other hand, it is possible that even when the instruction is poorly designed, learners might achieve acceptable results. It provokes the question: could there be certain individual characteristics that allow learners to perform even under unfavorable circumstances? According to previous research, the level of self-regulated learning skills equips learners with a variety of metacognitive strategies, such as planning, monitoring, and controlling their learning process, which help learners overcome challenging learning environments [2]. However, if extraneous load is a base for making monitoring judgments, it increases invested effort without adding to learning performance [36]. Although self-regulated learning skills are used to overcome challenging environments, a study by Lange et al. [37] showed that learner-controlled asynchronous environments containing high levels of extraneous load produced a weaker relationship between self-regulated effort and germane load, while the relationship was much stronger in low extraneous load learner-controlled environments. The same study showed a negative relationship between germane load and extraneous load and a positive relationship between self-regulated effort and germane load, as expected. While these relationships were expected, it was somewhat surprising to see that self-regulated effort had a much weaker effect during poorly designed instructional situations within a learner-controlled environment.

Learner control presumes that students take specific actions requiring self-regulated learning skills, which, in turn, are expected to lead to the availability of more germane resources that are beneficial for learning [1]. Germane cognitive load represents the availability of resources that can be directed toward intrinsic load, and self-regulated learners tend to take advantage of the resources that foster greater understanding [38]. It is no surprise, therefore, that highly self-regulated learners are better able to direct their resources more efficiently and be more productive. Previous research has shown that in learner-controlled environments, students have to apply self-regulated learning skills, i.e., monitoring, self-assessing their progress, and making their choices based on that assessment [39]. This is due to the fact that the effort associated with self-regulated learning is often context-dependent [40], and learners generally apply effort at higher levels in environments in which they believe they can succeed [41]. This is evident in learner-control asynchronous online environments containing more autonomy, as self-regulated learners are more likely to apply more effort in such environments because they believe that taking advantage of the autonomous nature will lead to success [42]. Empirical evidence supports the notion that self-regulated learning leads to a better understanding of the content through better planning and effort in the use of learner-control options in multimedia online environments [43,44].

3. The Current Study

Previous research lacks clarity in describing the interaction between prior knowledge, self-regulated learning skills, and cognitive load in learner-controlled environments. While there is empirical evidence that learner control might induce additional cognitive load and that students with higher self-regulated learning skills perform better in learner-controlled settings, the nature of relationships between these constructs remains unexplored. The current research is thus aimed at investigating the fundamental characteristics of connections that emerge between prior knowledge, different types of cognitive load, and self-regulated learning skills and how they might impact performance in a learner-controlled environment. The present study attempts to answer the following questions and test the associated hypotheses:

Q1. *What is the relationship between prior knowledge and the differing aspects of cognitive load in a learner-controlled environment?*

H1a. *Prior knowledge has a negative relationship with intrinsic load.*

H1b. *Prior knowledge has a negative relationship with extraneous load.*

H1c. *Prior knowledge has a positive relationship with germane load.*

Q2. *What is the relationship between self-regulation and the differing aspects of cognitive load in a learner-controlled environment?*

H2a. *Self-regulation has a negative relationship with intrinsic load.*

H2b. *Self-regulation has a negative relationship with extraneous load.*

H2c. *Self-regulation has a positive relationship with germane load.*

Q3. *What is the relationship between the differing aspects of cognitive load and performance in a learner-controlled environment?*

H3a. *Intrinsic load has a negative relationship with performance.*

H3b. *Extraneous load has a negative relationship with performance.*

H3c. *Germane load has a positive relationship with performance.*

4. Materials and Methods

The primary focus of this research is to examine the relationship between students' prior knowledge, different types of cognitive load, and their academic performance in a learner-controlled online environment. The study was conducted with an introductory class of a course on Data Protection Law and designed for university students who take the course during the first and second years of their master's degree in law.

4.1. Participants

The participants gave their consent to participate in the research. They were informed that the performance assessment, which accounted for 20% of their final course grade, was a component of the study. The participants were studying at a large urban university in Russia. Although 108 students were enrolled in the course, only 97 attended the class, and 1 student decided not to answer the questionnaires. With an average age of 23, the group consisted of 56 females and 41 males. All students enrolled in the course had acquired a bachelor's degree in law, which provided them with a deep understanding of the core legal concepts of the Civil Code, which is fundamental for learning data protection law. Such a sampling approach helped to ensure approximately an equal level of knowledge in the field of legal studies among the respondents.

4.2. Materials

Cognitive load was assessed using a student self-report survey (10 items using an 11-point Likert scale). The survey used a tool developed by Leppink et al., [30] with all items translated into Russian with respect to commonly used vocabulary for such learning situations. The items measuring cognitive load are listed in Appendix B. This survey aims to measure the three types of cognitive load: intrinsic, extraneous, and germane cognitive

load. The internal consistency of the three types of cognitive load was estimated using Cronbach's alpha. Cronbach's alpha for the three types of cognitive load are intrinsic load (0.865), extraneous load (0.782), and germane load (0.913) (Table 1).

Self-regulated learning skills consist of several components, including self-regulation comprising metacognitive strategies and effort management. The latter was measured using an adaptation of the Motivated Strategies for Learning Questionnaire (MSLQ) developed by Pintrich and de Groot (1990) [16] and widely validated in the previous literature, consisting of 5 scales (44 items using a 7-point Likert scale). This survey is intended to measure the types of learning strategies and academic motivation used by college students; however, for the purpose of the current study, only one scale, "Self-regulation", was used, consisting of 6 items. Self-regulation, in this context, can be interpreted as the cognitive processes that students engage in to monitor and control their learning, thus generating a regulatory effort that is linked to mental effort and cognitive load. This scale, as well as its title, is referred to throughout the study; however, it is only one of the components of self-regulated learning. The items included in the scale are listed in Appendix A.

Table 1. Mean inter-item correlation for self-regulation and cognitive load scales.

Statistic	N Items	Mean Inter-Item Correlation	Cronbach's Alpha
Self-regulation	6	0.17	0.54
ICL	3	0.68	0.87
GCL	4	0.73	0.91
ECL	3	0.55	0.78

The learning and assessment materials were developed for this research based on the previous study of Costley et al. (2024) [45]. The study compared a problem-solving first sequence to a direct-instruction first instructional sequence in legal sciences. The materials developed for the current study have been further fine-tuned and adapted for asynchronous learning environments. The materials included a test on prior knowledge that included 10 multiple-choice questions covering various aspects of civil law that are relevant to the topics being studied. Three subtopics comprising four types of materials each were covered: a video lecture providing explicit instruction on the topic; a video presenting a worked example illustrating a step-by-step procedure on how the broadly defined legislative interpretation of personal data can be implemented for a particular situation and category of data; a problem-solving task in the form of a similar legal case; and finally the solution to this legal case that follows the same step-by-step procedure. This multidimensionality resulted in a relatively low Cronbach's alpha (0.42), with the mean inter-item correlation equaling 0.082, which is a common problem of wide-ranging conceptual knowledge tests [46].

Academic performance was assessed through a set of legal cases. Students were required to give a well-reasoned legal analysis for each case, determining whether certain data should be classified as personal or not, especially when the law is unclear or lacks specific guidance on the matter. While solving each case, students had to give the correct solution, implement the correct legal terminology, and use reasoning based on the existing legislature. Learning outcomes were evaluated in three dimensions: solution (whether the data was qualified correctly—2 points), argumentation (how substantive were the arguments that supported the solution—3 points), and terminology (whether the legal terminology was appropriately applied—1 point). These dimensions represent important skills for legal students [45]. The maximum score was 36 for all 6 cases for all three dimensions. Both the instructor and his teaching assistant graded all documents for reliability. If there was any discrepancy in scores, the work was discussed until a score was agreed upon. During the assessment, learners had no access to other materials.

4.3. Procedures

While the course took place offline and comprised standard lectures and seminars in smaller groups, the class for the research was organized in an online educational setting that would allow for learner control. The class aimed to explore the concept of personal data from a legal standpoint. While it was important to equip students with a fundamental comprehension of the types of information classified as personal data in accordance with legal regulations, considering regulatory perspectives and real-world legal applications, it was also essential to teach students to distinguish whether specific information qualifies as personal data when not directly stated in the law.

Pre-recorded learning materials and surveys were uploaded to the website specifically designed for the current research. The full illustration of all the materials is presented in Figure 1. Participants were instructed that the website interface and organization of the learning materials offer flexibility to select the preferred methods of engaging with the content. Students could take actions that they considered important to learn the new information and pass the final assessment within the time limit of 2,5 h. Thus, the learner control in this study included watching, rewinding, fast-forwarding, and pausing the pace of the videos with explicit instructions and worked examples; attempting to solve the cases representing problem-solving tasks; downloading the canonical solution to the case; and choosing the order of all materials.

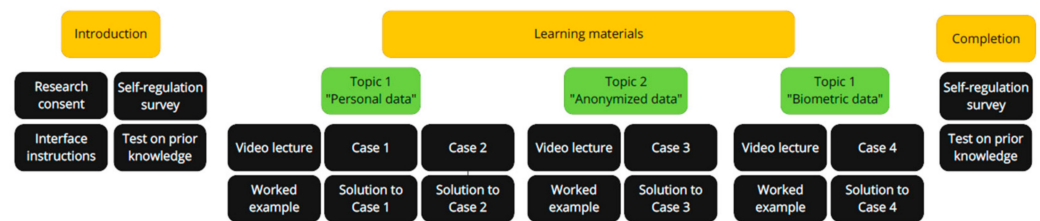


Figure 1. The organization of the learning materials.

5. Results

The means and standard deviations of the main variables of self-regulation, prior knowledge, intrinsic cognitive load, extraneous cognitive load, germane cognitive load, and performance can be found in Table 2.

Table 2. Descriptive statistics.

Statistic	N	Mean	St. Dev.	Min.	Max.
Self-regulation	96	4.79	0.95	2.50	6.67
Prior knowledge	96	7.61	1.64	2	10
ICL	96	4.06	2.12	0.00	10.00
GCL	96	8.37	1.68	0.00	10.00
ECL	96	2.13	1.91	0.00	10.00
Performance	97	20.97	4.89	7	34

To validate the proposed hypotheses, structural equation modeling (SEM) with a maximum likelihood estimator was employed [47]. This approach was chosen to comprehensively evaluate the intricate relationship between students' cognitive load and their performance while also controlling for their prior knowledge and self-regulation skills. Since our hypotheses required a more comprehensive approach to capture the complex interplay between cognitive load, prior knowledge, self-regulation skills, and performance, SEM allowed us to model these intricate relationships holistically, providing a more detailed and nuanced understanding of the factors influencing students' learning outcomes. We did not use mediation analysis because our primary goal was to understand the direct and indirect relationships among multiple indicators simultaneously rather than focusing on a single mediating pathway. To mitigate the potential limitations associated with a relatively

small sample size and the variables' distribution specifics, we employed the Bollen-Stine bootstrap method [48] to obtain robust estimates and reduce the risks of underpowered analysis. The use of p -values derived from the Bollen-Stine bootstrap further enhanced the validity and robustness of our analysis.

The findings from the structural equation modeling indicate that some aspects of students' cognitive load are determined by their previous knowledge and self-regulated learning skills (Figure 2). Specifically, students who entered the course with a higher level of previous knowledge demonstrated a significantly higher level of germane cognitive load ($\beta = 0.24$, p -value < 0.05). Additionally, students with more developed self-regulated learning skills also exhibited a significantly higher level of germane cognitive load ($\beta = 0.25$, p -value < 0.05). However, no significant relationship was found between other types of cognitive load and these variables. These findings provide support for Hypothesis 1c (Prior knowledge has a positive relationship with germane load.) and Hypothesis 2c (Self-regulated learning has a positive relationship with germane load.).

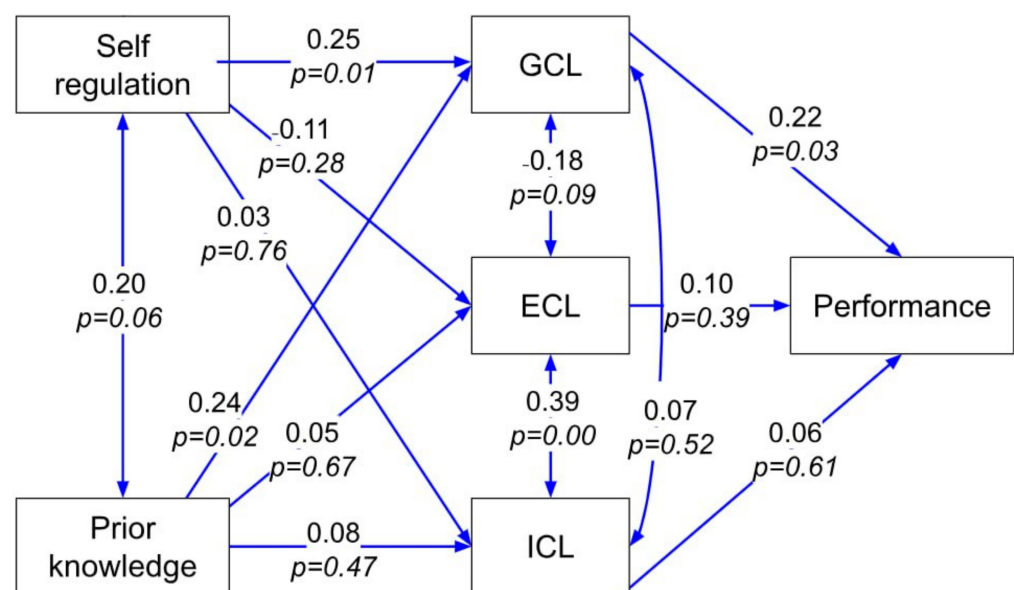


Figure 2. The results of structural equation modeling (the standardized estimates).

Furthermore, we explored the internal structure of cognitive load, examining relationships between its different aspects. The results reveal a significant positive relationship between intrinsic load and extraneous load ($\beta = 0.39$, p -value < 0.01). Additionally, a significant negative relationship was observed between extraneous load and germane load ($\beta = -0.18$, p -value < 0.10). Lastly, the study investigated the effect of different types of cognitive load on students' final performance. The results indicate that only germane cognitive load showed a significant positive linear relationship with student performance ($\beta = 0.22$, p -value < 0.05), supporting Hypothesis 3c (GCL has a positive relationship with performance.). However, no statistically significant relationships were found for other types of cognitive load. Finally, self-regulation showed a significant positive relationship with prior knowledge ($\beta = 0.20$, p -value < 0.10).

6. Discussion

This research investigated the complex relationships between prior knowledge, different types of cognitive load, and self-regulation, including metacognitive strategies and effort regulation as a component of self-regulated learning skills, and how they might correlate with performance in a learner-controlled environment. The findings of this study suggest that learners' individual characteristics, such as prior knowledge and self-regulated learning skills, are important factors affecting germane cognitive load in learner-controlled environments. On the other hand, our hypotheses on the direction and strength of the

relationships between intrinsic and extraneous types of cognitive load and prior knowledge, self-regulation, and performance were not confirmed by the results of the current study.

Intrinsic load was expected to have negative relationships with *prior knowledge*, *self-regulation*, and *performance* (Hypotheses 1a, 2a, and 3a). It was hypothesized that prior knowledge would have a negative relationship with intrinsic cognitive load, but this hypothesis was not confirmed. This does not fall in line with the fact that learning materials are likely to create fewer challenges for an expert learner compared to a novice learner [49], and learners with higher expertise are better able to manipulate the task, thus reducing content complexity during instruction [24]. Self-regulation was predicted to have a negative impact on intrinsic load due to the likelihood that self-regulated learners use learner control in ways that limit element interactivity [24], but the results of the current study are not in line with this assumption. Regarding the results involving the relationship that intrinsic load has with both prior knowledge and self-regulation, it could be that the learner-control options used in this study were effective in allowing learners to process a limited amount of elements at a single time regardless of prior knowledge and self-regulation levels. As for performance, it was assumed to have a negative relationship with intrinsic load, but the findings of the study do not support this hypothesis either. This may be accounted for by the level of complexity of the learning materials used in this study, as well as the general knowledge across the group. The average level of intrinsic load (4.06) reported by the participants could be regarded as optimal, i.e., neither too easy nor too difficult for learners. The results of the test on prior knowledge (7.61) might indicate that the group of students in this study have enough prior knowledge to successfully manage their intrinsic load.

It was anticipated that **extraneous load** would have a negative relationship with *prior knowledge*, *self-regulation*, and *performance* (Hypotheses 1b, 2b, and 3b). Previous research has shown that learners with substantial prior knowledge can effectively organize and integrate new information as well as effectively manage element interactivity through appropriate content selection in learner-controlled environments, resulting in reduced extraneous cognitive load [13,24]. Managing extraneous load becomes crucial when the content has high element interactivity and presents high intrinsic load for learners, but once the intrinsic complexity of the material does not overwhelm learners and impede information processing, managing extraneous load becomes achievable for students with different levels of expertise [28]. The intrinsic complexity of the materials for the current study is optimal, as the levels showed that the content was not overly simple or complex, and it may be the reason why prior knowledge has no relationship with extraneous cognitive load. The design of the materials was effective and did not require learners to use high levels of self-regulated learning skills, which is supported by the level of extraneous cognitive load (2.13). This fact might also explain why the results do not provide evidence of the negative relationship between extraneous load and performance. The instruction did not excessively strain working memory capacity and did not impede learning. In a learner-controlled environment, it could imply that even if learner control did impose additional load, overall working memory capacity was not exceeded [50], and thus, there are no relationships between extraneous cognitive load and performance.

It is of interest that the extraneous load in this study, while not intentionally manipulated, was low, and it had a relationship with the relatively low levels of intrinsic load. It is generally considered that because intrinsic load represents the difficulty of instruction and extraneous load represents the clarity of instructional delivery [12,13], the general consensus is that they are independent of each other, as they reflect different sources of load. In the case of the present study, extraneous load was significantly low with intrinsic load leaning toward the low end as well. Future research may want to consider direct manipulation of the loads. This could potentially solve two issues: 1) provide clarity as to what sources of loads are actually being examined, and 2) allow for the control of the amount of extraneous load in the environment to see how specific levels of extraneous load affect the relationships in this study.

Germane cognitive load refers to the actual working memory resources dedicated to managing intrinsic cognitive load [13]. We predicted that germane load would have positive relationships with all variables: *prior knowledge*, *self-regulation*, and *performance* (Hypotheses 3a, 3b, and 3c). Since prior knowledge considers mental schemas available in long-term memory, it is expected that having more schemas allows learners to allocate their mental resources to processing new information, rendering the mental effort germane. The results of this study provide insights into learner-controlled environments and generally support these expectations, showing a positive relationship between prior knowledge and germane cognitive load. Self-regulation shows a positive relationship with germane load as well; it can be attributed to the fact that self-regulated learners possess a repertoire of cognitive and metacognitive strategies that enable them to manage their learning processes effectively. Moreover, previous research [20] suggests that high levels of self-regulated learning skills imply that the skill is performed automatically, thus reducing cognitive load. Furthermore, self-regulated learners are more likely to engage in deep processing and employ effective learning strategies, including the strategic use of learner-control options when presented with such tools [51], which further contributes to the enhancement of germane cognitive load and better learning outcomes due to learners' ability to process information directly relevant to the learning task [52]. This is also confirmed by the findings of the current study, which show a positive relationship between germane load and performance. The allocation of cognitive resources toward germane load facilitates the integration of new information with prior knowledge, promoting the creation of schemas and ultimately leading to better performance.

Our results show no relationship between germane load and intrinsic load, going against research that states germane load is actually the availability of resources directed at intrinsic load. However, looking at the mean difference between the two loads in this study, 8.37 for germane and 4.06 for intrinsic, it can be inferred that although no relationship was established, the relatively low amount of intrinsic load presented the learners with more cognitive capacity for germane processes.

The relatively low levels of cognitive effort created through low intrinsic and extraneous processing most likely left more room to process relevant information that would lead to understanding. It may be that because the instruction was generally not unclear/confusing or difficult, this would leave the learners with more room to direct their processes to understanding the material. Along the lines of what Leppink et al. (2014) [12] have postulated, it could be said that because the material was not confusing or difficult, there were more resources available to try to manage intrinsic load, which, from the participants' perspective, meant that the material was more understandable. This is not to say that germane load, as presented in this study, represents the effort to understand the material. That effort can be seen in the relationship with self-regulation and germane load, showing that the characteristics of motivation for learners allowed them to generate the effort to better understand it. This falls in line with Schnotz (2010) [38] and Kalyuga (2011) [31]. Furthermore, empirical evidence supports this notion [35], showing that the effort attributed to self-regulation allowed learners to overcome high levels of intrinsic load and generate higher levels of germane load within learner-controlled environments. Finally, the positive relationship between self-regulation and germane load might be explained through the S2D2 framework, which suggests that effort directed at desirable difficulties ultimately leads to better learning outcomes [53]. It is likely that the germane load was directed at overcoming the intrinsic difficulty of the learning material, so a positive correlation with effort management, or as referred to in this study, self-regulation, should be expected.

One possible explanation for the observed discrepancies in the hypotheses regarding intrinsic and extraneous cognitive load and the confirmation of all hypotheses relating to germane load is that learners in learner-controlled environments may have adopted specific strategies that effectively converted much of their cognitive load from intrinsic and extraneous to germane [9], thereby freeing up cognitive resources. This optimization of cognitive load could have enabled learners to better leverage their self-regulation abilities to take advantage

of the available cognitive resources, deepening their understanding of the material. All learners, regardless of their expertise or self-regulation, were able to successfully mitigate some of the negative effects of cognitive load in the learner-controlled environment.

The research findings are of value for researchers and practitioners, as they provide insights into how prior knowledge and self-regulated learning skills interact with different types of cognitive load and to what extent these factors matter for learning outcomes. Regarding practical implications, the correlations between the variables underline the importance of providing learners with options to control their learning, taking into account their prior knowledge and level of self-regulated learning skills. Another significant finding of this study implies that it becomes even more essential to focus on fostering germane load within learner-controlled environments, as germane load not only leads to better performance but also helps decrease the negative side of some of the cognitive load effects. Our study offers a unique and original contribution by demonstrating how low levels of intrinsic and extraneous cognitive load can enhance comprehension in learner-controlled environments. Our findings underscore the importance of fostering germane load to improve learning outcomes, highlighting the value of learner control in effectively mitigating cognitive load and promoting deeper understanding.

The current study has some limitations. The most important one concerns the measurement tools used for the current study, which rely on self-reporting of both self-regulated learning skills and cognitive load that might distort the results. It may be that the wording of the items in Leppink et al. (2013) [30] led the participants to confound the concepts. Items address intrinsic load as complex concepts and extraneous load as unclear explanation, thus potentially leading students to be unable to distinguish between the two; unclear explanations in their mind potentially meant the concepts being explained were complex. This is a possibility when subjective measures are used and was discussed by Cierniak et al. (2009) [54], who concluded that the distinction between load types related to subjective cognitive load measures needs clarification. Concerning the germane load items, it would be logical to conclude that high levels of germane load represent high levels of understanding, which has been postulated by previous research [12]. The germane load items used in this study are specifically worded as the material enhancing “my understanding”. As for the self-regulated learning measurement tool, firstly, this quantitative approach is likely to measure the number of times a specific self-regulation strategy was used, not the quality of the learning approach [55]. Secondly, it is suggested by previous research that self-reported questionnaires tend to better reflect overall metacognitive skills rather than the use of specific strategies at a micro-level [56]. Other limitations include the narrow scope of the research conditioned by the small sample, and restrictions imposed by the legal domain might not allow scaling of the research findings. Finally, it would be of high interest to explore causal relations of the variables as well as to understand specific differences between learner-controlled and instructor-controlled environments; however, this research does not provide these insights due to the research design.

Future research should take into consideration these effects when choosing the measurement tool. It should also compare different conditions as well as look into specific aspects of learner control and its impact on different types of cognitive load. It would also be interesting to explore which specific strategies are employed in learner-controlled environments that make students’ mental effort germane and what other factors arise within learner control that could assist students in mitigating cognitive load effects.

7. Conclusions

The present study aimed at discovering the relationships between individual characteristics of students, such as prior knowledge and self-regulated learning skills and different types of cognitive load and understanding, and whether they have significant correlations with learning outcomes in learner-controlled environments. The results show that in learner-controlled environments, both prior knowledge and self-regulated learning skills are significantly correlated with germane load, which has a positive relationship

with performance. This implies that when learners with little prior knowledge and low self-regulated learning skills face the absence of guidance and structure, they might not be able to use efficient learning strategies and struggle with increased cognitive load. This underscores the importance of considering the interplay between prior knowledge, learner control, and cognitive load in the design of effective learner-controlled environments, taking into account students' individual characteristics.

Author Contributions: Conceptualization, methodology, original draft preparation, A.G.; original draft preparation, C.L.; original draft preparation, materials preparation, A.S.; formal analysis, original draft preparation, K.A.; supervision, project administration, review and editing, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: The article was prepared within the framework of the HSE University Basic Research Program.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of HSE University (protocol from 15 February 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are available on request by the author.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Self-Regulation Scale of MSLQ (Pintrich and de Groot, 1990) [16]

1. I ask myself questions to make sure I know the material I have been studying.
2. I work on practice exercises and answer end of chapter questions even when I don't have to.
3. Even when study materials are dull and uninteresting, I keep working until I finish.
4. Before I begin studying I think about the things I will need to do to learn.
5. When I'm reading I stop once in a while and go over what I have read.
6. I work hard to get a good grade even when I don't like a class.

Appendix B. Items for Cognitive Load Measurement (Leppink et al., 2013) [30]

1. The topics covered in the activity were very complex.
2. The activity covered information I perceived as very complex.
3. The activity covered concepts and definitions that I perceived as very complex.
4. The instructions and/or explanations during the activity were very unclear.
5. The instructions and/or explanations were, in terms of learning, very ineffective.
6. The instructions and/or explanations were full of unclear language.
7. The activity really enhanced my understanding of the topics covered.
8. The activity really enhanced my knowledge and understanding of the class subject.
9. The activity really enhanced my understanding of concepts and definitions.
10. The activity really enhanced my understanding of the content covered.

References

1. Scheiter, K. The Learner Control Principle in Multimedia Learning. In *The Cambridge Handbook of Multimedia Learning*, 3rd ed.; Mayer, R.E., Fiorella, L., Eds.; Cambridge University Press: Cambridge, UK, 2021; pp. 418–429. [CrossRef]
2. Vandewaetere, M.; Clarebout, G. Cognitive Load of Learner Control: Extraneous or Germane Load? *Educ. Res. Int.* **2013**, *2013*, 1–11. [CrossRef]
3. Kinzie, M.B. Requirements and benefits of effective interactive instruction: Learner control, self-regulation, and continuing motivation. *Educ. Technol. Res. Dev.* **1990**, *38*, 5–21. [CrossRef]
4. Kopcha, T.J.; Sullivan, H. Learner preferences and prior knowledge in learner-controlled computer-based instruction. *Educ. Technol. Res. Dev.* **2007**, *56*, 265–286. [CrossRef]
5. Niederhauser, D.S.; Reynolds, R.E.; Salmen, D.J.; Skolmoski, P. The Influence of Cognitive Load on Learning from Hypertext. *J. Educ. Comput. Res.* **2000**, *23*, 237–255. [CrossRef]
6. Jansen, R.S.; van Leeuwen, A.; Janssen, J.; Conijn, R.; Kester, L. Supporting learners' self-regulated learning in Massive Open Online Courses. *Comput. Educ.* **2020**, *146*, 103771. [CrossRef]

7. Zimmerman, B.J. Becoming a Self-Regulated Learner: An Overview. *Theory Into Pract.* **2002**, *41*, 64–70. [CrossRef]
8. Young, J.D. The effect of self-regulated learning strategies on performance in learner controlled computer-based instruction. *Educ. Technol. Res. Dev.* **1996**, *44*, 17–27. [CrossRef]
9. Hatsidimitris, G.; Kalyuga, S. Guided self-management of transient information in animations through pacing and sequencing strategies. *Educ. Technol. Res. Dev.* **2013**, *61*, 91–105. [CrossRef]
10. Castro-Alonso, J.C.; de Koning, B.B.; Fiorella, L.; Paas, F. Five Strategies for Optimizing Instructional Materials: Instructor- and Learner-Managed Cognitive Load. *Educ. Psychol. Rev.* **2021**, *33*, 1379–1407. [CrossRef]
11. Hasler, B.S.; Kersten, B.; Sweller, J. Learner control, cognitive load and instructional animation. *Appl. Cogn. Psychol.* **2007**, *21*, 713–729. [CrossRef]
12. Leppink, J.; Paas, F.; van Gog, T.; van der Vleuten, C.P.M.; van Merriënboer, J.J.G. Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learn. Instr.* **2014**, *30*, 32–42. [CrossRef]
13. Sweller, J.; van Merriënboer, J.J.G.; Paas, F. Cognitive Architecture and Instructional Design: 20 Years Later. *Educ. Psychol. Rev.* **2019**, *31*, 261–292. [CrossRef]
14. van Merriënboer, J.J.G.; Sweller, J. Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educ. Psychol. Rev.* **2005**, *17*, 147–177. [CrossRef]
15. Chen, O.; Paas, F.; Sweller, J. A Cognitive Load Theory Approach to Defining and Measuring Task Complexity Through Element Interactivity. *Educ. Psychol. Rev.* **2023**, *35*, 63. [CrossRef]
16. Pintrich, P.R.; De Groot, E.V. Motivational and self-regulated learning components of classroom academic performance. *J. Educ. Psychol.* **1990**, *82*, 33–40. [CrossRef]
17. Bhat, Z.H. Does Learner Control Prove Effective in the Systems of e-Learning? A Review of Literature. In *Learning How to Learn Using Multimedia*; Mishra, D., Chuang, Y., Eds.; In Lecture Notes in Educational Technology; Springer: Singapore, 2021; pp. 133–147. [CrossRef]
18. Mayer, R. Using multimedia for e-learning. *J. Comput. Assist. Learn.* **2017**, *33*, 403–423. [CrossRef]
19. de Bruin, A.B.H.; Roelle, J.; Carpenter, S.K.; Baars, M.; Mre, E. Synthesizing Cognitive Load and Self-regulation Theory: A Theoretical Framework and Research Agenda. *Educ. Psychol. Rev.* **2020**, *32*, 903–915. [CrossRef]
20. Wirth, J.; Stebner, F.; Trypke, M.; Schuster, C.; Leutner, D. An Interactive Layers Model of Self-Regulated Learning and Cognitive Load. *Educ. Psychol. Rev.* **2020**, *32*, 1127–1149. [CrossRef]
21. Sweller, J. Cognitive load theory and individual differences. *Learn. Individ. Differ.* **2024**, *110*, 102423. [CrossRef]
22. Biard, N.; Cojean, S.; Jamet, E. Effects of segmentation and pacing on procedural learning by video. *Comput. Hum. Behav.* **2018**, *89*, 411–417. [CrossRef]
23. Sweller, J. Cognitive load theory and educational technology. *Educ. Technol. Res. Dev.* **2020**, *68*, 839982. [CrossRef]
24. Gerjets, P.; Scheiter, K.; Catrambone, R. Designing Instructional Examples to Reduce Intrinsic Cognitive Load: Molar versus Modular Presentation of Solution Procedures. *Instr. Sci.* **2004**, *32*, 33–58. [CrossRef]
25. Eitel, A.; Endres, T.; Renkl, A. Self-management as a Bridge Between Cognitive Load and Self-regulated Learning: The Illustrative Case of Seductive Details. *Educ. Psychol. Rev.* **2020**, *32*, 1073–1087. [CrossRef]
26. Sweller, J. Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educ. Psychol. Rev.* **2010**, *22*, 123–138. [CrossRef]
27. Kalyuga, S.; Ayres, P.; Chandler, P.; Sweller, J. The Expertise Reversal Effect. *Educ. Psychol.* **2003**, *38*, 23–31. [CrossRef]
28. Chen, O.; Kalyuga, S.; Sweller, J. The Expertise Reversal Effect is a Variant of the More General Element Interactivity Effect. *Educ. Psychol. Rev.* **2017**, *29*, 393–405. [CrossRef]
29. Chen, O.; Retnowati, E.; Chan, B.K.Y.; Kalyuga, S. The effect of worked examples on learning solution steps and knowledge transfer. *Educ. Psychol.* **2023**, *43*, 914–928. [CrossRef]
30. Leppink, J.; Paas, F.; Van Der Vleuten, C.P.M.; Van Gog, T.; Van Merriënboer, J.J.G. Development of an instrument for measuring different types of cognitive load. *Behav. Res. Methods* **2013**, *45*, 1058–1072. [CrossRef] [PubMed]
31. van Merriënboer, J.J.G.; Ayres, P. Research on cognitive load theory and its design implications for e-learning. *Educ. Technol. Res. Dev.* **2005**, *53*, 5–13. [CrossRef]
32. de Koning, B.B. Best of Both Worlds? Combining Physical and Mental Self-Management Strategies to Support Learning from Split-Attention Examples. *Educ. Sci.* **2024**, *14*, 284. [CrossRef]
33. Schunk, D.H.; Greene, J.A. (Eds.) *Handbook of Self-Regulation of Learning and Performance*, 2nd ed.; Educational Psychology Handbook Series; Routledge, Taylor & Francis Group: New York, NY, USA, 2018.
34. Wang, T.-H. Developing an assessment-centered e-Learning system for improving student learning effectiveness. *Comput. Educ.* **2014**, *73*, 189–203. [CrossRef]
35. Lange, C.; Costley, J. The moderating effects of intrinsic load on the relationship between self-regulated effort and germane load. *J. Comput. Assist. Learn.* **2018**, *34*, 652–660. [CrossRef]
36. Baars, M.; Wijnia, L.; de Bruin, A.; Paas, F. The Relation Between Students' Effort and Monitoring Judgments During Learning: A Meta-analysis. *Educ. Psychol. Rev.* **2020**, *32*, 979–1002. [CrossRef]
37. Lange, C.; Costley, J.; Han, S.-L. The Effects of Extraneous Load on the Relationship Between Self-Regulated Effort and Germane Load Within an E-Learning Environment. *Int. Rev. Res. Open Distrib. Learn.* **2017**, *18*, 64–83. [CrossRef]
38. Schnotz, W. Reanalyzing the expertise reversal effect. *Instr. Sci.* **2010**, *38*, 315–323. [CrossRef]

39. Kostons, D.; van Gog, T.; Paas, F. Self-assessment and task selection in learner-controlled instruction: Differences between effective and ineffective learners. *Comput. Educ.* **2010**, *54*, 932–940. [CrossRef]
40. Duncan, T.G.; McKeachie, W.J. The Making of the Motivated Strategies for Learning Questionnaire. *Educ. Psychol.* **2005**, *40*, 117–128. [CrossRef]
41. Paas, F.; Tuovinen, J.E.; van Merriënboer, J.J.G.; Darabi, A.A. A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educ. Technol. Res. Dev.* **2005**, *53*, 25–34. [CrossRef]
42. Pi, Z.; Liu, C.; Meng, Q.; Yang, J. Co-learner presence and praise alters the effects of learner-generated explanation on learning from video lectures. *Int. J. Educ. Technol. High. Educ.* **2022**, *19*, 58. [CrossRef]
43. Moos, D.C. Hypermedia learning: Considering Cognitive Load and self-regulated learning. *J. Educ. Multimed. Hypermedia* **2013**, *22*, 39–61.
44. Moos, D.C.; Azevedo, R. Monitoring, planning, and self-efficacy during learning with hypermedia: The impact of conceptual scaffolds. *Comput. Hum. Behav.* **2008**, *24*, 1686–1706. [CrossRef]
45. Costley, J.; Gorbunova, A.; Savelyev, A.; Shcheglova, I.; Lange, C. Using worked examples to support problem solving of legal cases in an inductive instructional sequence. *Innov. Educ. Teach. Int.* **2024**, 1–15. [CrossRef]
46. Taber, K.S. The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Res. Sci. Educ.* **2018**, *48*, 1273–1296. [CrossRef]
47. Rosseel, Y. lavaan: AnRPackage for Structural Equation Modeling. *J. Stat. Softw.* **2012**, *48*, 1–36. [CrossRef]
48. Bollen, K.A.; Stine, R.A. Bootstrapping Goodness-of-Fit Measures in Structural Equation Models. *Sociol. Methods Res.* **1992**, *21*, 205–229. [CrossRef]
49. Sweller, J. Cognitive Load Theory: Recent Theoretical Advances. In *Cognitive Load Theory*, 1st ed.; Plass, J.L., Moreno, R., Brünken, R., Eds.; Cambridge University Press: Cambridge, UK, 2010; pp. 29–47. [CrossRef]
50. Paas, F.; Sweller, J. An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educ. Psychol. Rev.* **2012**, *24*, 27–45. [CrossRef]
51. Saqr, M.; Matcha, W.; Uzir, N.A.; Jovanovic, J.; Gašević, D.; López-Pernas, S. Transferring effective learning strategies across learning contexts matters: A study in problem-based learning. *Australas. J. Educ. Technol.* **2023**, *39*, 35–57. [CrossRef]
52. Kalyuga, S. Cognitive Load Theory: How Many Types of Load Does It Really Need? *Educ. Psychol. Rev.* **2011**, *23*, 1–19. [CrossRef]
53. de Bruin, A.B.H.; Biwer, F.; Hui, L.; Onan, E.; David, L.; Wiradhany, W. Worth the Effort: The Start and Stick to Desirable Difficulties (S2D2) Framework. *Educ. Psychol. Rev.* **2023**, *35*, 41. [CrossRef]
54. Cierniak, G.; Scheiter, K.; Gerjets, P. Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Comput. Hum. Behav.* **2009**, *25*, 315–324. [CrossRef]
55. Wirth, J.; Leutner, D. Self-Regulated Learning as a Competence: Implications of Theoretical Models for Assessment Methods. *J. Psychol.* **2008**, *216*, 102–110. [CrossRef]
56. Rovers, S.F.E.; Clarebout, G.; Savelberg, H.H.C.M.; de Bruin, A.B.H.; van Merriënboer, J.J.G. Granularity matters: Comparing different ways of measuring self-regulated learning. *Metacogn. Learn.* **2019**, *14*, 1–19. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Effects of Observing Urban and Natural Scenes on Working Memory Depletion and Restoration: An EEG Study

Lorenzo Consalvi ^{1,†}, Kim Ouwehand ^{1,*} and Fred Paas ^{1,2,3}¹ Department of Psychology Education and Child Studies (DPECS), Erasmus School of Social and Behavioural Sciences (ESSB), Erasmus University Rotterdam, 3062 PA Rotterdam, The Netherlands; paas@essb.eur.nl (F.P.)² School of Education/Early Start, University of Wollongong, Northfields Ave., Wollongong, NSW 2522, Australia³ School of Education, University of New South Wales, Sydney, NSW 2308, Australia

* Correspondence: ouwehand@essb.eur.nl

† Deceased author.

Abstract: Cognitive load theory focuses on the limited capacity of working memory (WM) to encapsulate information. While the original theory postulated a fixed capacity of working memory, research in the last decade has provided evidence for the depletion hypothesis. This hypothesis holds that WM becomes depleted after effortful cognitive operations that reduce its capacity, providing a framework for the restorative effects of rest periods. Rest periods during which natural scenery is observed have been found to replenish working memory after it has been subjected to depletion. In the present study, participants observed pictures depicting either a natural or an urban environment, after completing a cognitively depleting task. For this study, we obtained EEG measures of working memory by analyzing alpha and theta wave amplitudes. The motivation behind this choice was to derive a continuous index of WM capacity and address the lack of electrophysiological data regarding the depletion hypothesis. Previous research identified a decrease in alpha amplitude, and a simultaneous increase in theta activity with increasing WM load. Our findings partially replicated these results, as we observed a decrease in alpha amplitude with increasing cognitive load but no significant difference in theta power. Moreover, average signal amplitudes did not differ between the natural and the urban conditions, contrary to our hypothesis. These results suggest an absence of the expected environmental effect, opposing the outcome of existing research on the topic. The absence of this effect could also be attributed to similarities between the two conditions in certain factors thought to elicit differential physiological responses.

Keywords: cognitive load theory; working memory resource depletion; stress; EEG; environment

Citation: Consalvi, L.; Ouwehand, K.; Paas, F. Effects of Observing Urban and Natural Scenes on Working Memory Depletion and Restoration: An EEG Study. *Educ. Sci.* **2024**, *14*, 1204. <https://doi.org/10.3390/educsci14111204>

Academic Editor: Daniel Muijs

Received: 27 June 2024

Revised: 25 August 2024

Accepted: 25 October 2024

Published: 1 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The common cultural perception assuming that contact with natural environments induces beneficial effects on mental and spiritual well-being has spurred the emergence of a considerable amount of research unveiling the mechanisms that lie behind the relationship between the human brain and nature. The body of work that resulted was able to report the benefits of the interaction with nature on mental health, cognitive functions and psychophysiological processes [1–4].

The aim of the present study was to investigate the effect of observing natural scenery, as opposed to urban scenery, on the restoration of working memory after depletion by a cognitive effortful task. The intent was to substantiate the existing literature on the topic by providing psychophysiological measurements, using the electroencephalogram (EEG), and self-rated assessment of cognitive load. Furthermore, the research aimed to investigate the role of stress in the aforementioned relationship by complementing subjective ratings with heart rate measurements. It is important to consider objective physiological measures alongside subjective personal ratings, as they provide a more comprehensive understanding

of cognitive processes, such as cognitive load, and arousal states, such as increased stress. Additionally, these measures allow for continuous monitoring throughout the execution of an experiment, offering more detailed insights into participants' responses [5,6].

2. Working Memory Resource Depletion as an Extension of Cognitive Load Theory

Cognitive load theory (CLT) is a framework aimed at understanding how information processing load induced by a learning task affects the brain's capacity to process new information. It is based on the assumption that human cognitive processing is limited by our working memory capacity, which can only hold a limited number of elements simultaneously [7]. Recent work around cognitive load theory has suggested the notion that working memory capacity is not fixed but varies depending on the learning conditions [8]. The hypothesis worked out by Chen and associates (2018) [8] suggests that working memory becomes fatigued after sustained cognitive processing, decreasing its capacity to process further information. This aligns with previous conceptualizations of WM depletion. However, their work also indicates that this depletion effect can be reversed with adequate rest. The conditions under which working memory can efficiently replenish remain an open question posed by Chen and associates. Fortunately, the literature has extensively investigated related phenomena [2,3,7]. In the present study, we focused on the attention restoration theory (ART), first proposed by Kaplan (1995) [2], and the stress reduction theory (SRT), advanced by Ulrich (1981) [3].

2.1. Attention Restoration Theory

Attention restoration theory [2] distinguishes between two main processes of attention. The first process, involuntary attention, is naturally attracted by salient or fascinating stimuli. The second process, directed attention, is guided by cognitive control and involves the suppression of distracting elements. Kaplan (1995) [2] referred to the latter as "directed attention". According to ART, involuntary attention is immune to fatigue because it requires no conscious effort. This allows individuals engaging in involuntary attention to let their directed attention rest, providing an opportunity for recovery. ART proposes interventions to facilitate the recovery of effortful attention allowing cognitive rest [9]. The primary intervention is sleep. Alternatively, meditation serves as an effective aid for cognitive restoration. The third method, and the focus of the present study, is contact with environments that spontaneously capture involuntary attention, like natural environments.

2.2. Stress Reduction Theory

Stress reduction theory [3] posits that contact with natural environments provides greater benefits than environments lacking natural elements. Ulrich (1981) [3] observed that exposure to natural scenes can reduce arousal following stress, as evidenced by emotional states reports, heart rate and alpha rhythm amplitude. The mechanisms that underlie this arousal reduction are explained by traditional cognitive theories of emotion [10,11]. According to these theories, the first step in the appraisal of an environment is generalized affect (e.g., liking, interest, fear), which can be triggered by visual perception, imagination or memory. This initial affect reaction then influences cognitive evaluation of the scene [10]. Cognitive processes assess the stimuli (i.e., the environment) in terms of their significance for well-being. Cognitive theories of emotion suggest that the affective reaction of the observer is modeled by the preceding cognitive evaluation, refining the initial generalized affect and potentially giving rise to other emotions. These processes bring about physiological arousal and subjective feelings [10,12]. In the case of natural environments, affective reactions are thought to stem from interest and positive appraisal of their adaptive qualities, like the absence of threats and the availability of resources [13]. This hypothesis is rooted in an evolutionary perspective; in other words, scenes and sounds from nature are believed to be positively appraised because they represent environments where humans historically thrived. Ratcliffe and colleagues used semi-structured interviews to investigate which types of environment participants felt would facilitate recovery from stress and negative

affect. In the twenty interview transcripts analyzed, they identified 186 examples of natural sounds, particularly birdsong (35%). Participants associated these sounds with positively valenced memories and feelings of safety [13].

These two theoretical perspectives offer valuable insights into the underlying mechanisms through which natural scenes impact cognition and appraisal, which were central to our investigation in this study. Distinguishing the individual effects of stress and attention decline can be challenging, as these phenomena often occur simultaneously [2]. The scope of this research was not to evaluate the validity of the assumptions underlying these theories but rather to replicate the co-occurrence of these cognitive processes as outlined in Kaplan's (1995) [2] integrative framework. We suggest that future research should delve further into exploring the causal relationship between attention depletion and stress.

2.3. Integration

In light of the two paradigms proposed above, we will now focus on why WM is an important variable to assess. According to CLT, cognitive load rests on the capacity of working memory of withholding and processing information [7]. Only a limited amount of novel information can be remembered at the same time, and the amount decreases when this information has to be combined and elaborated on [14]. Attention and stress both influence WM depletion and restoration. To integrate WM into the ART framework, it is useful to examine the construct of executive functions. Executive functions (EFs) are a set of effortful mental processes required for concentration and selective attention. There are three main executive functions, namely inhibition (behavioral and cognitive inhibition), cognitive flexibility, and working memory [15]. Directed attention and WM are sustained by largely overlapping prefrontal–parietal systems, that enable top-down modulation. Top-down modulation is the mechanism that allows one to focus attention on task-relevant inputs while ignoring irrelevant ones. This same architecture is crucial for prioritizing stimuli during WM tasks [16].

When examining the relationship between stress and cognition, it has been observed that both everyday stressors and major stressful life events can bring about a decrease in WM capacity and attentional resources [17]. The study by Klein and Boals (2001) [17] supported the hypothesis that stressors represent involuntary thoughts that require WM resources to suppress. To test this hypothesis, the researchers administered a word memory task to participants, consisting of six sets of increasing difficulty. The study aimed at predicting the performance on this task based on participants' state anxiety and reported stress from negative life events, measured through questionnaires. The results indicated that participants reporting higher stress obtained a lower score on the word memory task on average. Additionally, the negative correlation between task performance and stress from life events became more pronounced with more demanding task requirements. Klein and Boals (2001) [17] concluded that, unlike simple distractions, thoughts about adverse life events do not diminish as task demands increase.

Keeping these connections in mind, in the present research, we tried to investigate whether WM resources could be depleted by inducing stress and directed attention fatigue, leading to an increase in cognitive load. This was achieved by the administration of a cognitively effortful mathematical task. Moreover, we aimed at reversing the effect of stress and fatigue on WM resources, thus on cognitive load, through the presentation of natural scenarios. The extent to which we managed to increase and subsequently decrease cognitive load was assessed by means of performance on a WM task, administered once after the completion of the cognitively depleting task and once after scenario presentation.

2.4. Physiological Measures of Cognitive Load: The Alpha and Theta Rhythms

Electroencephalographic (EEG) measures of cognitive load were central to the aim of the present study. EEG is a non-invasive neuroimaging technique that measures electrical activity in the brain with the use of electrodes placed on the scalp [5]. The EEG signal is the result of a number of oscillations that operate at different frequencies. These oscillations are

believed to represent information processing and transfer between clusters of neurons [18]. The two oscillations taken into consideration in this research, alpha and theta, have been observed to react to differential cognitive load demands [6,18]. Alpha is the dominant frequency in the adult human EEG. It manifests itself with a “peak” in the conventional alpha region (7.5–12.5 Hz) in spectral analysis and rhythmic alpha waves [6]. This particular frequency is most easily recorded at posterior sites and it is commonly observed when the individual is in a state of “relaxed wakefulness” [18]. In contrast, the theta frequency reaches a peak within a frequency of 4–7.5 Hz, with a maximum amplitude around the frontal midline [19]. It is most prominent when the individual exerts focused and extensive attention on a task, but it is also induced during nocturnal sleep (especially in REM and in stage 1 NREM sleep). Alpha and theta respond in their own characteristic ways to task demands. It has been found that, when comparing EEG power in a resting condition with power in a condition of mental activation, alpha power decreases (desynchronizes) while theta power increases (synchronizes) [6]. As described by Klimesch (1999) [6] in his review, physiologically, EEG power is a reflection of the number of neurons that discharge in synchronization.

The two waveforms we were interested in observing have also been relevant in research in the field of environmental psychology. Particularly, researchers have shown interest in assessing the behavior of alpha and theta brainwaves at the presentation of natural scenes, as opposed to urban scenes [3,20]. In his study on the effect of natural and urban scenes on stress levels, Ulrich (1981) [3] was able to observe higher alpha amplitudes in his subjects when they were asked to look at scenes that included vegetation, rather than when viewing urban scenes, indicating a more wakefully relaxed state in the former condition. Similarly, Chen and associates (2020) [20] reported higher alpha–theta power in the occipital lobes during exposure to a restorative, natural environment.

2.5. Objectives and Research Questions

This study was designed to examine posterior alpha and frontal-midline theta activities as indicators of WM resource depletion and restoration following exposure to either a natural or an urban environment. The research aimed to integrate findings on the physiological implications of cognitive load (measured by alpha and theta wave power; Chen et al., 2020; Ulrich, 1981) [3,20] with existing literature on the EEG response to restorative environmental experiences (Berman et al., 2008; Kaplan, 1995) [1,2]. ART and SRT provided useful frameworks for understanding and discussing the expected results. As mentioned, testing the validity and predictive quality of these two theories, as well as directly comparing them, was beyond the scope of this study.

The research questions this study aimed to answer were whether posterior alpha and frontal-midline theta activities are sensitive to WM resource depletion and environmental restoration and what role stress plays in this effect. To address these questions, it was necessary to formulate directly testable hypotheses. First, we hypothesized that alpha desynchronization and theta synchronization would occur after completing a WM-depleting task compared to baseline, along with an expected increase in heart rate due to the task. When exposed to urban scenery, participants were predicted to exhibit the same pattern of (de-)activation in line with the principles of ART and SRT. Conversely, after exposure to a natural environment we expected to observe the opposite pattern: alpha synchronization, theta desynchronization and a heart rate decrease compared to the post-depletion signal. Lastly, task performance and self-reported measures of cognitive load and stress were hypothesized to reflect the same pattern as the EEG measures and to show a significant association with them.

3. Methods

3.1. Participants

An estimation of the needed sample size for this study was derived from a meta-analysis of existing studies on restorative environments [21]. According to this meta-

analysis, the restorative effect of an environment on self-reported cognitive fatigue yielded an effect size of 0.76 (Hedges' g). Based on this effect size, the preferred sample size, as calculated using G*Power 3.1 under an a priori scenario, was 58 (29 for each independent group). Bowler and colleagues (2010) [21] also reported an effect size of 0.23 for the impact of restorative environments on cognitive performance measures of attention, which would have required a much larger sample size for the purpose of this study. Therefore, the planned sample size was set at $N = 58$. This more restricted sample size also facilitated the measurement of neural correlates of WM load and physiological assessment of stress. The final sample that attended and completed the experiment consisted of 62 healthy subjects (mean age $M = 20.3$, $SD = 1.92$; 66.7% females, 33.3% males), recruited from the population of psychology students at Erasmus University Rotterdam (Rotterdam, the Netherlands). Participants were recruited through ERAS, an administration and support system, and were awarded course credits as compensation for participating.

3.2. Materials and Measures

3.2.1. Working Memory Depletion and Assessment Tasks

The study made use of two tasks related to WM capacity. The first task, the Trier Mental Challenge Test (TMCT) [22], was used to induce working memory fatigue and stress in participants, before the presentation of natural or urban pictures. The TMCT is a computer-based task in which participants have five seconds in each trial to solve arithmetic operations. The task is divided into three blocks of three minutes each, in which participants need to try to answer to as many trials as possible. Difficulty of the trials increased according to performance. Initially, participants were presented with simple summation or subtraction problems (e.g., " $2 + 2$ "); as they accumulated correct responses, problems became longer and more complex, comprising different types of arithmetic operations (e.g., " $8/4 - 15/5 + 24/4$ "). Participants gained 1 point for each correct answer and lost 1 point for each incorrect or late response.

The second task, the digit span task, was used to assess WM capacity after depletion and manipulation. This type of task is a very common measure in research on the relationship between nature and cognitive processes. For example, Berman and associates (2008) [1] used a similar digit span paradigm to infer the effects of viewing pictures of natural scenes on directed attention abilities. Van Oordt et al. (2022) [23] used the digit span task as well, after inducing WM load through the TMCT, in their study on the restorative effects of natural scenes in a similar experimental setup. Traditionally, the digit span test is divided into two blocks, the forward and the backward block. In the forward digit span, participants need to recall the numbers in the same order as they were presented one by one on the screen. In the backward task, participants are instructed to recall the numbers in the opposite order than the one they were presented in. The first digit span block was administered following the cognitive depleting task, while the second block was presented after the picture viewing, in a counterbalanced order. Each block comprised 14 trials, with trial lengths adjusted based on performance to ensure adequate WM taxation. In the initial trial of each block, participants were required to recall a sequence of three digits, presented sequentially on the screen. With each successful recall, the sequence length increased by one digit, up to a maximum of seven digits. Both digit span blocks were visually presented on a monitor using E-Studio 3.0, the software employed for the experimental design, which is part of the E-Prime 3.0 package. Data for analysis were extracted using E-Data Aid, a software tool from the same package that facilitates data filtering and organization. The scores for both digit span blocks used in the main analysis were calculated by dividing the number of successfully recalled sequences by the total number of sequences presented to each participant.

3.2.2. Cognitive Load and Stress Questionnaires

For the self-assessment of cognitive load, the study relied on two items that rated the extent to which participants found the task difficult or mentally effortful, using a 9-point

Likert scale (Paas, 1992) [24]. Participants were asked to rate the difficulty with the prompt “Indicate on this scale how difficult you found the task” and assign a numerical value from 1 to 9 after task completion. Similarly, for mental effort, participants were asked, “Indicate on this scale how much mental effort it cost you to complete the task”. A value of 1 indicated an easy task with low mental effort, while a value of 9 indicated a difficult task with high mental effort. These two items were administered in the same order, three times throughout the experiment: after the cognitive depleting task (TMCT), and after the two digit span tasks. The reliability of the scale, measured with Cronbach’s alpha in Paas’ study (1992) [24], was established at $\alpha = 0.90$.

Stress was measured using the Zuckerman Inventory of Personal Reactions (ZIPERS) [25]. This questionnaire was administered twice in each condition throughout the course of the experiment, the first time after the completion of the first digit span task (in order of time) and the second after the picture-viewing manipulation. The aim was to observe the differential effect of natural and urban scenarios on levels of induced stress. The present study only included the ZIPERS scale measuring stress states. The stress states scores are obtained through responses to 13 items on a 5-point Likert scale. Each item describes the state respondents felt during the preceding experimental phase (e.g., “I felt carefree or playful”, “I felt attentive or concentrating”). Importantly, the questionnaire aimed at measuring scores on 5 different subscales (fear arousal, positive affect, anger and aggression, attentive coping, and sadness). Scores on the 13 5-point items were summed into 5 separate compound scores, that corresponded to the different subscales, according to the instructions found in the development study by Zuckerman (1977) [25]. Internal (split-half) reliability was tested over 10 situations, with coefficients never dropping below $\alpha = 0.80$ (Zuckerman, 1977) [25]. Test–retest reliability analyses showed a low to moderate coefficient of $\alpha = 0.34$, this result was predicted as the scale measures provisional stress states. Once again, the E-Studio software came in useful in administering the questions and allowing participants to provide the answers. The software E-Data Aid allowed us to extrapolate the score on each question for every participant.

3.2.3. EEG Recording and Data Preprocessing

Electroencephalographic activity was measured using a BioSemi ActiveTwo system. A set of 32 Ag/AgCl pin electrodes was attached to the scalp through a head cap according to the 10/20 international system. Four additional flat electrodes were used to record the electrooculogram (EOG): two placed above and below the left eye, to measure vertical eye movement, and two positioned on the outer canthi of both eyes, for horizontal eye movement. Moreover, two flat electrodes were placed on the opposite mastoids, and one other electrode was used to record the participants’ heart rate from beneath the rib. The sampling rate was set to 4096 Hz, with a decimation pre-set to 1/8, resulting in an effective sample rate of 512 Hz.

The data were pre-processed with the BrainVision Analyzer 2.0 program. First of all, the data were referenced to the “linked mastoids” (EXG5-EXG6 electrodes). The second step consisted in inspecting the data for abnormal signal ranges. We used the automatic Raw Data Inspector function, that allowed us to apply the same thresholds to all subjects. The Raw Data Inspector is mainly used to detect “flatliners” (channels that did not record any data). Hence, only the scan for “Low Activity” was allowed (below 0.1 μV) for all the channels, except VEOG, HEOG and reference channels. Data were high-passed at 0.1 Hz and low-passed at 60 Hz throughout each channel using an IIR filter. The next step involved the segmentation of the data to compare the wave amplitudes at different points in time in the experiment. Segments were created based on the marker position, for Analyzer to cut the signal around the starting point and ending point of the condition. The segmentation consisted in dividing the signal into five blocks, following the timeline of the experiment, in order, (1) baseline with participants being instructed to keep their eyes open, which lasted 1 min (or two blocks of 30 s); (2) baseline with participants instructed to keep their eyes closed, also lasting 1 min; (3) first administration of the digit span task,

that took on average 4.5 min; (4) urban or natural manipulation, of 5 min in total; and (5) second administration of the digit span task, also 5 min on average. This division allowed us to test for differences in the signal in adjacent segments, according to our hypotheses. Once the data were segmented, we dealt with artifacts and bad electrodes using the “Reject Artifacts” function. Every channel except the EOG channels were included in this transformation. To analyze the spectral content of the EEG data we used Fourier transformation. This method translates data from the time domain (temporal changes) into the frequency domain (variations across different frequency bands). To minimize the impact of the edge effects, due to the segmentation being carried out before the Fourier transformation, a Hanning window was applied. The final step involved averaging the data across different time points to reduce noise and identify differences in the signal.

A parallel process was conducted on the heart rate data. A single electrode, positioned below the participant’s left ribcage, was employed to record the signal. Noise was filtered out through the IIR Filter function, with a low cutoff set at 0.1 Hz and a high cutoff at 40 Hz. An additional notch filter was used to remove activity at 50 Hz. The data were subsequently segmented into four portions according to marker positions, following the experimental blocks. These segments resembled the ones that were obtained from the recording of EEG activity, except that no differences in heart rate were expected during baseline based on whether participants had their eyes open or closed. Hence, the segments for this measure were, in order, baseline, first digit span administration, picture viewing and second digit span administration. To extract electrocardiogram (ECG) features from the signal for analysis, we employed the ECG Markers function. This function detected the five standard ECG components (P, Q, R, S, T) based on the positions of the different peaks in the recorded cardiac activity.

After data pre-processing, the figures corresponding to alpha and theta amplitude and heart rate (calculated in beats per minute; bpm) were exported, to allow for statistical analysis. In the case of alpha and theta amplitude, the original data, measured in microvolts (μV), were log-transformed to compensate for skewness of the distribution. To obtain mean amplitude and mean heart rate, the signal was averaged across time for each block and the data extracted for each of the two frequency bands, as well as for the heart rate measurement. This way, we obtained mean alpha, mean theta and mean beats per minute in each individual segment for every electrode site that we retained. Before the main analyses were conducted, we averaged the value we got on each site to obtain a single, more comprehensive, mean amplitude to be compared across phases (in the case of heart rate, this step was not necessary, as we only used one electrode).

3.3. Stimuli and Equipment

Participants were divided into two conditions and presented with a set of 30 pictures of either a natural or an urban environment (Figure 1). To ensure the images were representative of realistic scenarios for the sample population, the pictures were taken in green areas or central neighborhoods of a Dutch city using a 9 MP digital camera. The pictures had a resolution of 3024×3024 pixels and were adjusted to a square format to maintain uniformity among the stimuli. Weather conditions, points of observation and luminosity were kept constant to eliminate the potential influence of these factors on mood or attentional processes. The pictures in both conditions were carefully selected to exclude people, as their presence could draw attention and affect the results [26]. During picture presentation, participants were instructed to simply observe the carousel of images as they scrolled. No specific instruction on how to engage with the images was provided. No additional tasks were administered during this phase.



Figure 1. Below is a Sample of the Pictures Used for the Manipulation. The Three Pictures on the First Row are Among those Used in the Natural Condition, While the Three at the Bottom Were Retrieved From the Urban Condition. The Full Set of Pictures is Provided in Appendix D.

The participants were presented with the pictures, as well as the questionnaires and tests, with E-Studio, from the E-Prime 3.0 package, on a 22" monitor, with a screen resolution of 1280×1024 pixels and a 120,000 Hz refresh rate. The test equipment also included a keyboard and a mouse, to allow participants to record their responses.

3.4. Procedure

Participants were informed about the procedure of the experiment in advance through e-mail. The e-mail contained instructions about the day and time of their reserved session, that required participants to be present at the Erasmus Behavioral Lab at Erasmus University. For the EEG data recording to be carried out successfully, participating students needed to show up with clean and dry hair and should have avoided consuming caffeine before the start of their session, for it could have compromised the results of the heart rate measurement. At the time participants were introduced to the lab, they needed to sign the informed consent form, informing them about the content and the goal of the experiment, as well as its duration and nature. The experiment unfolded in three phases, in the following order: the depletion phase, the manipulation phase and the assessment phase (Figure 2). Before the start of the tasks, students were wired to the BioSemi ActiveTwo equipment; after that, the depletion phase took place. All students, independently of their assigned condition (exposure to natural or urban pictures), completed the TMCT, after the measurement of physiological parameters at baseline. Baseline measurements were recorded with participants being presented an empty screen for the initial two minutes (alternating times when they needed to keep their eyes open with times in which they were asked to keep them closed). After the completion of the TCMT, subjects completed either the forward or the backwards block of the digit span task, based on a counterbalanced order. Subsequently, they were exposed to the manipulation phase. According to their assigned group, participants either viewed 30 pictures of a natural or an urban environment in random order. Each picture remained present on the screen for 10 s. In total, the manipulation phase lasted 5 min. Picture viewing was preceded and followed by the administration of the ZIPERS questionnaire. Next, the experiment proceeded with the assessment phase, in

which participants carried out the second digit span block. The TCMT and both digit span blocks were followed by the two self-rating questions concerning the difficulty and the required mental effort in relation to each task just completed. Each phase was introduced by an instructional text informing participants about the questionnaires or tasks about to be administered and the modality of response. Every participant provided their responses to the questionnaires and the tasks with their right hand, according to the setup of the working station, comprising a keyboard and a mouse. Responding to the questionnaires and to the TCMT was possible using the mouse; the digit span tasks required subjects to enter a response with the keyboard. Answers on the TCMT and on the digit span tasks were followed by feedback screens lasting 1000 ms after each trial. At the end of the three phases participants were thanked, disconnected from the BioSemi system, debriefed and dismissed after being given a chance to clean themselves up. This experimental procedure shared similar characteristics with previously cited studies on the restorative effects of nature [1,20,23,26].

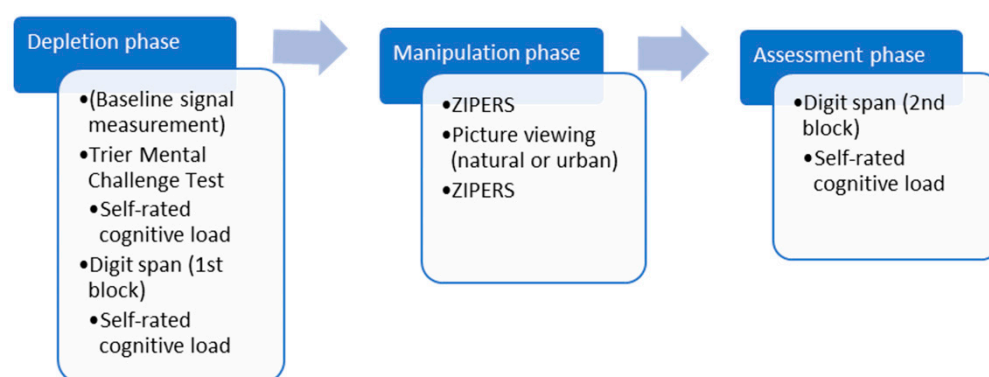


Figure 2. Experiment Flow With the Tasks Carried Out in Each Phase.

3.5. Processing and Statistical Analyses

At the end of data collection, the scores for each of the two conditions included accumulated points in the TCMT, pre- and post-manipulation ZIPERS scores, frontal-midline theta and occipital alpha average power (log-transformed), together with average heart rate (as measured at baseline, during depletion and during manipulation) and proportion of accuracy on the digit span task. Data processing and statistical analyses were all executed in IBM SPSS Statistics 27. The scores obtained were inspected for normality, homogeneity and outlying cases for each condition. Outliers were inspected considering Cook's distance and standardized residuals. Cases with a Cook's distance larger than 1 or standardized residuals more extreme than -3 or 3 (in task results and questionnaire scores) were removed. After inspection and data cleaning, a mean score was computed for each measure for both groups.

Differences in heart rate were investigated using a 2×4 factorial ANOVA. In this analysis, the two picture conditions were retained as a between-subjects factor, while the four experimental phases (namely, baseline, depletion, manipulation and assessment) functioned as a within-subject factor. Alpha and theta power were the dependent variables of two separate 2×5 factorial ANOVAs (one for each wave type). The same factors applied to these twin analyses, with the only difference being the separation of the baseline measurement into a period in which participants kept their eyes open with another one in which they were requested to close them. The aim was to confront differences in signal at different stages of the experiment (using each segment obtained during pre-processing as a measurement point) and at the same stage between groups. Fourteen electrodes out of the original thirty-two were retained for the analysis of variance of both alpha and theta, to infer statistical significance of the differences in signal amplitude. Electrode selection was based on previous studies in the context of research on cognitive load and its effect on the EEG spectrum [3,18,19,27,28]. According to their findings, the desynchronization

of the alpha wave following an increase in cognitive load mainly involves the occipital, occipitoparietal and parietocentral regions; on the other hand, a more emphatic increase of theta amplitude, connected to augmented cognitive demands, was observed along the frontal-midline area. Therefore, the fourteen electrodes of interest for the analysis of the alpha wave included C3, CP1, CP5, P3, Pz, PO3, O1, Oz, O2, PO4, P4, CP6, CP2 and C4. The other fourteen, of relevance for comparing theta amplitude, consisted in Fp1, AF3, F7, F3, FC1, FC5, FC6, FC2, F4, F8, AF4, Fp2, Fz and Cz.

Mixed ANOVAs were also the chosen method to observe the hypothesized differences in ZIPERS scores, self-assessed cognitive load and mean accuracy between the two digit span tasks. In the first case, data on the questionnaire were collected on two occasions, namely after the depletion and after the manipulation. In the second, the scores on cognitive load reports after each one of the digit span blocks were considered for the main analysis.

In all factorial ANOVAs, Bonferroni-corrected post hoc comparisons were applied to test the hypotheses that working memory load and stress are higher in the urban, compared to the natural, condition. In order to infer the validity of the bioindicators, the associations between stress self-reports and heart rate and between alpha/theta power, memory performance and self-rated cognitive load were measured using bivariate correlations. If the physiological data were a valid indication of the level of cognitive load and stress, we would expect a significant linear association between biomarkers, task performance and subjective ratings. A significance level of $\alpha = 0.05$ will be chosen for every statistical test.

4. Results

4.1. Descriptive Statistics

The descriptive statistics for each of the two conditions on every dependent variable in the experimental phases are reported in Table 1. An overview of the physiological measurements at baseline is found in Appendix A (Table A1).

Table 1. Descriptive Statistics of the Physiological and Behavioral Measurements in each experimental phase: Alpha and Theta Oscillations (log-transformed), Heart Rate (bpm), Digit Span (proportion score), Perceived Difficulty (1–9 Likert), Mental Effort (1–9 Likert) and scores on the ZIPERS items (Fear Arousal, Positive Affect, Anger and aggression, Attentive Coping and Sadness; sum score).

Experimental Phase	Depletion		Manipulation		Assessment	
	Natural	Urban	Natural	Urban	Natural	Urban
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Alpha	1.30 (1.69)	1.30 (0.81)	1.57 (0.74)	1.60 (0.85)	1.36 (0.75)	1.39 (0.85)
Theta	1.51 (0.62)	1.47 (0.58)	1.48 (0.62)	1.44 (0.57)	1.54 (0.61)	1.52 (0.55)
Heart rate	81.02 (13.11)	77.09 (9.18)	73.50 (10.67)	73.85 (10.14)	79.72 (12.46)	80.54 (11.50)
Digit span	0.68 (0.15)	0.63 (0.11)			0.67 (0.15)	0.66 (0.11)
DS difficulty	5.93 (1.39)	5.75 (1.63)			6.21 (1.80)	6.03 (1.90)
DS mental effort	6.83 (1.44)	6.47 (1.54)			6.79 (1.99)	6.81 (1.58)
Fear arousal	6.03 (1.92)	5.28 (1.93)	3.75 (1.96)	3.50 (1.01)		
Positive affect	7.86 (2.46)	8.09 (2.74)	12.17 (3.36)	7.81 (3.70)		
Anger and aggression	6.10 (2.66)	5.06 (2.02)	3.68 (1.72)	3.81 (1.40)		
Attentive coping	7.79 (1.63)	7.59 (1.68)	5.00 (1.89)	5.09 (1.89)		
Sadness	1.69 (0.93)	1.56 (0.88)	1.62 (0.98)	1.56 (0.80)		

Note: DS = digit span.

4.2. Physiological Measures of Stress and Cognitive Load

To measure the effect of depletion and environmental restoration on physiological indicators of cognitive load, this study focused on the difference in average alpha and theta wave amplitude between the two conditions. The results of the ANOVA conducted on the alpha signal showed no effect of condition $F(1, 56) = 0.35, p = 0.556, \eta_p^2 = 0.01$. However, a main effect of phase was found $F(2.72, 152.06) = 106.22, p < 0.001, \eta_p^2 = 0.66$ (Figure A2).

From the post hoc pairwise comparisons, it emerged that, in the natural condition, alpha amplitude significantly decreased from baseline to the depletion phase, as expected. Subsequently, we observed a significant increase during the manipulation compared to the depletion phase. Finally, alpha power in the natural condition decreased again as a result of increased cognitive load demands in the assessment phase; once again, this last difference was significant. Similar fluctuations in alpha power were observed among participants in the urban condition. In this group, alpha wave amplitude decreased from baseline to the depletion phase, followed by an increase during the manipulation phase, and then a subsequent decrease from manipulation to assessment. All observed differences in the urban group were statistically significant (a more detailed outline of this section, mean differences and significance level are reported in Table A2). The interaction effect between condition and phase was not significant $F(2.72, 151.06) = 2.56, p = 0.063, \eta_p^2 = 0.04$.

The other analysis of variance aimed at inferring differences in theta frequency amplitude. Similarly to the previous analysis, participants in the natural and in the urban conditions did not differ in their average theta amplitude $F(1, 56) = 0.12, p = 0.730, \eta_p^2 < 0.01$. However, the average signal of the two groups was significantly affected by experimental phase $F(3.46, 193.99) = 37.89, p < 0.001, \eta_p^2 = 0.40$ (Figure A2). In the natural group, post hoc pairwise comparisons highlighted a more pronounced theta amplitude in the depletion phase compared to baseline; this increase was significant. Theta amplitude then followed a decreasing tendency during the manipulation phase and a subsequent increase while participants completed the second digit span block, that is, during the assessment phase; the last two effects described were not significant. Theta frequency amplitude followed a similar trend in participants in the urban condition. Among those in this group, frontal-midline theta activity significantly increased from baseline to the depletion phase. The decrease observed during manipulation and the increase that occurred during assessment both fell below the significance level (mean difference values are reported in Table A3). In this case too, we encountered a non-significant interaction effect between condition and phase, meaning that the difference in signal between the two conditions did not change as an effect of phase $F(3.46, 193.99) = 1.76, p = 0.148, \eta_p^2 = 0.03$. For the heart rate measurement we encountered an increase (across both conditions) in beats per minute in the depletion phase compared to baseline. This effect was then followed by a decrease in bpm during picture viewing and a second increase in the assessment phase. Analysis of contrasts revealed a significant cubic effect of phase on bpm, $F(1, 49) = 49.66, p < 0.001, \eta_p^2 = 0.50$, but no effect of picture condition $F(1, 49) = 0.26, p = 0.612, \eta_p^2 < 0.01$ (Table A3).

4.3. Cognitive Performance

A third 2×2 ANOVA was used to infer the difference in digit span performance between the two groups after the depletion and after the manipulation. Hence, here we used condition as a between-subjects factor and time as a within-subject factor. Results showed neither main effects of picture condition, $F(1, 58) = 1.22, p = 0.274, \eta_p^2 = 0.02$, nor time, $F(1, 58) = 0.60, p = 0.442, \eta_p^2 = 0.01$, nor an interaction effect between picture condition and time $F(1, 58) = 1.92, p = 0.171, \eta_p^2 = 0.03$. These results illustrate that digit span performance did not change as a factor of time and that participants in the natural condition did not score on average higher nor lower than those in the urban condition on digit span accuracy. Lastly, there was no difference in scores from a measurement point to the other in either of the groups.

4.4. Subjective Reports of Stress and Cognitive Load

Statistical significance of the ZIPERS questionnaire difference scores was inferred by a series of mixed 2×2 ANOVA designs, with picture condition as a between-subjects factor and time as a within-subject factor. ANOVAs were carried out on each of the five subscales making up the questionnaire.

The only subscale that presented a significant difference between conditions was positive affect, with participants in the natural condition scoring 2.06 points higher on average

than their counterparts in the urban condition ($F(1, 59) = 10.44, p = 0.002, 95\% \text{ CI } [0.79, 3.34]$). Additionally, we observed changes in scores in all subscales, except sadness, between the first and the second measurement time. Fear arousal decreased by 2.11 points on average ($F(1, 57) = 61.56, p < 0.001, 95\% \text{ CI } [1.57, 2.65]$); similarly, anger and aggression showed a 2.07-point decrease ($F(1, 55) = 40.60, p < 0.001, 95\% \text{ CI } [1.42, 2.72]$). Positive affect increased by 2.02 points ($F(1, 59) = 17.76, p < 0.001, 95\% \text{ CI } [-2.97, -1.06]$), while attentive coping underwent the largest decrease with 2.65 points ($F(1, 59) = 96.44, p < 0.001, 95\% \text{ CI } [2.11, 3.19]$). The interaction effect between measurement time and condition was significant in two out of five cases, namely in positive affect, $F(1, 59) = 23.06, p < 0.001, \eta_p^2 = 0.23$, and anger and aggression $F(1, 55) = 5.25, p = 0.026, \eta_p^2 = 0.09$. In these two cases the difference in scores between conditions was dependent on measurement time. Scores were adjusted for multiple comparisons using the Bonferroni correction.

The 2×2 ANOVA designs were used to investigate the difference in cognitive load self-reports. On this occasion, we kept the same two conditions as a between-subjects factor and time as a within-subject factor. Even though participants were asked to assess their cognitive load on three occasions, namely after the depletion task and after each one of the two digit spans, only the scores on the last two assessments were included in the analysis. That is because we were mainly interested in the difference in cognitive load resulting from the manipulation. The analysis was performed on the two subscales measuring perceived task difficulty and invested mental effort. Participants in the natural and in the urban conditions did not report any significant difference in scores on difficulty, $F(1, 58) = 0.02, p = 0.880, \eta_p^2 = 0.00$, nor on mental effort $F(1, 58) = 0.50, p = 0.481, \eta_p^2 = 0.01$. On the former, scores did not differ as an effect of time $F(1, 58) = 1.16, p = 0.287, \eta_p^2 = 0.02$. The interaction effect between condition and time in the difficulty subscale was not significant $F(1, 58) = 0.01, p = 0.915, \eta_p^2 = 0.00$. Similar results were obtained from the analysis of the mental effort subscale, where time did not exert a significant effect on the scores, $F(1, 58) = 1.67, p = 0.201, \eta_p^2 = 0.03$, and the interaction effect between the between- and the within-subject factors remained below the significance level $F(1, 58) = 0.42, p = 0.839, \eta_p^2 = 0.00$. In these cases as well, Bonferroni correction was applied to adjust for multiple comparisons.

4.5. Associations Between Physiological Measurements, Digit Span Performances and Subjective Ratings

We used Pearson correlations to infer the strength and direction of associations between the variables in our analyses. The bivariate correlations between alpha and theta amplitudes and digit span performance during the depletion phase were inconclusive, as were the associations between these brain oscillations and self-reported cognitive load. In addition, no significant correlations were found between heart rate and digit span accuracy or between heart rate and any of the pre-manipulation ZIPERS subscales. Heart rate did not appear to reliably predict any of the ZIPERS scores obtained after the environmental manipulation (Figure A1).

In the assessment phase, the bivariate correlation analyses also revealed no significant associations between self-reports, performance measurements and mean alpha and theta amplitudes in either condition. However, among participants in the natural condition, digit span accuracy correlated negatively with self-reported task difficulty, depicting a moderate effect ($r = -0.607, p < 0.001$), while in the urban condition, digit span accuracy showed a mild to moderate negative correlation with perceived mental effort ($r = -0.350, p = 0.05$). Correlation tables highlighting all the described relationships are reported in Appendix E.

5. Discussion

The main question this research aimed to answer was whether posterior alpha and frontal theta activities are sensitive to WM demands and whether these could be alleviated by observing natural scenery. The role of stress was also considered as a paradigm to possibly explain the effect. The motivation behind this research was to extend the current understanding of the role that natural and urban environments can play in learning and

memory performance, as proposed by ART by Kaplan (1995) and SRT by Ulrich (1981) [2,3]. The results indicated an absence of a significant difference between the group exposed to natural environments and the group exposed to urban environments across all measures of cognitive load. This suggests either a lack of the expected effect from the environmental type or excessive similarity between the two types of stimuli. It is possible that the selected scenario for the manipulation phase did not possess certain characteristics necessary for modulating directed attention. However, self-reported measures of stress highlighted a significant increase in positive affect in the natural environment condition compared to the urban environment condition. Differences in heart rate, fear arousal, attentive coping, anger and aggression and sadness remained below significance levels.

5.1. Cognitive Repletion Through Environmental Exposure, the Role of Brain Oscillations and Alternative Explanations

One of the practical aims of this paper was to bridge the findings that support the notion that exposure to a restorative environment, such as a natural environment, can benefit a range of cognitive functions and well-being [2,3,23] with those that highlight the role of alpha and theta oscillations in modulating these cognitive functions and emotional states, such as elevated stress [6,18,19]. The study design aimed to provide psychophysiological data as real-time measures of cognitive load and stress, thereby validating subjective reports typically used in most studies on the topic.

To understand the role of the alpha and theta frequencies in EEG spectrum analysis, we refer to the findings of Klimesch and colleagues (1999) [6] on the functional significance of these two waves. According to their review, alpha oscillations observed at posterior sites are associated with a cognitive state of “relaxed wakefulness” and tend to synchronize with decreased mental workload [6]. Conversely, theta oscillations, observable at frontal sites (particularly on the “frontal midline”), are most prominent when a subject is engaged in focused attention [6,19]. Understanding the function of these parameters is crucial for interpreting our results within the context of Kaplan’s ART (1995) [2]. Kaplan’s theory highlights the restorative characteristics of natural environments, proposing that these environments capture involuntary attention—a cognitive mechanism requiring no effort. When observers are in a state of effortless attention, their cognitive functions requiring focused effort are allowed to rest. However, the analysis of the spectral components of the EEG and heart rate in the present research did not reveal a significant restorative effect from exposure to a natural environment. These results stand in marked contrast to those reported in the existing literature [1,20,23,26].

Berman and colleagues (2008) [1] investigated the restorative role of nature on cognitive functioning, finding support for Kaplan’s ART through consistent differences on performances on a backward digit span task and restorative experience self-ratings between participants that observed natural scenery and those that did not. In a second study, conducted within the same research, they tested the assumption that the characteristic peacefulness of natural environments would bring about the restorative effect they observed. They reached the conclusion that peacefulness could not be retained as the decisive factor, as participants observing different scenes in the same laboratory settings registered significantly different scores on the dependent variables, highlighting, once again, the restorative characteristics of natural scenery. Similar findings were replicated by van Oordt et al. (2022) [23], who used research settings similar to the present study. Van Oordt and colleagues also aimed at corroborating Kaplan’s framework by testing WM capacity. They found significantly different performance on a digit span task between participants exposed to natural scenery and those who observed an urban or a third control setting. In the studies by Grassini et al. (2019) [26] and Chen et al. (2020) [20], these results were substantiated by EEG spectral analysis, in addition to participants’ self-reports and performance evaluations. Grassini and colleagues observed a positive increase in alpha wave power, associated with increased relaxation, when participants viewed static images of natural scenery. Conversely, the theta band showed a decrease in power while

viewing the same scenes, which they connected with a decreased state of arousal. Chen and associates (2020) [20] found an increased activity in the alpha band in the occipital lobes elicited by exposure to a restorative natural environment.

The results of the present research indicated that both the natural and urban manipulations were effective in reducing fear arousal, anger and aggression, while also increasing attentive coping, as demonstrated by the analysis of the ZIPERS scores. Additionally, the natural condition significantly increased positive affect, with participants in this condition reporting feeling more carefree, affectionate or elated and more likely to act friendly compared to participants in the urban condition [25]. Both conditions also appeared to decrease heart rate, an indicator of reduced stress arousal and anxiety [3]. However, definite conclusions about the causal relationship between environmental exposure and a reduction of stress cannot be drawn, due to the absence of a proper control group. The effect might have been brought about by simple rest from a working period. During the manipulation, in both the natural and the urban group, we were also able to observe an increase in alpha wave amplitude. To interpret this result in the context of our findings, we turn once again to previous studies investigating the functional significance of alpha wave suppression [3,6,28]. Ulrich (1981) [3] had used alpha wave amplitude as an operationalization of stress, describing it as a valid measure of cortical arousal, that correlates with states of alertness, showing a negative association with physiological arousal and feelings of anxiety. Following his definition, we could interpret the increase in alpha activity found during the manipulation phase as a reflection of a decreased state of anxiety and reduced physiological arousal. This interpretation is backed up by the results reported above of the ZIPERS stress questionnaire and heart rate measurement. It is, once again, important to keep in mind that it is, at present, not possible to attribute this effect to the exposition to environmental sceneries, as a no-condition control group was not included in the design. Klimesch (1999) [6] and Gevins and colleagues (1997) [28] found alpha wave amplitude to reflect directed attention and to be sensitive to WM demands and cognitive load. In order to reach a similar conclusion, we would have expected the increase in alpha amplitude observed during the manipulation phase to be accompanied by better performance on the second digit span block, as compared to the first, and a decrease in self-reported cognitive load, other than by a significant decrease in theta amplitude during said phase. The mentioned results were not observed in the present study. Hence, we cannot propose a decrease in cognitive load during the manipulation, despite the increase in alpha amplitude.

As mentioned in the opening section of the Discussion, we failed to replicate the positive effects of natural environments reported in previous research on the topic. The mixed ANOVAs that we used to compare our participants in the two different groups did not reveal any restorative effect of exposure to a natural environment, as opposed to an urban one. One possible explanation for the absence of this effect, aside from the argument that such an effect might not exist, is that environments may need to differ significantly on various parameters to elicit differential effects on stress and cognitive load. In his 1995 review of empirical findings in support of ART, Kaplan discusses several components that distinguish restorative environments from non-restorative ones. These components retained their importance in eliciting cognitive rest through the previously described mechanism. In the first place, nature provides a sense of “being away”; secondly, natural scenery elicits what is called “fascination”, which is characteristic of those environments that can capture attention but in an undramatic way. When in a state of fascination, the observer is relieved from effortful attention and is, then, free to think of other things. The third component listed by Kaplan is “extent”. Extent is typical of natural areas, even if they do not include vast extensions of land. Many natural environments contain elements (e.g., paths and trails) that provide the feeling of being in a much more extended area than one actually is. Kaplan calls this feeling a “connection to a larger world”. The fourth and last element that allows cognitive rest in a natural environment is referred to as “compatibility”. This last principle asserts that there exists a special resonance between natural settings and human beings. This is due to the fact that, for many people,

functioning in the natural environment is perceived as much less effortful than functioning in an urban setting, despite the greater familiarity with the latter [2]. A similar description of visual properties influencing aesthetic preference and interest is also found in a review by Ulrich (1983) [12]; these visual properties are thought to be related to more specific mechanisms occurring within the affective response process of people interacting with a given environment. Environments can differ in their complexity and structural properties. Complexity refers to the number of independently perceived elements in a scene [29]. Similarly, structural properties pertain to the extent to which an environment is structured in patterns. Another variable that enriches this framework is the notion of focality, that refers to the degree to which a scene contains a focal point [30]. Next, Ulrich describes ground surface texture as a landscaped feature that can modulate preference. Lastly, environments are categorized as containing, or not containing, elements of threat or risk. In terms of these properties, a particular environment should elicit higher preference if it shows moderate complexity (if complexity is too high or too low, preference tends to decline), has structural properties that include a focal point and other examples of patterning or clustering, the ground surface texture is homogeneous and elements of threat or tension are negligible or absent. In addition to these properties, preference can also increase as a function of depth, presence of a deflected vista and presence of water [12]. The material in the present study was produced and selected in order to prioritize ecological validity over a representation of two qualitatively distinct environments. The motivation that guided selection was to directly confront natural elements and urban ones to enquire whether the simple presence of these elements in an observed scene could bring about the desired replenishing effect. As a result, elements of threat or tension (e.g., traffic, large crowds, garbage, etc.) were omitted. Both conditions presented a diverse array of either urban or natural elements, with the intent of building up a more comprehensive picture sample. A side effect of this approach could have been the reduction of highly preferred elements (e.g., water) to a minimum or insufficient amount. Finally, picture complexity, depth and texture homogeneity were not inferred.

Ulrich had already hinted at the role that preference plays in the affective reaction to different types of environments in his paradigm-defining paper of 1981 [3]. On that occasion, he observed a difference in relaxed wakefulness, operationalized as alpha wave amplitudes, between participants observing vegetation and water scenes and those observing urban scenes. Ulrich was interested in observing differences in alpha amplitude as correlates of feelings of anxiety and stress arousal. However, our psychophysiological data, which appear to challenge Ulrich's conclusions, suggest that there is no differential effect of the type of environment on stress arousal. Our heart rate analysis, like Ulrich's research, showed no significant differences between groups, suggesting that physiological responses to environmental exposure may vary based on individual characteristics, previous emotional state, personality and daily experiences. This explanation finds support in the research of Zajonc (1980) [11], which proposes that affective reactions, like the ones considered in this framework, do not depend on cognition. According to Zajonc's theory [11], the first stage of response to a stimulus is characterized by global affects related to preferences. In other words, the quality and intensity of the affective reactions are subjected to the influence of internal states of the observer, together with previous experiences of the observer with stimuli of the same class [12]. With these facts in mind, we suggest that future research includes participants' preferences, as well as a more thorough examination of the aesthetic characteristics of the environment, in the theoretical model. Techniques that might be useful in guiding the aesthetic assessment include multiscale entropy (MSE) and monocular depth estimation (MDE). This approach may help clarify the complex interactions between environment type and affective responses.

5.2. Limitations

The first limitation to address is the lack of a true control group in the study design. Ideally, this control group would have been exposed to neutral scenery to test the causal

assumption that environmental settings can regulate stress arousal, as demonstrated by previous results [23]. The choice of a two-condition design in the present study was motivated by the expected sample size and power analysis. Furthermore, it was beyond the scope of this study to investigate the effects of specific components or aesthetic qualities of each environment, which might be central in evoking the hypothesized affective response and cognitive repletion. Instead, the methodological choice was to provide participants with ecologically valid stimuli. Since the sample was recruited from a population of students from the Dutch city of Rotterdam, the pictures selected for the manipulation were taken from the same area. The original idea was to test the restorative (or depleting) effects of the participants' habitual environment. We ensured that the pictures in both conditions did not contain elements that could influence affective reactions or appraisals through different mechanisms (e.g., people, traffic, advertisement), in an effort to directly contrast nature with urban settings. Future research should address the cognitive and affective responses elicited by scenery that differ in the domains described by Ulrich (1983) [12] and Kaplan (1995) [2]. Another methodological improvement be the use of techniques that allow observation of the specific elements participants pay attention to (e.g., eye tracking), to better understand how these elements affect the processes of interest.

6. Conclusions

In conclusion, the results showed that both urban and natural pictures elicited correlates of reduced stress and anxiety, as evidenced by a significant increase in alpha wave amplitude and a significant reduction in heart rate during picture viewing. However, we cannot conclude that these desirable effects were caused by the presentation of environmental pictures, as the study did not include a control group to test this hypothesis. Concerning the difference in activation between the urban and the natural group, we did not find any effect of type of environment on the restoration of WM. Although participants in the natural group reported feeling significantly more carefree, affectionate or elated, and more likely to act friendly than those in the urban group, we cannot recommend it for a general reduction in stress due to the inconclusiveness of the majority of the stress measures. Participants in the natural group did not perform better on a post-manipulation digit span task, nor did they report greater cognitive load or demonstrate greater stress, than the urban group, as hypothesized. Analyses of variance of heart rate and alpha/theta band power did not produce significant results with regard to the differences between the two groups, in contrast to findings from previous research [1,20]. We speculate that the hypothesized effects are only elicited by environments that differ on specific parameters (e.g., fascination, presence of threats). Future research is encouraged to investigate the influence of these parameters in more depth.

Author Contributions: Conceptualization, L.C., K.O. and F.P.; methodology, L.C., K.O. and F.P.; formal analysis, L.C.; investigation, L.C.; resources, L.C.; data curation, L.C. and K.O.; writing—original draft preparation, L.C. and K.O.; writing—review and editing, K.O. and F.P.; visualization, L.C.; supervision, K.O. and F.P.; project administration, L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of Erasmus University Rotterdam (protocol code ETH2223-0666, June 2023).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original data presented in the study are openly available in PsychArchives at <https://doi.org/10.23668/psycharchives.15015>.

Acknowledgments: We express our gratitude to the Erasmus Behavioral Lab for their support with preparing and conducting the research and Kristel de Groot for her assistance in the data analysis. We wish to honor the memory of our first author, Lorenzo Consalvi, whose talent and kind spirit

profoundly shaped this work. Lorenzo passed away shortly after completing the revisions for this manuscript, and we dedicate this article to his lasting contribution and enthusiasm for this field.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Baseline measurement for alpha and theta (log-transformed) and heart rate (bpm).

	Natural	Urban
	<i>M (SD)</i>	<i>M (SD)</i>
Alpha (eyes open)	1.35 (0.86)	1.58 (0.91)
Alpha (eyes closed)	2.00 (0.93)	2.23 (0.88)
Theta (eyes open)	1.33 (0.72)	1.37 (0.63)
Theta (eyes closed)	1.17 (0.68)	1.28 (0.64)
Heart rate	73.36 (9.52)	71.96 (9.11)

Appendix B

Table A2. Mean Differences in Alpha Amplitude Between the Different Phases.

Natural			Urban		
Phase (I)	Phase (J)	Mean Difference (I-J) ^b	Phase (I)	Phase (J)	Mean Difference (I-J)
1	2	−0.643 *	1	2	−0.665 *
	3	0.052		3	0.242 *
	4	−0.211 *		4	−0.067
	5	−0.011		5	0.156
2	1	0.643 *	2	1	0.665 *
	3	0.695 *		3	0.907 *
	4	0.432 *		4	0.598 *
	5	0.632 *		5	0.821 *
3	1	−0.052	3	1	−0.242 *
	2	−0.695 *		2	−0.907 *
	4	−0.264 *		4	−0.309 *
	5	−0.063		5	−0.086
4	1	0.211 *	4	1	0.067
	2	−0.432 *		2	−0.598 *
	3	0.264 *		3	0.309 *
	5	0.201 *		5	0.223 *
5	1	0.011	5	1	−0.156
	2	−0.632 *		2	−0.821 *
	3	0.063		3	0.086
	4	−0.201 *		4	−0.223 *

Note. The numbers correspond to the phases in order of completion (1 = baseline with eyes open; 2 = baseline with eyes closed; 3 = depletion; 4 = manipulation; 5 = assessment). *. Mean difference significant at the 0.05 level.

^b. Adjustment for multiple comparisons: Bonferroni.

Table A3. Mean Differences in Theta Amplitude Between the Different Phases.

Natural			Urban		
Phase (I)	Phase (J)	Mean Difference (I-J) ^b	Phase (I)	Phase (J)	Mean Difference (I-J)
1	2	0.162 *	1	2	0.096
	3	−0.177 *		3	−0.109 *
	4	−0.152 *		4	−0.089
	5	−0.215 *		5	−0.170 *
2	1	−0.162 *	2	1	−0.096
	3	−0.339 *		3	−0.205 *
	4	−0.314 *		4	−0.185 *
	5	−0.377 *		5	−0.266 *

Table A3. Cont.

Natural			Urban		
Phase (I)	Phase (J)	Mean Difference (I-J) ^b	Phase (I)	Phase (J)	Mean Difference (I-J)
3	1	0.177 *	3	1	0.109 *
	2	0.339 *		2	0.205 *
	4	0.025		4	0.020
	5	−0.038		5	−0.061
4	1	0.152 *	4	1	0.089
	2	0.314 *		2	0.185 *
	3	−0.025		3	−0.020
	5	−0.064		5	−0.081
5	1	0.215 *	5	1	0.170 *
	2	0.377 *		2	0.266 *
	3	0.038		3	0.061
	4	0.064		4	0.081

Note. The numbers correspond to the phases in order of completion (1 = baseline with eyes open; 2 = baseline with eyes closed; 3 = depletion; 4 = manipulation; 5 = assessment). *. Mean difference significant at the 0.05 level.
^b. Adjustment for multiple comparisons: Bonferroni.

Appendix C

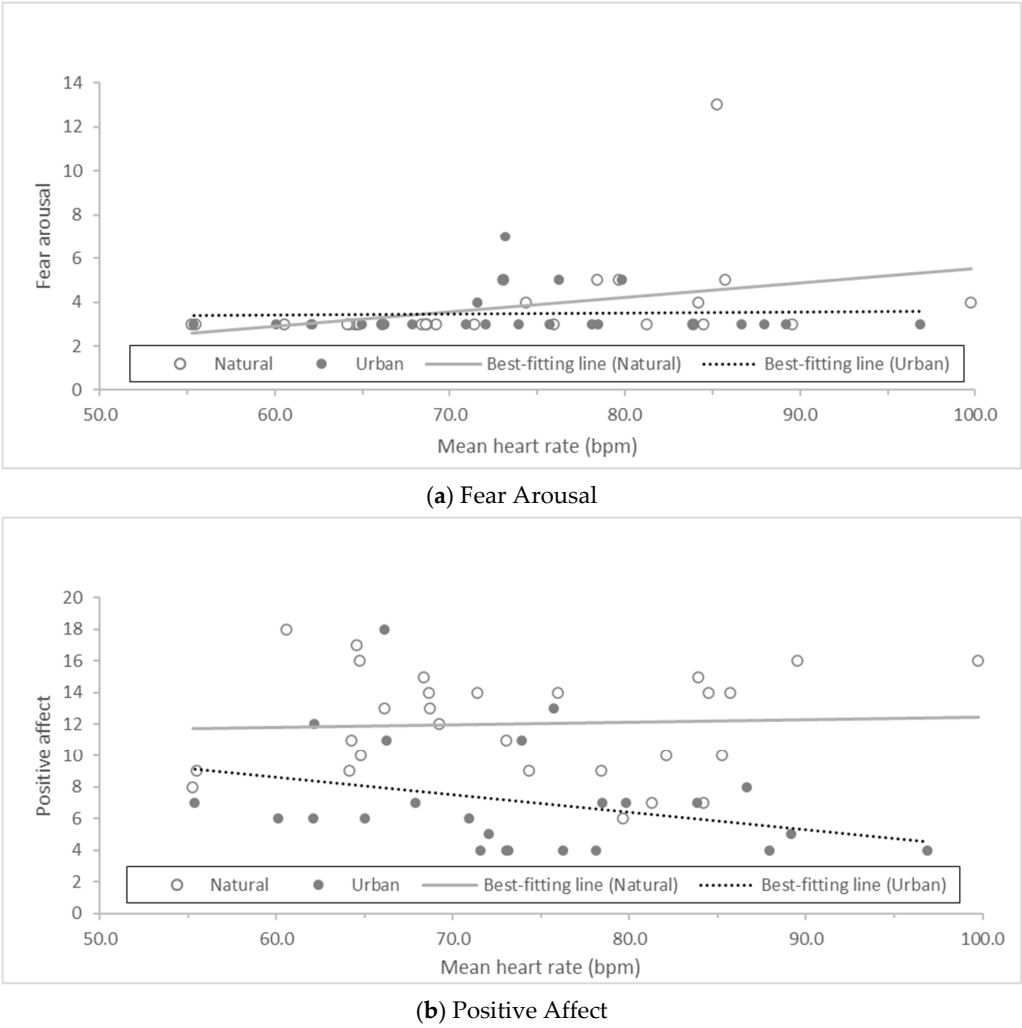
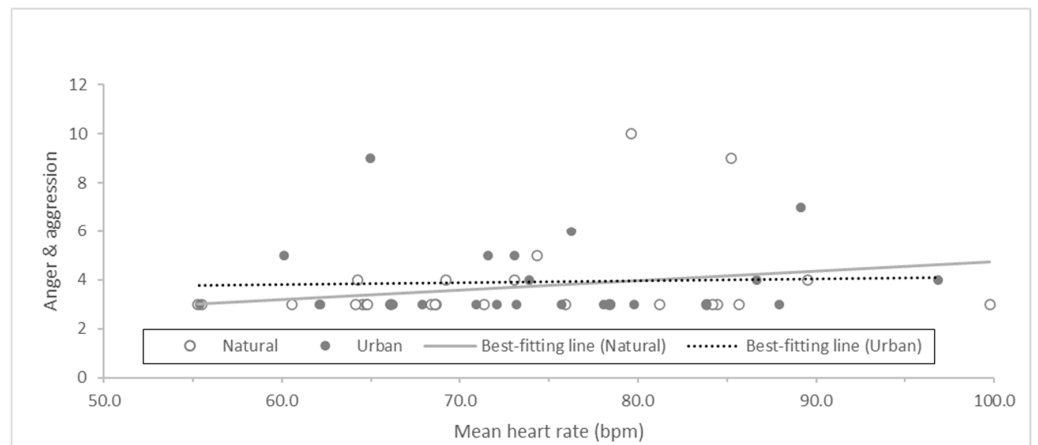
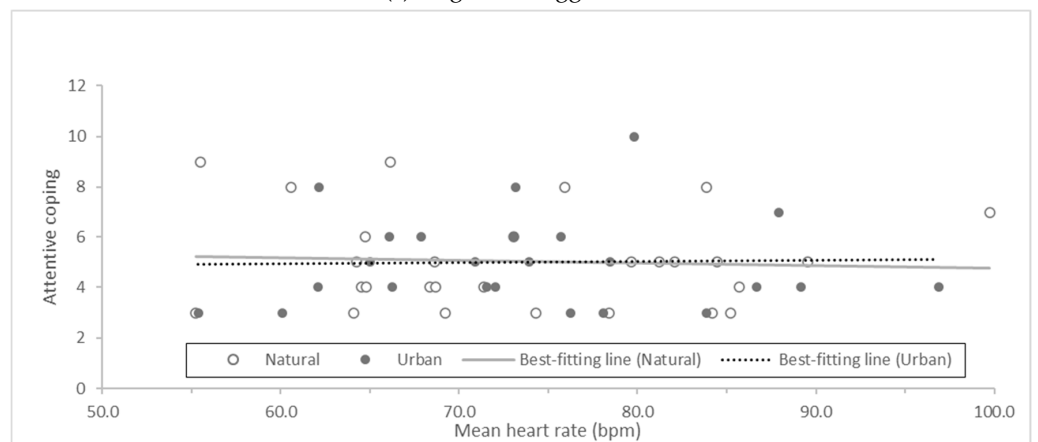


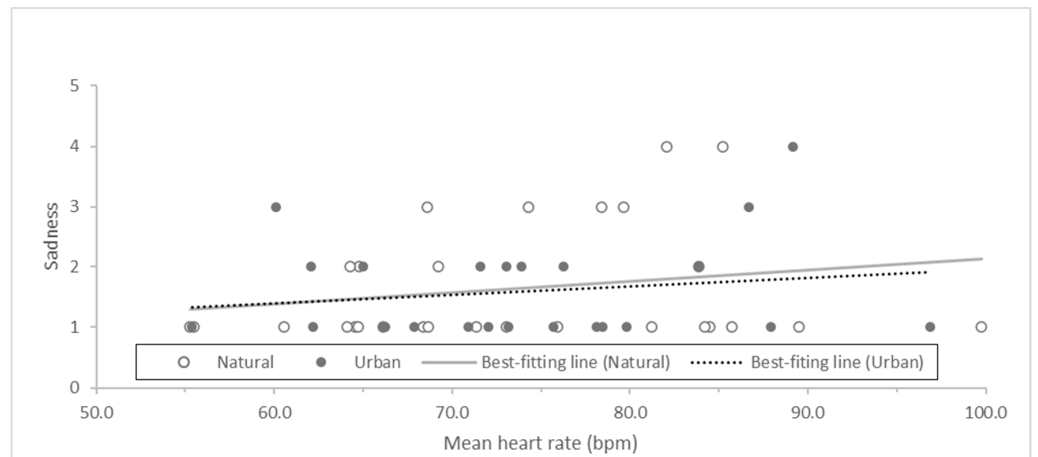
Figure A1. Cont.



(c) Anger and Aggression

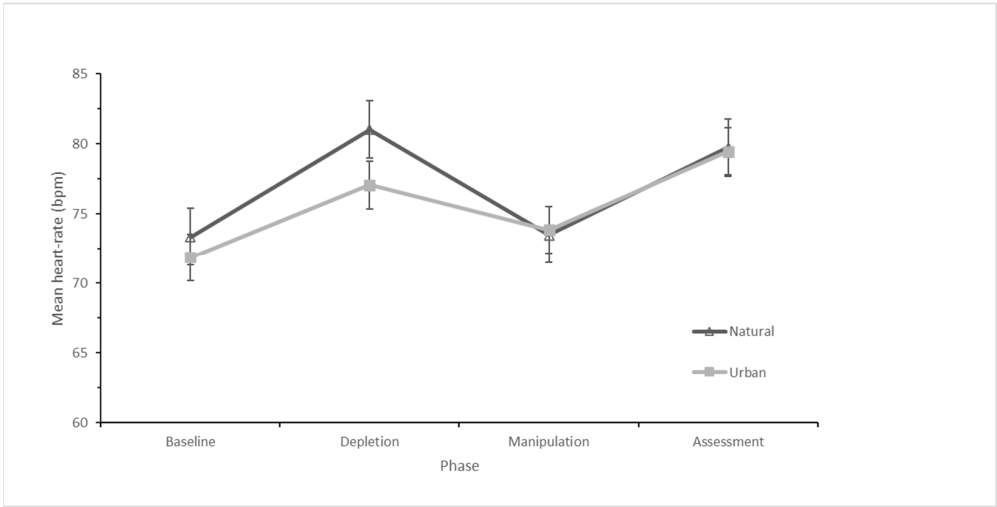


(d) Attentive Coping

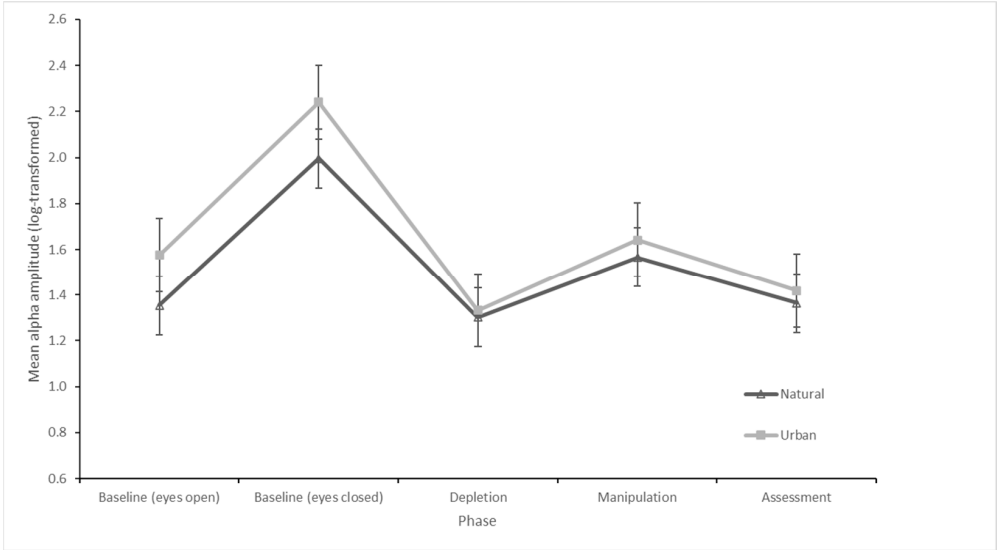


(e) Sadness

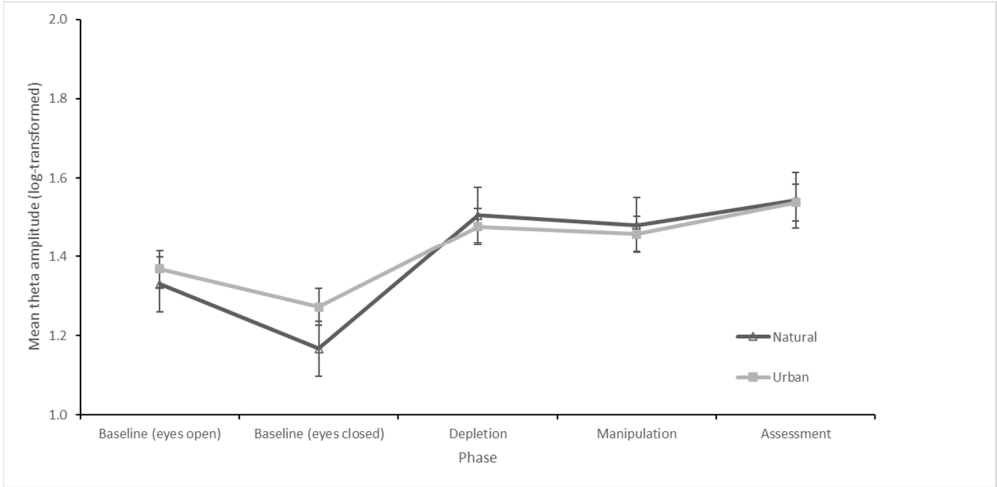
Figure A1. Linear Associations Between Heart Rate and Post-Manipulation ZIPERS Subscores.



(a) Changes in Heart Rate as a Function of Phase and Condition



(b) Changes in Alpha Amplitude as a Function of Phase and Condition



(c) Changes in Theta Amplitude as a Function of Phase and Condition

Figure A2. Note. Error bars represent standard errors of the mean.

Appendix D



Figure A3. Cont.



Figure A3. *Cont.*



Figure A3. Pictures used in the manipulation phase for the natural condition.

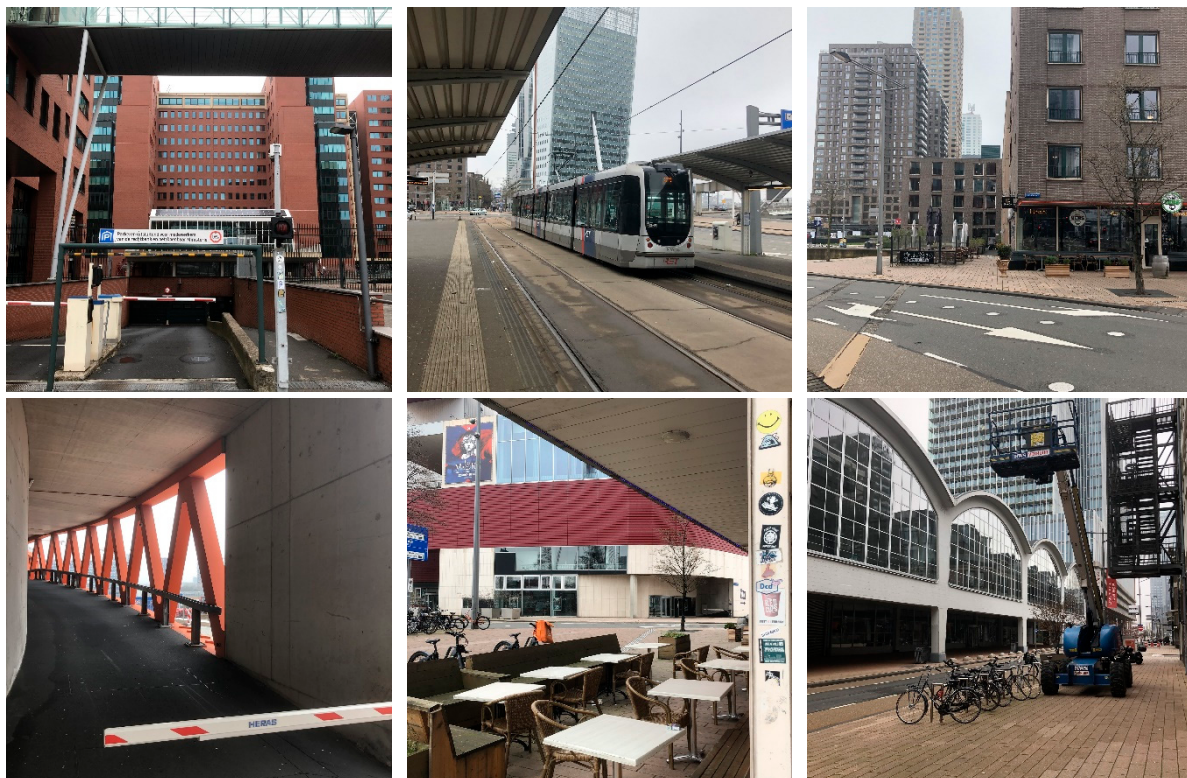


Figure A4. Cont.



Figure A4. Cont.



Figure A4. Pictures used in the manipulation phase for the urban condition.

Appendix E

Table A4. Pearson Correlations Between Alpha, Theta, Digit Span Performance and Cognitive Load Self-Reports in the Depletion Phase.

		Alpha	Theta	Digit Span	Mental Effort	Difficulty
Alpha	Pearson Correlation Sig. (2-tailed)	1	0.801 ** <0.001	−0.014 0.914	0.197 0.134	0.001 0.996
Theta	Pearson Correlation Sig. (2-tailed)	0.801 ** <0.001	1	−0.136 0.305	0.176 0.182	0.022 0.868
Digit span	Pearson Correlation Sig. (2-tailed)	−0.014 0.914	−0.136 0.305	1	−0.115 0.377	−0.283 * 0.027
Mental effort	Pearson Correlation Sig. (2-tailed)	0.197 0.134	0.176 0.182	−0.115 0.377	1	0.639 ** <0.001
Difficulty	Pearson Correlation Sig. (2-tailed)	0.001 0.996	0.022 0.868	−0.283 * 0.027	0.639 ** <0.001	1

* Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed).

Table A5. Pearson Correlation Between Heart Rate, Digit Span Performance and ZIPERS Subscales Scores in the Depletion Phase.

		Heart Rate	Digit Span	Fear Arousal	Positive Affect	Anger and Aggression	Attention	Sadness
Heart Rate	Pearson Correlation Sig. (2-tailed)	1	0.016 0.911	0.072 0.612	0.093 0.512	−0.056 0.691	−0.041 0.773	0.016 0.909
Digit Span	Pearson Correlation Sig. (2-tailed)	0.016 0.911	1	−0.042 0.748	−0.006 0.961	−0.288 * 0.025	0.421 ** 0.001	−0.283 * 0.027
Fear Arousal	Pearson Correlation Sig. (2-tailed)	0.072 0.612	−0.042 0.748	1	0.088 0.501	0.234 0.070	−0.015 0.910	0.169 0.194
Positive Affect	Pearson Correlation Sig. (2-tailed)	0.093 0.512	−0.006 0.961	0.088 0.501	1	−0.069 0.599	0.057 0.661	−0.003 0.984
Anger and Aggression	Pearson Correlation Sig. (2-tailed)	−0.056 0.691	−0.288 * 0.025	0.234 0.070	−0.069 0.599	1	−0.426 ** 0.001	0.287 * 0.025
Attention	Pearson Correlation Sig. (2-tailed)	−0.041 0.773	0.421 ** 0.001	−0.015 0.910	0.057 0.661	−0.426 ** 0.001	1	−0.430 ** 0.001
Sadness	Pearson Correlation Sig. (2-tailed)	0.016 0.909	−0.283 * 0.027	0.169 0.194	−0.003 0.984	0.287 * 0.025	−0.430 ** 0.001	1

* Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed).

Table A6. Pearson Correlation Between Heart Rate and ZIPERS Subscales Scores in the Manipulation Phase.

		Heart Rate	Fear Arousal	Positive Affect	Anger and Aggression	Attention	Sadness
Heart Rate	Pearson Correlation Sig. (2-tailed)	1	0.253 0.074	−0.099 0.485	0.160 0.262	−0.022 0.874	0.191 0.176
Fear Arousal	Pearson Correlation Sig. (2-tailed)	0.253 0.074	1	−0.162 0.216	0.499 ** <0.001	−0.076 0.565	0.424 ** 0.001
Positive Affect	Pearson Correlation Sig. (2-tailed)	−0.099 0.485	−0.162 0.216	1	−0.322 * 0.012	0.237 0.066	−0.236 0.067
Anger and Aggression	Pearson Correlation Sig. (2-tailed)	0.160 0.262	0.499 ** <0.001	−0.322 * 0.012	1	−0.222 0.089	0.662 ** <0.001

Table A6. Cont.

		Heart Rate	Fear Arousal	Positive Affect	Anger and Aggression	Attention	Sadness
Attention	Pearson Correlation	−0.022	−0.076	0.237	−0.222	1	−0.290 *
	Sig. (2-tailed)	0.874	0.565	0.066	0.089		0.024
Sadness	Pearson Correlation	0.191	0.424 **	−0.236	0.662 **	−0.290 *	1
	Sig. (2-tailed)	0.176	0.001	0.067	<0.001	0.024	

* Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed).

Table A7. Pearson Correlations Between Alpha, Theta, Digit Span Performance and Cognitive Load Self-Reports in the Assessment Phase (Divided by Condition).

	Condition		Alpha	Theta	Digit Span	Mental Effort	Difficulty
Natural	Alpha	Pearson Correlation	1	0.800 **	0.013	−0.095	−0.093
		Sig. (2-tailed)		<0.001	0.949	0.630	0.636
	Theta	Pearson Correlation	0.800 **	1	−0.101	−0.042	−0.133
		Sig. (2-tailed)	<0.001		0.609	0.832	0.501
	Digit Span	Pearson Correlation	0.013	−0.101	1	−0.129	−0.337
		Sig. (2-tailed)	0.949	0.609		0.504	0.074
	Mental Effort	Pearson Correlation	−0.095	−0.042	−0.129	1	0.709 **
		Sig. (2-tailed)	0.630	0.832	0.504		<0.001
	Difficulty	Pearson Correlation	−0.093	−0.133	−0.337	0.709 **	1
		Sig. (2-tailed)	0.636	0.501	0.074	<0.001	
Urban	Alpha	Pearson Correlation	1	0.813 **	−0.054	0.403 *	0.058
		Sig. (2-tailed)		<0.001	0.772	0.025	0.755
	Theta	Pearson Correlation	0.813 **	1	−0.220	0.356 *	0.138
		Sig. (2-tailed)	<0.001		0.235	0.049	0.460
	Digit Span	Pearson Correlation	−0.054	−0.220	1	−0.162	−0.279
		Sig. (2-tailed)	0.772	0.235		0.377	0.122
	Mental Effort	Pearson Correlation	0.403 *	0.356 *	−0.162	1	0.587 **
		Sig. (2-tailed)	0.025	0.049	0.377		<0.001
	Difficulty	Pearson Correlation	0.058	0.138	−0.279	0.587 **	1
		Sig. (2-tailed)	0.755	0.460	0.122	<0.001	

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

References

- Berman, M.G.; Jonides, J.; Kaplan, S. The cognitive benefits of interacting with nature. *Psychol. Sci.* **2008**, *19*, 1207–1212. [CrossRef] [PubMed]
- Kaplan, S. The restorative benefits of nature: Toward an integrative framework. *J. Environ. Psychol.* **1995**, *15*, 169–182. [CrossRef]
- Ulrich, R.S. Natural versus urban scenes: Some psychophysiological effects. *Environ. Behav.* **1981**, *13*, 523–556. [CrossRef]
- Ulrich, R.S.; Simons, R.F.; Losito, B.D.; Fiorito, E.; Miles, M.A.; Zelson, M. Stress recovery during exposure to natural and urban environments. *J. Environ. Psychol.* **1991**, *11*, 201–230. [CrossRef]
- Antonenko, P.; Paas, F.; Grabner, R.; van Gog, T. Using electroencephalography to measure cognitive load. *Educ. Psychol. Rev.* **2010**, *22*, 425–438. [CrossRef]
- Klimesch, W. EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Res. Rev.* **1999**, *29*, 169–195. [CrossRef] [PubMed]
- Sweller, J.; van Merriënboer, J.J.; Paas, F. Cognitive architecture and instructional design: 20 years later. *Educ. Psychol. Rev.* **2019**, *31*, 261–292. [CrossRef]
- Chen, O.; Castro-Alonso, J.C.; Paas, F.; Sweller, J. Extending cognitive load theory to incorporate working memory resource depletion: Evidence from the spacing effect. *Educ. Psychol. Rev.* **2018**, *30*, 483–501. [CrossRef]
- Kaplan, S.; Berman, M.G. Directed attention as a common resource for executive functioning and self-regulation. *Perspect. Psychol. Sci.* **2010**, *5*, 43–57. [CrossRef]
- Lazarus, R.S.; Kanner, A.D.; Folkman, S. Emotions: A cognitive–phenomenological analysis. *Theor. Emot.* **1980**, *1980*, 189–217. [CrossRef]
- Zajonc, R.B. Feeling and thinking: Preferences need no inferences. *Am. Psychol.* **1980**, *35*, 151–175. [CrossRef]
- Ulrich, R.S. Aesthetic and affective response to natural environment. *Behav. Nat. Environ.* **1983**, *6*, 85–125. [CrossRef]

13. Ratcliffe, E.; Gatersleben, B.; Sowden, P.T. Bird sounds and their contributions to perceived attention restoration and stress recovery. *J. Environ. Psychol.* **2013**, *36*, 221–228. [CrossRef]
14. Cowan, N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behav. Brain Sci.* **2001**, *24*, 87–114. [CrossRef]
15. Diamond, A. Executive functions. *Annu. Rev. Psychol.* **2013**, *64*, 135–168. [CrossRef]
16. Gazzaley, A.; Nobre, A.C. Top-down modulation: Bridging selective attention and working memory. *Trends Cogn. Sci.* **2012**, *16*, 129–135. [CrossRef]
17. Klein, K.; Boals, A. The relationship of life event stress and working memory capacity. *Appl. Cogn. Psychol. Off. J. Soc. Appl. Res. Mem. Cogn.* **2001**, *15*, 565–579. [CrossRef]
18. Klimesch, W.; Schack, B.; Sauseng, P. The functional significance of theta and upper alpha oscillations. *Exp. Psychol.* **2005**, *52*, 99–108. [CrossRef]
19. Inanaga, K. Frontal midline theta rhythm and mental activity. *Psychiatry Clin. Neurosci.* **1998**, *52*, 555–566. [CrossRef]
20. Chen, Z.; He, Y.; Yu, Y. Attention restoration during environmental exposure via alpha-theta oscillations and synchronization. *J. Environ. Psychol.* **2020**, *68*, 101406. [CrossRef]
21. Bowler, D.E.; Buyung-Ali, L.M.; Knight, T.M.; Pullin, A.S. A systematic review of evidence for the added benefits to health of exposure to natural environments. *BMC Public Health* **2010**, *10*, 1–10. [CrossRef] [PubMed]
22. Kirschbaum, C.; Pirke, K.M.; Hellhammer, D.H. The ‘Trier Social Stress Test’—A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* **1993**, *28*, 76–81. [CrossRef] [PubMed]
23. Van Oordt, M.; Ouwehand, K.; Paas, F. Restorative effects of observing natural and urban scenery after working memory depletion. *Int. J. Environ. Res. Public Health* **2022**, *20*, 188. [CrossRef] [PubMed]
24. Paas, F. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *J. Educ. Psychol.* **1992**, *84*, 429–434. Available online: <https://psycnet.apa.org/doi/10.1037/0022-0663.84.4.429> (accessed on 1 June 2023). [CrossRef]
25. Zuckerman, M. Development of a situation-specific trait-state test for the prediction and measurement of affective responses. *J. Consult. Clin. Psychol.* **1977**, *45*, 513–523. Available online: <https://psycnet.apa.org/doi/10.1037/0022-006X.45.4.513> (accessed on 1 June 2023). [CrossRef]
26. Grassini, S.; Revonsuo, A.; Castellotti, S.; Petrizzo, I.; Benedetti, V.; Koivisto, M. Processing of natural scenery is associated with lower attentional and cognitive load compared with urban ones. *J. Environ. Psychol.* **2019**, *62*, 1–11. [CrossRef]
27. Stipacek, A.; Grabner, R.H.; Neuper, C.; Fink, A.; Neubauer, A.C. Sensitivity of human EEG alpha band desynchronization to different working memory components and increasing levels of memory load. *Neurosci. Lett.* **2003**, *353*, 193–196. [CrossRef] [PubMed]
28. Gevins, A.; Smith, M.E.; McEvoy, L.; Yu, D. High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cereb. Cortex* **1997**, *7*, 374–385. [CrossRef]
29. Berlyne, D.E. Aesthetics and psychobiology. *J. Aesthet. Art Crit.* **1973**, *31*, 553. [CrossRef]
30. Ulrich, R.S. Visual landscape preference: A model and application. *Man-Environ. Syst.* **1977**, *7*, 279–293. Available online: <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc2&NEWS=N&AN=1979-31072-001> (accessed on 1 June 2023).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

The Moderating Role of Interest in the Relationship between Perceived Task Difficulty and Invested Mental Effort

Katrin Schuessler ^{1,*}, Vanessa Fischer ¹, Maik Walpuski ¹ and Detlev Leutner ²

¹ Institute for Chemistry Education, Faculty of Chemistry, University of Duisburg-Essen, Schützenbahn 70, 45127 Essen, Germany; vanessa.fischer@uni-due.de (V.F.); maik.walpuski@uni-due.de (M.W.)

² Instructional Psychology Department, Faculty of Education Sciences, University of Duisburg-Essen, Universitätsstr. 2, 45141 Essen, Germany; detlev.leutner@uni-due.de

* Correspondence: katrin.schuessler@uni-due.de

Abstract: Including motivational variables such as interest in the cognitive load framework is an ongoing process. Of particular interest is the question of how motivational variables influence the investment of mental effort. In this study, we investigated how topic interest affects the investment of mental effort in simple tasks. A total of 1543 students' judgments regarding invested mental effort, perceived task difficulty, and topic interest for 32 tasks of a chemistry test were analyzed at the task level based on item response theory parameters. Additionally, objective task difficulty was calculated. The Rasch parameters were used for correlation and moderated regression analyses. The results indicated that when perceived task difficulty was low, students invested more mental effort in solving tasks of low topic interest compared to tasks of high topic interest. With increasing perceived task difficulty, the amount of invested mental effort rose for tasks of low as well as high topic interest. However, the difference between tasks of low and high topic interest in the amount of invested mental effort decreased as perceived task difficulty increased and even vanished when perceived task difficulty roughly corresponded to students' performance capability. These results are in line with flow theory and the expectancy-value-cost model of motivation. When solving tasks that match their performance capability, students can experience a flow situation. However, when solving rather easy tasks of low interest, students can experience motivational costs in terms of additional effort, such as an increased need for motivational self-regulation. The results of this study provide a basis for systematically investigating and better understanding the relationship between interest, task difficulty, invested mental effort, flow experience, and emotional costs.

Keywords: cognitive load; invested mental effort; perceived task difficulty; topic interest; objective task difficulty; emotional costs; motivational self-regulation; chemistry test

Citation: Schuessler, K.; Fischer, V.; Walpuski, M.; Leutner, D. The Moderating Role of Interest in the Relationship between Perceived Task Difficulty and Invested Mental Effort. *Educ. Sci.* **2024**, *14*, 1044. <https://doi.org/10.3390/educsci14101044>

Academic Editor: Luca Tateo

Received: 29 June 2024

Revised: 20 September 2024

Accepted: 22 September 2024

Published: 24 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Cognitive load theory has long focused primarily on the cognitive aspects of learning. In recent years, however, several attempts have been made to include affective aspects, that is, motivational and emotional processes [1–4]. The present study contributes to the growing literature including motivational processes in the cognitive load framework by investigating whether topic interest has an impact on the investment of mental effort. We used a large data set ($N = 1543$) collected based on a multi-matrix design to evaluate a chemistry test at the end of lower secondary school. In addition to the chemistry tasks to be solved, the test featured additional questionnaire items regarding cognitive load and topic interest for each of the tasks [5]. All our follow-up analyses were performed on the level of tasks, with the units of analysis being the tasks and their characteristics (as opposed to participants and their characteristics). The analyses were based on ideas from cognitive load theory [6,7], flow theory [8], and expectancy-value-cost theory [9], suggesting that topic interest moderates the regression of invested mental effort on perceived task difficulty.

In the following, we will first present a short introduction to cognitive load theory before briefly introducing the core ideas of flow theory and the expectancy-value-cost model of motivation. Finally, we will discuss recent approaches to extend cognitive load theory to include motivational processes.

1.1. Cognitive Load Theory

Cognitive load theory [6,7,10–13] is based on the assumption that working memory, unlike long-term memory, has a limited capacity. Furthermore, information must be actively processed in working memory to be successfully stored in long-term memory. However, processing information in working memory causes cognitive load. Depending on whether this load contributes to learning or not, cognitive load theory distinguishes between intrinsic and germane cognitive load on the one hand and extraneous cognitive load on the other [12–19]. Intrinsic load directly contributes to learning and depends on the complexity of the learning content, that is, the number of interacting elements that are to be processed simultaneously in working memory (element interactivity). Germane cognitive load directly contributes to learning as well and depends on the specific way the elements are processed in terms of generative learning to construct a coherent mental representation of the learning content. Extraneous cognitive load, however, is based on cognitive processing that does not directly contribute to learning and can be caused, for example, by poor instructional design. There is ongoing discussion about whether germane cognitive load actually represents a type of load that is independent of intrinsic cognitive load [18]. Although a number of multidimensional cognitive load instruments seem to be able to measure germane cognitive load over and above intrinsic and extraneous cognitive load [15,20,21], a recent study found no correlation between germane cognitive load and immediate and delayed learning outcomes [22]. In said study, only intrinsic cognitive load was found to be negatively related to performance, which indicates that cognitive load is a two-dimensional concept. Working memory resources rather than germane cognitive load proved to be a relevant factor for learning success in situations with high intrinsic load. The authors of the study thus suggested investigating the extent to which affective factors are relevant to the availability of working memory resources.

The types of cognitive load add up to the total cognitive load imposed on working memory (additivity hypothesis [23]); when the total cognitive load exceeds the limited capacity of working memory, learning performance is reduced. Cognitive load theory therefore provides guidance on how extraneous cognitive load can be reduced so that sufficient capacity for germane cognitive load is available for a given intrinsic cognitive load. Cognitive load theory was originally developed to describe learning processes. However, the theory can be applied to all other processes involving working memory and the cognitive processing of information. One example is students' processing of the tasks of a subject matter test, such as a test on chemistry knowledge and skills.

Using questionnaire items, overall cognitive load can be measured in terms of perceived task difficulty and invested mental effort [14,16,24–27]. Perceived task difficulty depends on the complexity of the task and the availability of prior knowledge for processing the task, whereas invested mental effort denotes the cognitive capacity allocated to cope with the resulting demands. When tasks are not too difficult, perceived task difficulty and invested mental effort are positively correlated. That is, when task difficulty increases, invested mental effort increases as well (up to the point when a learner or problem solver gives up because a task is perceived as too difficult to be mastered [28,29]).

1.2. Flow Theory

Flow theory describes flow as a state in which a person's inherent abilities and the perceived challenge of an activity are in equilibrium [8,30]. In this case, the challenge posed by the activity is greater than the challenge posed by everyday activities but still appears manageable. The flow state creates harmony within the self so that attention is completely focused on the activity; there is simply no cognitive capacity left to focus on

other things. All attention is focused on the requirements of the activity and on what needs to be completed next. There is no room for doubt, uncertainty, and extraneous thoughts. The sense of time is also lost in this flow state. Flow experiences often occur in well-structured areas, such as skiing in powder snow, motorcycling on a mountain pass road, or computer programming. Such activities require clear goals and offer relatively short-term and unambiguous feedback, so skills and challenges can be easily controlled and varied. From a motivational point of view, flow experiences are usually actively sought out because the state itself is perceived as desirable. To maintain flow, the complexity of the activity must increase consistently so that new challenges are found and new skills can be developed (dynamic of flow). When skills and perceived challenges are not in balance, anxiety (high challenge, low skills) or relaxation/boredom (low challenge, high skills) can occur. In both cases, less than the full amount of cognitive capacity is devoted to the ongoing activity.

In terms of cognitive load theory, flow describes a situation in which the investment of mental effort is exclusively determined by the perceived difficulty of the task. That is, other motivational processes beyond flow play a subordinate role in the investment of mental effort in such a task. However, if the skills and challenges associated with a task are out of balance—leading, for example, to the experience of anxiety or boredom—other motivational processes come into play [8,31]. These include motivational self-regulation to cope, for instance, with reduced expectancy (in the case of anxiety), reduced value (in the case of boredom), and increased costs (in the case of both anxiety and boredom).

1.3. The Expectancy-Value-Cost Model of Motivation

According to the expectancy-value-cost model of motivation [9], motivation is composed of three components: expectancy (“Can I complete the task?”), value (“Do I want to complete the task?”), and cost (“What are the costs associated with completing the task and am I prepared to bear them?”). While expectation and value are positively related to each other, both are negatively related to cost [31–33].

All three components can be further differentiated. For example, interest is considered one of three sub-components of value (alongside achievement value and utility value) [34]. Interest represents the willingness to engage with a certain content or topic. How pronounced and stable this willingness is depends on how strong the interest already is in terms of repeated engagement and acquired knowledge. At higher levels of developed interest, the individual looks for opportunities to re-engage with the topic of their interest [34]. In a review article, Hidi [35] points out that research often distinguishes between situational interest and individual interest. Increased value, knowledge, and positive feelings characterize individual interest. Individual interest is described as a stable factor that develops over time. In contrast, situational interest arises spontaneously through interaction with the environment. Topic interest is described as another relevant interest construct for research and can be generated by both existing individual interest in a topic and spontaneously aroused situational interest in the topic. Regardless of whether topic interest can be traced back to individual interest or situational interest, it is described as relevant to affects and, thus, to persistence in learning and learning outcomes [35].

Regarding costs, negative perception seems to be a crucial aspect [31]. For example, although students described effort as a motivating factor for the class for which they were most motivated, excessive effort was a reason for disliking the class for which they were least motivated. However, too little effort was also cited as a de-motivating factor for their least motivated class. According to Flake et al. [31], and in line with flow theory, the two de-motivating factors for the least motivated classes are the perception of excessive effort (“This course is too demanding”) and the perception of too little effort (“This course is too boring”). Both signal an imbalance between skills and challenges, leading to emotional costs such as anxiety or boredom. In general, emotional costs correlate negatively with long-term topic interest, and costs have been shown to be a relevant predictor of performance, procrastination, and avoidance intentions in mathematics at the secondary school level [32],

among others. A recent study showed that the expectancy-value-cost model explains achievement better than an expectancy-value model in some cases; the study emphasizes that the situational context must be taken into account when investigating motivation [33].

Cognitive load, particularly extraneous cognitive load, can be considered a source of motivational cost [1] that is relevant to learners' motivation to solve similar tasks in the future and their expectation of success in solving such tasks. In their study, Feldon et al. investigated how the amount of (extraneous) cognitive load during instruction influenced participants' post-instruction self-efficacy [36]. Their findings suggest that self-efficacy is not only influenced by students' achievement in previous tasks [37] but also by the cognitive load imposed on them by the previous tasks, regardless of whether they achieved them [38].

1.4. Motivational Factors in the Cognitive Load Framework

When Paas [39] proposed the mental effort rating scale, he considered motivation a relevant factor for mental effort investment but assumed that the participants in his study were sufficiently and equally motivated so that motivational effects were negligible. Among others, concerns about high dropout rates in online learning courses led to a more serious consideration of motivation within the cognitive load framework. Paas et al. argued that "the motivation to achieve well" [40] (p. 26) is necessary, specifically for learning scenarios outside a laboratory and with a longer duration. Therefore, they recommended capturing learners' motivation in addition to invested mental effort and performance. They further stated that "[a]s long as a task is not too easy and not too difficult, ratings of task difficulty may correlate highly with ratings of invested mental effort. Most importantly, it is clear that mental effort is a voluntary mobilization process of resources, which depends on the task demands in relation to the amount of resources the learner is willing or able to allocate" [40] (p. 32).

Following these first thoughts, the Cognitive and Affective Theory of Learning with Media (CATLM; [4,41,42]) extended existing cognitive theories of learning with media by adding the explicit assumption that motivational factors influence learning by affecting cognitive engagement (a similar model was presented even earlier [43]). Mayer [44,45] also complemented his Cognitive Theory of Multimedia Learning by adding assumptions about motivational processes: motivation is described in terms of "the learner's willingness to exert effort to engage in appropriate cognitive processing during learning" [45] (p. 70), with interest as one of the driving forces of the learner's motivation (in addition to affect, beliefs, and feelings of social connection).

In addition to motivation, another approach to increasing learners' engagement is to focus on emotional factors in the design of learning tasks (emotional design). Some relevant studies varied the learning materials (for example, by introducing warm colors), which was effective in improving learning [46–48]. Other studies added elements to the learning materials (seductive details such as pointing out the usefulness of ATP (adenosine triphosphate) in different areas, such as sports in a biology lesson on ATP); however, this was generally not effective in improving learning [49,50].

In the literature, motivational and emotional factors are discussed as mediators or moderators in predicting students' cognitive engagement and learning [2]. On the one hand, motivation (or emotion) can mediate the effect of a specific instructional design factor (e.g., using warm colors) on invested mental effort and learning outcomes. On the other hand, motivation (or emotion) can moderate, for instance, the influence of perceived task difficulty on invested mental effort (as investigated in the present study). If, as an example of moderation, low motivation leads to the need to invest more mental effort than expected, this need would represent a source of additional cognitive load [3]. This can also be assumed concerning emotional effects. Thus, whereas Feldon and colleagues [1,36,38] investigated motivation as an *outcome* of cognitive load, in the present study we investigated (low) motivation as a *source* of (additional) cognitive load.

1.5. Research Question and Hypothesis

The present study investigates the following research question: does interest moderate the relationship between perceived task difficulty and invested mental effort? For any particular student who is working on a test at school, only a small number of test items or tasks will exactly match their skills; the majority of tasks will either be too easy or too difficult. According to flow theory, working on a task that matches a student's skills will cause a state of flow. In this state of flow, the entire cognitive capacity of the student is devoted to processing the task. However, facing a task that is perceived as too easy compared to the student's skills will cause relaxation or even boredom or disengagement. Nevertheless, assuming that a student taking a test at school is compliant and wants to perform well, they will not disengage and will persist in processing the task at hand (instead of deciding to stop), which might even create a feeling of anger. Persistence when feeling boredom and anger requires mental effort in terms of motivational self-regulation, in addition to the mental effort required for completing the task. However, this need to invest additional mental effort to cope with the emotional costs of boredom and anger should not be challenging because sufficient cognitive capacity is available given the ease of purely processing the task. In this situation, interest comes into play. If a student is highly interested in the topic of the task, this topic interest (i.e., the wish to engage with the content) will prevent them from feeling bored (and, perhaps, even angered) even though the task is not challenging. Consequently, compared to the uninterested student, the interested student will not need to invest additional mental effort into motivational self-regulation to cope with emotional costs. In other words, high topic interest, unlike low topic interest, will compensate for coping with emotional cost when processing tasks of low perceived difficulty.

Accordingly, we formulated the following hypothesis for our study: Topic interest moderates the effect of perceived task difficulty on invested mental effort, such that simple tasks that are perceived as interesting are associated with less invested mental effort than simple tasks that are not perceived as interesting. In line with flow theory, we assume that skills and challenges are not in balance for easy tasks. For interesting easy tasks, only a small amount of mental effort must be invested. In contrast, for uninteresting easy tasks, a comparatively high level of mental effort must be invested to cope with the additional emotional costs that are associated with processing the task.

2. Materials and Methods

2.1. The Original Study

The objective of the original study was to develop a test to assess students' chemistry content knowledge and their abilities to make chemistry-related decisions at the end of lower secondary school. Based on this assessment, the study aimed to determine the influence of affective factors, such as interest in the task and motivation when working on the task, on task difficulty. The participants were 1543 students (51.7% male, $M_{age} = 15.2$, $SD = 0.87$) in their final year of German lower secondary schools in 2015 [5]. Chemistry teachers and school headmasters decided to participate in the study with their classes, but every student and their parents had the chance to refuse participation without consequences. Although institutional review board approval is not necessary at German schools, the study design complied with the recommendations of the German Research Foundation for Good Scientific Practice. The data were collected in 2015 in German schools during regular chemistry lessons near the end of lower secondary school. Paper-pencil materials and a multi-matrix design were used to administer the 32 tasks of the test. According to this multi-matrix design, each student worked on two of the 32 chemistry tasks, resulting in approximately 96 students per task. Each task consisted of four items. Half of the tasks were developed to measure the competence area of content knowledge and the other half to measure the competence area of decision-making. The tasks were developed based on the educational standards for the subject of chemistry in lower secondary schools in Germany [51]. The tasks only required prior knowledge, which should have been acquired

on the basis of the educational standards in chemistry lessons. The information required to answer the task was presented to the students in the respective task stem. The items about dealing with specialist knowledge required the students to select, organize, or integrate information from the task stem. Before running the study, the psychometric quality of the test items was evaluated in a pre-study. Among others, a distractor analysis was carried out, and all distractors with a response frequency under 5% were replaced with suitable distractors.

Students' answers to each item were coded as correct (1) or incorrect (0). ConQuest was used for item response theory (IRT) analysis. The analysis was conducted on a person-centered basis to obtain IRT parameters at the item level. A two-dimensional Rasch model (Dimension 1: content knowledge; Dimension 2: decision-making) showed a good statistical fit to the data (item reliability = 0.96, $0.85 \leq \text{MNSQ} \leq 1.17$). The task difficulty parameter ($-2.59 \leq \text{item difficulty content knowledge} \leq 1.40$; $-1.85 \leq \text{item difficulty decision-making} \leq 2.67$) was determined as the mean of the item difficulty parameters for each task ($M_{\text{content knowledge}} = -0.63$; $M_{\text{decision-making}} = 0.83$). The instruments used to measure students' skills in the two areas of competence exhibited good statistical characteristics ($0.21 \leq \text{item discrimination} \leq 0.80$). Only one item from the content knowledge competence area had a discriminatory power below the recommended characteristic value ranges [52]. However, this could result from the fact that a high level of difficulty was measured for this item (item difficulty = 3.418); therefore, it was included in the calculations for further analyses.

Upon completion of each task, students were requested to provide subjective ratings regarding cognitive load and affective variables such as topic interest (because the other affective variables from the original study were not used for the present study, they are not presented in detail here). Cognitive load measures were collected by asking students about perceived task difficulty [53,54] (seven-point rating scale from "very easy" to "very difficult") and invested mental effort [39] (seven-point rating scale from "very low" to "very high"). The two-dimensional Rasch model for the rating scale assessing cognitive load (Dimension 1: perceived task difficulty; Dimension 2: invested mental effort) demonstrated satisfactory fit (item reliability = 0.93, $0.72 \leq \text{MNSQ} \leq 1.35$). Topic interest (3 items; item example: "I found the task on smoking interesting") [55–57] was measured using a five-point rating scale from "completely agree" to "completely disagree". The Rasch model for the rating scale assessing topic interest also demonstrated acceptable fit (item reliability = 0.97, $0.60 \leq \text{MNSQ} \leq 1.91$). Based on these analyses, the data set contained task-related values for objective task difficulty, perceived task difficulty, invested mental effort, and topic interest.

2.2. Data for the Present Study

For the present study, we used the Rasch parameters from the original study at the task level. Hence, our data set consisted of 32 tasks with Rasch parameters of objective task difficulty, perceived task difficulty, invested mental effort, and topic interest for each task.

Objective task difficulty was aggregated from students' right or wrong responses to the items forming one task of the chemistry test. The mean value of the Rasch parameters of the four associated items was calculated for each task. Tasks that were answered correctly by very few students were considered difficult tasks and associated with positive Rasch values for objective task difficulty.

Perceived task difficulty was the Rasch parameter of students' ratings regarding the perceived difficulty of the specific task. Tasks with positive Rasch values for perceived task difficulty were tasks that were perceived to be difficult (ratings of 6 or 7), whereas tasks with negative Rasch values for perceived task difficulty were perceived to be easy (ratings of 1 or 2).

Invested mental effort was the Rasch parameter of students' ratings regarding the mental effort invested in the specific task. Tasks with positive Rasch values for invested mental effort were tasks in which students invested considerable effort (ratings of 6 or 7), whereas

tasks with negative Rasch values for invested mental effort were tasks in which students invested little mental effort (ratings of 1 or 2).

Topic interest was the Rasch parameter of students' ratings regarding topic interest for the specific task. Again, positive Rasch values (ratings of 4 or 5) signaled high topic interest in the task, and negative values reflected low topic interest in the task (ratings of 1 or 2).

All ratings were collected after the students had completed the tasks. Accordingly, it is not possible to say whether interest corresponded to spontaneous interest (situational interest) that arose in the test situation or whether the students already had a stable, long-term interest (individual interest) in the topics, which had developed through repeated engagement. Both alone—and their combination—are conceivable. In both cases, topic interest can be expected to be positively related to student outcomes [35].

Importantly, the Rasch model has the property of specific objectivity [58]. That is, if the model holds for a set of data, the task or item parameters are generally independent of the specific sample of participants, and the person parameters (in which we were not interested) are generally independent of the specific sample of tasks or items. In other words, the Rasch parameters of our 32 chemistry tasks represented genuine properties of the tasks: the objective difficulty of a given task (based on students' task performance), the level of difficulty that students generally perceived for said task, the amount of effort that students generally reported having invested in the task, and the level of topic interest that students generally reported for the task. Thus, the calculation of the Rasch parameters for cognitive load ratings was task-specific and did not exhibit the problem of aggregated cognitive load ratings across multiple tasks described by Leppink and van Merriënboer [59].

2.3. Data Analysis

The Rasch parameters of the 32 tasks of the chemistry test were imported to SPSS (version 28). The tasks were the units of observation, and the Rasch parameters were the variables. To investigate the relationships between the variables, we computed linear correlations (Pearson's r). The SPSS PROCESS tool (version 4.2; [60]) was used to analyze the influence of interest on the relationship between invested mental effort and perceived task difficulty (PROCESS Model #1). In the results section, we report F -ratios, R^2 , and changes in R^2 for the model. Additionally, we report t -ratios, regression coefficients (B), and standard errors (SE) for the model parameters as well as the results based on Johnson–Neyman output. For all analyses, 95% bias-corrected and accelerated confidence intervals, based on 10,000 bootstrap samples for correlations and 50,000 bootstrap samples for the PROCESS parameters (95% CI), are reported in parentheses [61].

3. Results

3.1. Correlations

In the first step, linear correlations were computed to examine the relationships between the variables of interest (objective task difficulty, perceived task difficulty, invested mental effort, and topic interest; Table 1). As expected, invested mental effort and perceived task difficulty were highly correlated ($r = 0.78$, $p < 0.001$, 95% CI [0.56, 0.91]). Both invested mental effort and perceived task difficulty was correlated with objective task difficulty. Perceived task difficulty and objective task difficulty were almost perfectly related ($r = 0.91$, $p < 0.001$, 95% CI [0.85, 0.96]). In contrast, the correlation between invested mental effort and objective task difficulty was remarkably lower ($r = 0.69$, $p < 0.001$, 95% CI [0.45, 0.86]). The mental effort that students invested in solving a task therefore did not depend exclusively on the objective difficulty of the task.

Further correlation analyses indicated that only perceived task difficulty was positively correlated with topic interest, while the correlation between topic interest and invested mental effort was not significant. Thus, higher topic interest did not entail higher invested mental effort.

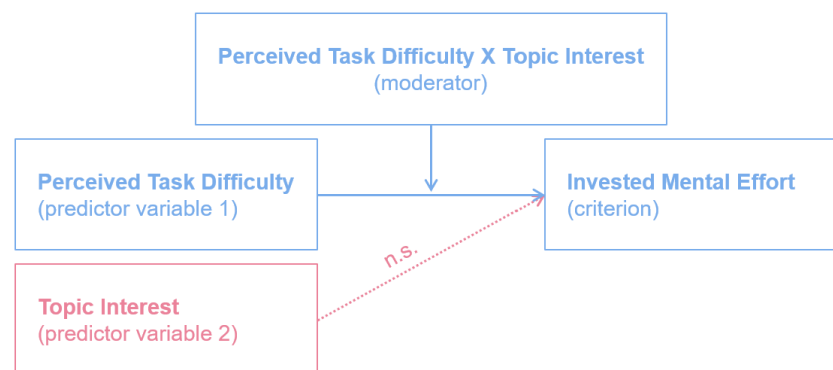
Table 1. Pearson correlations.

	Objective Task Difficulty	Perceived Task Difficulty	Invested Mental Effort	Topic Interest
Objective Task Difficulty				
Perceived Task Difficulty	0.91 (<0.001) [0.85, 0.96]			
Invested Mental Effort	0.69 (<0.001) [0.45, 0.87]	0.78 (<0.001) [0.56, 0.91]		
Topic Interest	0.41 (0.019) [0.10, 0.65]	0.54 (0.002) [0.29, 0.73]	0.24 (0.182) [−0.16, 0.58]	

Note. Pearson's r , (p), [95% BCa CI, based on 10,000 bootstrap samples], $N_{\text{tasks}} = 32$.

3.2. Analyzing the Impact of Interest

To investigate whether topic interest moderated the relationship between perceived task difficulty and invested mental effort, we used the SPSS Tool PROCESS to specify a linear regression model with perceived task difficulty and interest and their interaction as predictors and invested mental effort as the criterion (Figure 1).

**Figure 1.** Statistical model.

The results showed that perceived task difficulty was a significant predictor ($B = 0.43$, 95% CI [0.26, 0.57], $SE = 0.07$, $t(26) = 5.84$, $p < 0.001$), whereas topic interest was not ($B = -0.06$, 95% CI [−0.14, 0.04], $SE = 0.05$, $t(26) = -1.14$, $p = 0.263$). However, the interaction term (perceived task difficulty X interest) turned out to be a significant predictor ($B = 0.26$, 95% CI [0.03, 0.62], $SE = 0.12$, $t(26) = 2.23$, $p = 0.034$). Overall, the model accounted for 71% of the variance of invested mental effort, with the interaction term explaining 5% of the variance beyond the main-effects terms. The interaction indicated that when topic interest was low (1 SD beneath the mean), there was a significant but low positive relationship between perceived task difficulty and invested mental effort ($B = 0.38$, 95% CI [0.20, 0.56], $SE = 0.09$, $t(26) = 4.39$, $p < 0.001$). When topic interest was at the mean, the relationship between perceived task difficulty and invested mental effort was stronger ($B = 0.51$, 95% CI [0.37, 0.65], $SE = 0.07$, $t(26) = 7.53$, $p < 0.001$). When topic interest was high (1 SD above the mean), the relationship between perceived task difficulty and invested mental effort was strongest ($B = 0.63$, 95% CI [0.45, 0.82], $SE = 0.09$, $t(26) = 7.13$, $p < 0.001$). As expected, these results suggest that topic interest moderated the regression of invested mental effort on perceived task difficulty.

To visualize the moderated regression, we split the 32 tasks at the median of topic interest into interesting and uninteresting tasks. In a scatter plot (Figure 2a), the regression of

invested mental effort on perceived task difficulty is displayed separately for uninteresting (blue) and interesting (red) tasks. For tasks that were perceived to be rather easy, there is a gap between interesting and uninteresting tasks, with the latter requiring considerably more mental effort than interesting tasks. However, this difference in invested mental effort between interesting and uninteresting tasks vanishes as perceived task difficulty increases. Given that uninteresting tasks beyond the point of intersection of the two regression lines are missing, it remains unclear whether the effect is inverse for tasks perceived as more difficult.

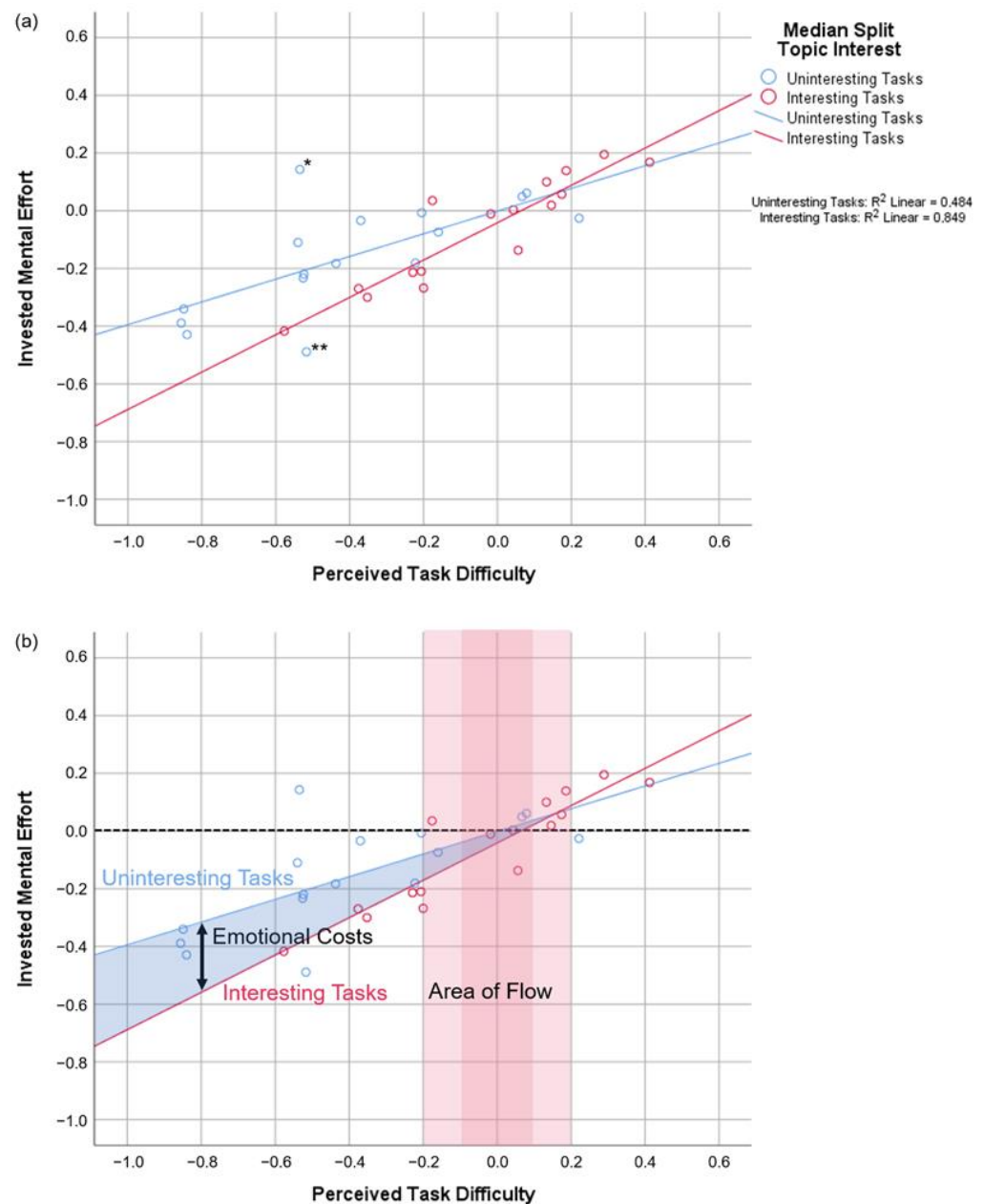


Figure 2. Interaction effect. (a) Scatterplot visualizing the moderated regression of invested mental effort on perceived task difficulty for interesting and uninteresting tasks; (b) Emotional costs.

Note. Invested mental effort and perceived task difficulty for the 32 tasks are represented as Rasch parameters on Logit scales.

Note. The blue area describes emotional costs (investment of additional mental effort). The red area describes a state of flow where perceived task difficulty roughly corresponds to students' performance capability.

There seem to be two outliers in Figure 2a regarding the regression line for the uninteresting tasks (blue line, the point marked (**) at $-0.54, 0.14$ and the point marked (*) at $-0.52, -0.49$). However, when these two points are excluded from the analysis, the interaction of perceived difficulty and interest remains significant ($B = 0.16$, 95% CI $[0.01, 0.57]$).

4. Discussion

The present study used an existing data set evaluating a chemistry test at the end of German lower secondary school. The test consisted of 32 tasks. In our analyses, the tasks were used as the units of observation, and the Rasch parameters of the tasks were the variables. The Rasch parameters relied on data from 1543 students who solved the chemistry test based on a multi-matrix design (objective task difficulty). Furthermore, students rated each task for perceived difficulty, invested mental effort, and topic interest. The Rasch parameters of our 32 chemistry tasks represented genuine properties of the tasks. All Rasch parameters were task-specific and, therefore, did not exhibit the problem of aggregated cognitive load ratings across multiple tasks [59].

The correlation analyses supported the expectation of van Gog and Paas [27] that invested mental effort and perceived task difficulty measure a similar construct (cognitive load) but not the same thing ($r = 0.78$). Perceived task difficulty correlated nearly perfectly with objective task difficulty ($r = 0.91$), in line with the findings of Ayres [53]. The smaller correlation ($r = 0.69$) between invested mental effort and objective task difficulty indicated that invested mental effort did not exclusively depend on cognitive aspects but also on motivational ones [40]. Unexpectedly, invested mental effort was not directly related to topic interest. Accordingly, there was no direct link between interest in a task and the effort invested in that task.

Nonetheless, a moderated regression analysis indicated that topic interest moderated the linear regression of invested mental effort on perceived task difficulty. Figure 2b helps to conceptualize this moderated regression: the tasks in the red area of the figure were perceived by participants with average ability as not too easy and not too difficult, as indicated by a Rasch parameter of 0.0 ± 0.2 for perceived task difficulty. According to flow theory [8,30], these tasks represent an equilibrium of perceived skills and challenges, resulting in an experience of flow. The lower left part of the figure shows tasks that were perceived to be easy, as indicated by negative Rasch parameters < -0.2 . These tasks lie outside the equilibrium of skills and challenges, probably resulting in an experience of boredom.

However, the decrease in invested mental effort as perceived task difficulty decreases differs for interesting and uninteresting tasks: interesting easy tasks are associated with a lower investment of mental effort than uninteresting easy tasks. Concerning interesting tasks, we assume that the students would like to interact with these tasks because they were interested in the tasks' topic and did not mind that these tasks were not perceived to be challenging. Consequently, they would not feel bored, and there would be no (or very low) emotional costs with which to cope. This result is in line with findings from interest research, which support the hypothesis that interest leads to spontaneous attention to interesting aspects, which are thus processed more quickly [35].

Conversely, uninteresting easy tasks were linked to a higher investment of mental effort. Students were not interested in interacting with these tasks (although they were required to do so), and they did not perceive them as a challenge. Therefore, we assume that they would feel bored and the emotional costs to cope with would be high. This is in line with the findings of Flake et al. [31], who found that a lack of challenge (which caused boredom, i.e., emotional costs) was a reason why students perceived a course as not motivating. We hypothesize that the additional amount of mental effort is caused by the need for motivational self-regulation, such as convincing and continuously reminding oneself not to become distracted and to continue working even though the task is boring. Accordingly, the additional effort invested in uninteresting easy tasks can be considered an emotional cost given that students have to self-regulate their emotions.

As perceived task difficulty increases, the difference in mental effort investment between interesting and uninteresting tasks decreases (center of Figure 2b). The reduction in additional mental effort (i.e., emotional costs) with increasing perceived task difficulty supports the idea that flow situations, where skills and challenges are balanced, have no emotional costs and, therefore, do not require motivational self-regulation. This idea is supported by the fact that the point of intersection of the two regression lines is close to the middle of the two scales. At this point, invested mental effort is not influenced by topic interest but exclusively relies on perceived task difficulty.

Overall, the findings are in line with the hypothesis that easy tasks that are interesting are associated with lower cognitive load (i.e., invested mental effort) than easy tasks that are not interesting. Still, because there are no uninteresting tasks in Figure 2b beyond the point of intersection, the given sample of tasks does not allow us to discuss situations where students are over-challenged by very difficult tasks, resulting in negative emotions such as anxiety (as predicted by flow theory).

Furthermore, the results generally substantiate Grund et al.'s [62] suggestion to differentiate the allocation of effort by cause: effort by complexity, effort by need frustration, and effort by allocation. In this sense, the mental effort invested in easy tasks is initially determined by effort by complexity. For uninteresting tasks, additional effort (effort by need frustration) must be integrated, resulting in a higher total effort for these tasks.

Limitations and Further Research

Analyzing existing data comes with some limitations. First, all analyses were performed with tasks as the unit of observation (task level), not students (student level). This allowed us to make general statements about tasks, including that easy, uninteresting tasks are associated with a higher level of invested mental effort than easy, interesting tasks. However, our analyses at the task level did not allow us to draw conclusions about individual psychological processes. Future studies analyzing data at the student level are thus needed. Second, our results are limited to tasks of a chemistry test at the end of German lower secondary school. They should therefore be replicated by intervention studies in other domains (following, for example, a 2×2 design, varying topic interest and task difficulty as factors, and measuring invested mental effort as the dependent variable and perceived task difficulty as a mediator). Third, measuring emotional costs and motivational self-regulation would be necessary given that we interpreted our results with respect to these two variables without corresponding measures in our data set. Fourth, in our data, topic interest was measured after students had processed the tasks, which limited our findings. Future studies should measure topic interest before the students process the tasks and differentiate between stable individual interest and spontaneously aroused situational interest. Fifth, the data set lacks tasks that are uninteresting and difficult. Both interesting and uninteresting difficult tasks would be necessary to test whether interest is a relevant factor for the investment of mental effort even in situations in which the demands of the tasks are higher than students' abilities (which would probably lead to emotional costs such as anxiety). To understand individual differences in dealing with tasks for which challenges and skills are not balanced, analyses at the student level are necessary.

Our analyses used perceived task difficulty as an indication of the balance between ability and challenge. Even though perceived task difficulty correlated almost perfectly with objective task difficulty ($r = 0.91$), the two values are not identical. One possible reason could be that although the order of the Rasch parameters of task difficulty on the logit scale can be assumed to be valid for all individuals, the distance between an individual's ability parameter and the mean of the item difficulty parameters on the logit scale varies. In our data set, perceived task difficulty was always lower than objective task difficulty: on the regression line for objective task difficulty and perceived task difficulty, the value for perceived task difficulty was -0.18 when the objective task difficulty was 0.00 on the logit scale. Further research is thus needed to determine whether perceived task difficulty

is more important to a flow experience than objective task difficulty and what role the difference between the two plays.

Our assumption that the higher invested mental effort for easy, uninteresting tasks is caused by motivational self-regulation is an interpretation of our results based on theoretical considerations and previous research. However, this interpretation could not be tested empirically in our study because the necessary data were not available in the data set. Consequently, further research should consider how other constructs (prior knowledge, objective task difficulty, self-efficacy) influence the individual investment of mental effort. In interpreting the data, we assumed that there was a balance between ability and challenge for certain tasks. At the task level, the data set makes statements about the balance between average perceived task difficulty and average invested mental effort. At the individual level, this does not necessarily always equate to a student's ability and challenge.

Finally, our results raise the question of whether the additional cognitive load caused by the extra mental effort needed when working on uninteresting easy tasks should be considered extraneous cognitive load. This view is supported by the fact that this additional load appears unnecessary, unlike for interesting easy tasks. Alternatively, the additional load may be necessary to show persistence and solve these uninteresting easy tasks successfully. Following this line of reasoning, the additional load would be better described as a germane cognitive load or as a pertinent factor for activating relevant working memory resources [22]. Corresponding measuring instruments (i.e., multidimensional cognitive load measures and working memory resources) would be required to settle this question.

5. Conclusions

The present study contributes to research seeking to integrate motivation into cognitive load theory. Specifically, we investigated whether there is an interaction between perceived task difficulty and interest that affects invested mental effort. To do so, we used an existing data set from a chemistry test. All analyses were performed at the task level using task-specific Rasch parameters, which were calculated from students' responses. The results showed that the mental effort invested in uninteresting easy tasks was greater than for interesting easy tasks. At first glance, our finding that interest can compensate for investing additional mental effort to cope with emotional costs when processing unchallenging test items or tasks may be counter-intuitive to cognitive load researchers. This is because other researchers attempted to increase invested mental effort by arousing students' interest. However, the two ideas need not contradict each other. Our analysis only describes the area in which skills are greater than challenges (left side of Figure 2b); for the area where challenges are greater than skills (right side of Figure 2b), the idea that interest could increase students' willingness to invest mental effort would be consistent with our findings.

Overall, our results shed light on the role of motivational variables in the context of cognitive load theory. Our results have both theoretical and practical implications. On the theoretical side, our results indicate that the relationship between motivation and cognitive load is not simply uni-directional but bidirectional. Whereas Feldon and colleagues [1,36,38] showed that motivation can be an outcome of cognitive load, our results show that motivation (in this case: low motivation) can be a source of cognitive load (in this case: a source of additional cognitive load when processing easy tasks). This theoretical insight opens up new research questions, such as questions concerning possible chained effects of low motivation as a predictor of higher cognitive load (in terms of increased mental effort), which in turn might have differential effects on students' motivation. On the practical side, our results indicate that teachers would be well advised to focus on raising students' interest, especially when students are expected to process tasks or problems that are very easy for them. In addition to the need to replicate the results and test our interpretation that higher mental effort is caused by motivational self-regulation, a central question is whether additional mental effort represents an emotional cost and how this is related to extraneous and germane cognitive load and the experience of flow.

Author Contributions: Conceptualization, K.S. and D.L.; methodology, K.S. and D.L.; formal analysis, K.S. and D.L.; investigation, V.F.; resources, M.W. and V.F.; data curation, V.F.; writing—original draft preparation, K.S. and D.L.; writing—review and editing, V.F. and M.W.; visualization, K.S.; supervision, M.W.; funding acquisition, M.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by German Research Foundation (DFG), grant number IMBliCK (MA1792/6-1, Jürgen Mayer; WA2829/5-1, Maik Walpuski; SU187/12-1 Elke Sumfleth).

Institutional Review Board Statement: Ethical review and approval were waived for this study for the following reasons: approval by the Institutional Review Board is not required at German schools; the design of the original study complied with the recommendations of the German Research Foundation for Good Scientific Practice.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon request.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study, the collection, analyses, or interpretation of the data, the writing of the manuscript, or the decision to publish the results.

References

1. Feldon, D.F.; Callan, G.; Juth, S.; Jeong, S. Cognitive load as motivational cost. *Educ. Psychol. Rev.* **2019**, *31*, 319–337. [CrossRef]
2. Leutner, D. Motivation and emotion as mediators in multimedia learning. *Learn. Instr.* **2014**, *29*, 174–175. [CrossRef]
3. Plass, J.L.; Kalyuga, S. Four ways of considering emotion in cognitive load theory. *Educ. Psychol. Rev.* **2019**, *31*, 339–359. [CrossRef]
4. Moreno, R. Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Comput. Assist. Learn.* **2006**, *22*, 149–158. [CrossRef]
5. Fischer, V. *Der Einfluss von Interesse und Motivation auf die Messung von Fach- und Bewertungskompetenz im Fach Chemie*; Logos Verlag Berlin GmbH: Berlin, Germany, 2019; Volume 277.
6. Sweller, J.; van Merriënboer, J.J.G.; Paas, F.G.W.C. Cognitive architecture and instructional design. *Educ. Psychol. Rev.* **1998**, *10*, 251–296. [CrossRef]
7. Sweller, J.; van Merriënboer, J.J.G.; Paas, F. Cognitive architecture and instructional design: 20 years later. *Educ. Psychol. Rev.* **2019**, *31*, 261–292. [CrossRef]
8. Shernoff, D.J.; Csikszentmihalyi, M.; Shneider, B.; Shernoff, E.S. Student engagement in high school classrooms from the perspective of flow theory. *Sch. Psychol. Q.* **2003**, *18*, 158–176. [CrossRef]
9. Barron, K.E.; Hulleman, C.S. Expectancy-value-cost model of motivation. In *International Encyclopedia of the Social & Behavioral Sciences*, 2nd ed.; Wright, J.D., Ed.; Elsevier: Amsterdam, The Netherlands, 2015; pp. 503–509, ISBN 9780080970875.
10. Paas, F.; Sweller, J. Implications of cognitive load theory for multimedia learning. In *The Cambridge Handbook of Multimedia Learning*, 2nd ed.; Mayer, R.E., Ed.; Cambridge University Press: Cambridge, UK, 2014; pp. 27–42, ISBN 9781139547369.
11. Sweller, J. Cognitive load during problem solving: Effects on learning. *Cogn. Sci.* **1988**, *12*, 257–285. [CrossRef]
12. Sweller, J. Cognitive Load Theory: What we learn and how we learn. In *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy*, 1st ed.; Spector, J.M., Lockee, B.B., Childress, M.D., Eds.; Springer: Cham, Switzerland, 2023; pp. 137–152, ISBN 978-3-319-17460-0.
13. Sweller, J. Cognitive load theory and individual differences. *Learn. Individ. Differ.* **2024**, *110*, 102423. [CrossRef]
14. Krieglstein, F.; Beege, M.; Rey, G.D.; Ginns, P.; Krell, M.; Schneider, S. A systematic meta-analysis of the reliability and validity of subjective cognitive load questionnaires in experimental multimedia learning research. *Educ. Psychol. Rev.* **2022**, *34*, 2485–2541. [CrossRef]
15. Krieglstein, F.; Beege, M.; Rey, G.D.; Sanchez-Stockhammer, C.; Schneider, S. Development and validation of a theory-based questionnaire to measure different types of cognitive load. *Educ. Psychol. Rev.* **2023**, *35*, 9. [CrossRef]
16. Paas, F.; Tuovinen, J.E.; Tabbers, H.; van Gerven, P.W.M. Cognitive load measurement as a means to advance cognitive load theory. *Educ. Psychol.* **2003**, *38*, 63–71. [CrossRef]
17. Paas, F.; Renkl, A.; Sweller, J. Cognitive load theory and instructional design: Recent developments. *Educ. Psychol.* **2003**, *38*, 1–4. [CrossRef]
18. Sweller, J. Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educ. Psychol. Rev.* **2010**, *22*, 123–138. [CrossRef]
19. Paas, F.; Sweller, J. Implications of cognitive load theory for multimedia learning. In *The Cambridge Handbook of Multimedia Learning*; Mayer, R.E., Fiorella, L., Eds.; Cambridge University Press: Cambridge, UK, 2021; pp. 73–81, ISBN 9781108894333.
20. Leppink, J.; Paas, F.; van der Vleuten, C.P.M.; van Gog, T.; van Merriënboer, J.J.G. Development of an instrument for measuring different types of cognitive load. *Behav. Res. Methods* **2013**, *45*, 1058–1072. [CrossRef]

21. Klepsch, M.; Schmitz, F.; Seufert, T. Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Front. Psychol.* **2017**, *8*, 1997. [CrossRef] [PubMed]
22. Greenberg, K.; Zheng, R. Revisiting the debate on germane cognitive load versus germane resources. *J. Cogn. Psychol.* **2023**, *35*, 295–314. [CrossRef]
23. Moreno, R.; Park, B. Cognitive load theory: Historical development and relation to other theories. In *Cognitive Load Theory*; Plass, J.L., Moreno, R., Brünken, R., Eds.; Cambridge University Press: Cambridge, UK, 2012; pp. 9–28, ISBN 9780521677585.
24. Brünken, R.; Seufert, T.; Paas, F. Measuring cognitive load. In *Cognitive Load Theory*; Plass, J.L., Moreno, R., Brünken, R., Eds.; Cambridge University Press: Cambridge, UK, 2012; pp. 181–202, ISBN 9780521677585.
25. Sweller, J. Measuring cognitive load. *Perspect. Med. Educ.* **2018**, *7*, 1–2. [CrossRef]
26. Sweller, J.; Ayres, P.; Kalyuga, S. Measuring cognitive load. In *Cognitive Load Theory: Explorations in the Learning Sciences, Instructional Systems, and Performance Technologies*; Sweller, J., Ayres, P., Kalyuga, S., Eds.; Springer: New York, NY, USA, 2011; pp. 71–85, ISBN 978-1-4419-8125-7.
27. van Gog, T.; Paas, F. Instructional efficiency: Revisiting the original construct in educational research. *Educ. Psychol.* **2008**, *43*, 16–26. [CrossRef]
28. Belmont, J.M.; Mitchell, D.W. The general strategies hypothesis as applied to cognitive theory in mental retardation. *Intelligence* **1987**, *11*, 91–105. [CrossRef]
29. Yerkes, R.M.; Dodson, J.D. The relation of strength of stimulus to rapidity of habit-formation. *J. Comp. Neurol. Psychol.* **1908**, *18*, 459–482. [CrossRef]
30. Csikszentmihalyi, M. The flow experience and its significance for human psychology. In *Optimal Experience: Psychological Studies of Flow in Consciousness*, 1st ed.; Csikszentmihalyi, M., Csikszentmihalyi, I.S., Eds.; Cambridge University Press: Cambridge, UK, 1988; pp. 15–35.
31. Flake, J.K.; Barron, K.E.; Hulleman, C.; McCoach, B.D.; Welsh, M.E. Measuring cost: The forgotten component of expectancy-value theory. *Contemp. Educ. Psychol.* **2015**, *41*, 232–244. [CrossRef]
32. Jiang, Y.; Rosenzweig, E.Q.; Gaspard, H. An expectancy-value-cost approach in predicting adolescent students' academic motivation and achievement. *Contemp. Educ. Psychol.* **2018**, *54*, 139–152. [CrossRef]
33. Muenks, K.; Miller, J.E.; Schuetze, B.A.; Whittaker, T.A. Is cost separate from or part of subjective task value? An empirical examination of expectancy-value versus expectancy-value-cost perspectives. *Contemp. Educ. Psychol.* **2023**, *72*, 102149. [CrossRef]
34. Hidi, S.; Renninger, K.A. The four-phase model of interest development. *Educ. Psychol.* **2006**, *41*, 111–127. [CrossRef]
35. Hidi, S. Interest, reading, and learning: Theoretical and practical considerations. *Educ. Psychol. Rev.* **2001**, *13*, 191–209. [CrossRef]
36. Feldon, D.F.; Franco, J.; Chao, J.; Peugh, J.; Maahs-Fladung, C. Self-efficacy change associated with a cognitive load-based intervention in an undergraduate biology course. *Learn. Instr.* **2018**, *56*, 64–72. [CrossRef]
37. Bandura, A. *Self-Efficacy: The Exercise of Control*; W. H. Freeman: New York, NY, USA, 1997.
38. Feldon, D.F.; Brockbank, R.; Litson, K. Direct effects of cognitive load on self-efficacy during instruction. *J. Educ. Psychol.* **2023**. [CrossRef]
39. Paas, F. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *J. Educ. Psychol.* **1992**, *84*, 429–434. [CrossRef]
40. Paas, F.; Tuovinen, J.E.; van Merriënboer, J.J.G.; Darabi, A. A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *ETR D* **2005**, *53*, 25–34. [CrossRef]
41. Moreno, R. Instructional technology: Promise and pitfalls. In *Technologie-Based Education: Bringing Researchers and Practitioners Together*; Pyllykzillig, L.M., Bodvarsson, M., Bruning, R., Eds.; Information Age: Greenwich, CT, USA, 2005; pp. 1–20.
42. Moreno, R.; Mayer, R. Interactive multimodal learning environments. *Educ. Psychol. Rev.* **2007**, *19*, 309–326. [CrossRef]
43. Astleitner, H.; Wiesner, C. An integrated model of multimedia learning and motivation. *J. Educ. Multimed. Hypermedia* **2004**, *13*, 3–21.
44. Mayer, R.E. Cognitive theory of multimedia learning. In *The Cambridge Handbook of Multimedia Learning*, 2nd ed.; Mayer, R.E., Ed.; Cambridge University Press: Cambridge, UK, 2014; pp. 43–71, ISBN 9781139547369.
45. Mayer, R.E. Cognitive theory of multimedia learning. In *The Cambridge Handbook of Multimedia Learning*; Mayer, R.E., Fiorella, L., Eds.; Cambridge University Press: Cambridge, UK, 2021; pp. 57–72, ISBN 9781108894333.
46. Heidig, S.; Müller, J.; Reichelt, M. Emotional design in multimedia learning: Differentiation on relevant design features and their effects on emotions and learning. *Comput. Hum. Behav.* **2015**, *44*, 81–95. [CrossRef]
47. Mayer, R.E.; Estrella, G. Benefits of emotional design in multimedia instruction. *Learn. Instr.* **2014**, *33*, 12–18. [CrossRef]
48. Plass, J.L.; Heidig, S.; Hayward, E.O.; Homer, B.D.; Um, E. Emotional design in multimedia learning: Effects of shape and color on affect and learning. *Learn. Instr.* **2014**, *29*, 128–140. [CrossRef]
49. Magner, U.I.; Schwonke, R.; Aleven, V.; Popescu, O.; Renkl, A. Triggering situational interest by decorative illustrations both fosters and hinders learning in computer-based learning environments. *Learn. Instr.* **2014**, *29*, 141–152. [CrossRef]
50. Park, B.; Moreno, R.; Seufert, T.; Brünken, R. Does cognitive load moderate the seductive details effect? A multimedia study. *Comput. Hum. Behav.* **2011**, *27*, 5–10. [CrossRef]

51. KMK. Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss (Jahrgangsstufe 10): [Educational Standards in Chemistry for the Intermediate School-Leaving Certificate (Year 10)]. Beschluss vom 16.12.2004. Available online: https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Chemie.pdf (accessed on 9 September 2024).
52. Adams, R. Scaling PISA cognitive data. In *Programme for International Student Assessment (PISA)*; Adams, R., Wu, M., Eds.; OECD: Paris, France, 2003; pp. 99–108, ISBN 9789264199514.
53. Ayres, P. Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learn. Instr.* **2006**, *16*, 389–400. [CrossRef]
54. Kalyuga, S.; Chandler, P.; Tuovinen, J.; Sweller, J. When problem solving is superior to studying worked examples. *J. Educ. Psychol.* **2001**, *93*, 579–588. [CrossRef]
55. Fechner, S. *Effects of Context-Oriented Learning on Student Interest and Achievement in Chemistry Education*; Logos Verlag: Berlin, Germany, 2009; ISBN 9783832598679.
56. van Vorst, H. *Kontextmerkmale und ihr Einfluss auf das Schülerinteresse im Fach Chemie: [Contextual Features and Their Influence on Student Interest in Chemistry]*; Logos Verlag: Berlin, Germany, 2013; ISBN 3832533214.
57. Haugwitz, M. Kontextorientiertes Lernen und Concept Mapping im Fach Biologie. Eine Experimentelle Untersuchung zum Einfluss auf Interesse und Leistung unter Berücksichtigung von Moderationseffekten Individueller Voraussetzungen beim Kooperativen Lernen: [Context-Oriented Learning and Concept Mapping in Biology. An Experimental Study on the Influence on Interest and Performance Considering Moderation Effects of Individual Preconditions in Cooperative Learning]. 2009. Available online: http://duepublico.uni-duisburg-essen.de/servlets/DerivateServlet/Derivate-23401/Dissertation_Haugwitz.pdf (accessed on 19 September 2024).
58. Rasch, G. On Specific Objectivity. An Attempt at Formalizing the Request for Generality and Validity of Scientific Statements. *Dan. Yearb. Philos.* **1977**, *14*, 58–93. [CrossRef]
59. Leppink, J.; van Merriënboer, J.J.G. The beast of aggregating cognitive load measures in technology-based learning. *Int. Forum Educ. Technol. Soc.* **2015**, *18*, 230–245.
60. Hayes, A.F. PROCESS Procedure for SPSS Version 4.2. 2022. Available online: <https://haskayne.ucalgary.ca/CCRAM/resource-hub> (accessed on 19 September 2024).
61. Field, A. *Discovering Statistics Using IBM SPSS Statistics*, 5th ed.; SAGE: Los Angeles, CA, USA; London, UK; New Delhi, India; Singapore; Washington, DC, USA; Melbourne, Australia, 2018; ISBN 9781526419521.
62. Grund, A.; Fries, S.; Nückles, M.; Renkl, A.; Roelle, J. When is learning “effortful”? Scrutinizing the concept of mental effort in cognitively oriented research from a motivational perspective. *Educ. Psychol. Rev.* **2024**, *36*, 11. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

The Worked-Example Effect and a Mastery Approach Goal Orientation

Hee Min Lee * and Paul Ayres *

School of Education, University of New South Wales, Sydney, NSW 2052, Australia

* Correspondence: heemin.lee@unsw.edu.au (H.M.L.); p.ayres@unsw.edu.au (P.A.)

Abstract: This study aimed to explore the impact of a mastery approach goal orientation (MAGO) on learning from worked examples. In this experiment ($N = 98$, mean age = 13.9 years), learners had their MAGO measured, and received instruction in mathematics, either through a worked-example or a problem-solving strategy. The study demonstrated that the worked-examples approach resulted in enhanced retention (but not transfer) and decreased cognitive load when compared to the problem-solving approach. However, there was a significant interaction between instructional strategy and the MAGO, indicating that only learners with a high MAGO benefited from worked examples. Learners with higher MAGO levels also experienced less cognitive load than learners with a lower MAGO. These results indicate a moderating role of MGO in enhancing the effectiveness of worked examples. This study also found that prior knowledge was the only factor influencing transfer performance, highlighting the importance of studying its impact.

Keywords: worked examples; cognitive load theory; mastery approach goal orientation

1. Introduction

The *worked example effect* occurs when greater learning is achieved through studying worked-out solutions to problems rather than through problem-solving methods. The effect has been extensively researched within the field of cognitive load theory (CLT; see [1,2]), finding substantial evidence that for novice learners in particular, explicit instruction based on studying expert solutions to problems is more effective for acquiring new knowledge than unguided problem-solving methods (see [3,4]). Evidence supporting the worked example effect has been found in a wide spectrum of learning domains and under various conditions (for summaries, see [4,5]).

From the perspective of factors that moderate the worked-example effect, research has mainly focused on the impact of prior knowledge. When learners gain expertise in a domain, worked examples lose their effectiveness and can become equal or even inferior to problem solving. This phenomenon is known as the expertise reversal effect [6,7]). However, other individual differences that may moderate the effect have received less investigation. The main aim of the present research was to fill this gap and therefore broaden the worked-examples research by investigating the moderating impact of having a mastery goal orientation [8]. With few exceptions, notably studies by Belenky and Nokes-Malach [9,10], little research has been conducted that investigates worked examples combined with a mastery goal orientation.

2. Cognitive Load Theory and Worked Examples

According to cognitive load theory (CLT; see [5]), cognitive load is considered to be the working memory load expended during learning, and three categories have been identified ([1]). If the materials to be learnt involve many interacting elements, they are considered to be high in element interactivity ([11]) and generate *intrinsic cognitive load*. If the instructional materials are designed poorly, they can create *extraneous cognitive load*.

Citation: Lee, H.M.; Ayres, P. The Worked-Example Effect and a Mastery Approach Goal Orientation. *Educ. Sci.* **2024**, *14*, 597.

<https://doi.org/10.3390/educsci14060597>

Academic Editor: Daniel Muijs

Received: 21 January 2024

Revised: 31 March 2024

Accepted: 24 May 2024

Published: 1 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

by requiring learners to process information unrelated to learning. A third cognitive load (germane) has also been identified that corresponds to the mental resources invested directly in schema acquisition ([1]). In more recent studies, researchers have argued that germane load is not an independent load as such (see [12]). For example, Paas and van Merriënboer [13] define germane processing as the resources devoted to intrinsic cognitive load, and thus are dependent upon this load. Accordingly, total cognitive load is calculated by adding together intrinsic and extraneous loads only, rather than the three loads as previously defined (see [13,14]). Even though the concept of germane load has evolved over the years, CLT researchers have successfully designed instructional materials to accommodate the different types of cognitive load, especially in lowering extraneous cognitive load, in order to maximize learning ([1]).

Worked examples provide step-by-step solutions to a problem, and, according to CLT, allows the learner to devote more cognitive resources to learning about the problem domain than unguided problem solving [5]. In the case of learning through problem solving, learners may eventually solve the problem but not engage in schema acquisition due to a heavy cognitive load generated by inefficient problem-solving strategies [15]. Without domain-specific prior knowledge, a problem solver's limited working memory capacity becomes stretched to deal with both solving the problem and learning from it [2,4]. In contrast, worked examples provide the learner with a detailed expert problem-solving model [3,16], which reduces the negative effects of extraneous cognitive load while allowing for a more directed focus on learning and the subsequent generation of germane cognitive load.

Original CLT studies into the worked example effect were in the domain of mathematics [17,18]. But since this time, the worked-example effect has been found in a much broader range of learning domains, including non-scientific areas [4]. These diverse studies have consistently shown that worked examples leads to greater learning with lowered cognitive load. A number of additional learning strategies have also been successfully incorporated into the experimental designs, such as variability [19], collaboration [20,21], fading strategies [22], process explanations [23], isolating elements [24], individual solution steps [25], and self-explanations [26]. These studies have shown that other learning strategies can be successfully incorporated into worked-examples to enhance their influence further. Research has also identified a number of factors that moderate the effectiveness of worked examples.

Moderating Factors of the Worked Example Effect

As described in the previous section, the worked example effect has been identified in many learning domains; however, its effectiveness is highly dependent upon the learner's relevant prior knowledge. For learners with high prior knowledge, worked examples have been found to be less effective, in contrast to learners with low prior knowledge whom benefit most from them. This phenomenon has been called the *expertise reversal effect* [6]. Learners with high levels of prior-knowledge do not need step-by-step instructions, which they can find redundant, leading to increases in extraneous cognitive load and reduced learning [27], whereas high-knowledge learners can benefit from fewer guided instructions such as that provided by problem-solving strategies [7] due to enhanced domain-specific working memory capacity, which allows learners to cope and benefit from the higher cognitive demands of problem solving [2,4]. It is domain-specific knowledge that makes the difference, rather than more general working-memory capacity (see [28]).

In addition to prior knowledge, a recent meta-analysis by Barbieri et al. [29] found the correctness of examples, timing of administration (practice or skill acquisition), and the use of self-explanation prompts to be moderating factors, whereas these factors were significant moderators and were implemented by the study designers, and were not necessarily dependent upon individual characteristics per se.

Although not exhaustive, the above section describes some of the factors that moderate the worked example effect, although few can be considered to be individual traits. However,

of interest to the present study is the finding that a mastery goal orientation—an individual motivational trait—has a potential moderating effect [9,10]. Before describing this research, some theoretical background is provided on goal orientation theory.

3. Goal Orientation Theory

Goal orientation theory is part of the broader field of motivation, which has been shown to be a crucial factor in effective learning (see [30–32]). Motivational theories are commonly used to explain student engagement, persistence, and performance. According to early goal orientation theory, a learner can either hold a mastery goal orientation (sometimes referred to as a learning goal orientation) or a performance goal orientation (see [8,33]). A mastery goal oriented (MAGO) learner aims to increase competence and has a desire to learn. A performance goal oriented (PGO) learner aims to prove their competence and has a desire to display their competence compared to others [34]. However, achievement goal orientation is now commonly recognised as consisting of four distinct achievement goals as a 2×2 framework: mastery approach, mastery avoidance, performance approach, and performance avoidance [35,36].

Studies examining the links between achievement goal orientation and achievement have found that a MAGO has been frequently viewed as the ideal goal orientation (see [37]). Individuals who have a MAGO have been found to be more enthusiastic about learning and more cognitively engaged [38]. Furthermore, Lee et al. [39] found that a MAGO can be associated with a greater working memory capacity in learning mathematics. However, some studies (see [40]) have failed to find a positive relationship between a MAGO and learning.

In contrast to a MAGO, a PGO has often been correlated with low academic achievement. Lee et al. [41] examined personality characteristics from self-determination theory and its influence on performance and levels of enjoyment based on achievement goal levels from university undergraduate students. Lee et al. [41] found mastery and performance approach goal orientation facilitated focus by focusing on a positive mindset. In addition, PGO learners have been found to withdraw from tasks when faced with the possibility of failure, show little interest in difficult tasks, and have the tendency to choose less challenging materials to increase their possibility of success [41]. However, research findings into PGO have also been inconsistent. For example, Elliot et al. [35] examined achievement goal orientation as predictors of study strategies and tested the study strategies as mediators between achievement goal orientation and exam performance on undergraduate students. They found a PGO to be positive predictors of surface processing, persistence, exam performance, and were not significantly associated with any study strategies; and mastery goals were positive predictors of deep processing and persistence, and were unrelated to surface processing and disorganisation.

Pintrich [42] investigated the role of mastery and performance goals and their associations with motivation, affect, strategy utilization, and academic performance among junior high school students. The results revealed that students who had both high mastery goal orientation and high performance goal orientation demonstrated an increased use of cognitive strategies for learning. Contrary to initial expectations, significant differences in anxiety, self-handicapping, and risk taking were not associated between students exhibiting high levels of both mastery goal orientation and performance goal orientation. However, it was noteworthy that the students with both high mastery goal orientation and performance goal orientation exhibited higher perception of task value, including perceived utility and significance of mathematics, compared to students who were high mastery goal orientation and low performance goal orientation.

3.1. Research Combining CLT and Mastery Goal Orientation

The present study focuses on only one aspect of achievement goal orientation theory: Mastery Approach Goal Orientation (MAGO) and how it impacts cognitive load. Some limited studies have investigated this relationship specifically. For example, a study by

Steele-Johnson et al. [43] found that a MAGO had beneficial effects on tasks with high cognitive load, whereas a performance-based orientation benefitted more from low cognitive load tasks. Cook et al. [44] found that in a medical simulation training activity, a MAGO was positively correlated with germane cognitive load. In other words, this goal orientation led to more mental effort applied to knowledge acquisition.

Studies into worked examples and goal orientations have also received some research interest. For example, Crippen et al. [45], using worked examples in a web-based learning study, found that although a MAGO was the best predictor of performance, it was not related to how often students used worked examples when given a choice. In contrast, a PGO was positively related to the use of worked examples. It was argued that students with a PGO preferred simpler tasks (as provided by worked examples), which enabled them to gain confidence through perceived lower cognitive load and allowed them to demonstrate success. For MAGO students, it was argued that they may not engage with worked examples if they believed that they would not lead to deeper understanding.

A small number of studies have examined the impact of goal orientations on learning from both worked examples and problem-solving strategies. A study by Belenky and Nokes-Malach [46] investigated goal orientations when comparing a *tell-and-practice* instructional approach (a type of worked examples) with an invention approach (a type of problem solving) on transfer outcomes. They also investigated the impact of an additional worked example imbedded in a test prior to a transfer task. They found that regardless of the original instructional format, the single worked example located before the transfer task enabled the task to be solved more easily than without a worked example. They also found that tell-and-practice (initial worked examples) instruction led to better procedural skill, but invention (problem solving) led to better transfer regardless of orientation (mastery or performance). However, an initial MAGO was required for successful transfer using the initial tell-and-practice approach.

A later study by Belenky and Nokes-Malach [9] also found that regardless of the original instructional format, the single worked example before a transfer task enabled the task to be solved more easily than without a worked example. In addition, students with a MAGO benefitted more from the worked example than a PGO. Evidence also emerged that the problem-solving instruction (invention) led to greater understanding and care about their answers than the worked example instruction. Students with high mastery levels had success with transfer regardless of instruction, whereas those with low mastery levels benefited more from an invention approach. It was argued that the problem-solving instruction led to the adoption of mastery-type goals and greater conceptual understanding, whereas the worked examples led to less attention on conceptual understanding due to a focus on the repetition of the non-transfer tasks. A further study by Belenky and Nokes-Malach [10] found that only a MAGO predicted transfer performance compared to the structure and framing of the learning tasks.

In terms of the benefits of worked examples and problem-solving methods, the findings reported in the previous section are somewhat mixed, but nevertheless indicate that a learner's goal orientation has been linked to the outcomes of both worked examples and problem solving. Hence, there is a strong motivation for investigating this topic further using a cognitive load theory framework.

3.2. The Current Study

The current study extends the research into worked examples by investigating how a mastery approach goal orientation interacts with the worked example effect. Previous research by Belenky and Nokes-Malach [9,10] into this topic used a sparse number of worked examples embedded in the tasks, whereas the present study uses a full set of worked examples during instruction consistent with most CLT studies (see [5]). Furthermore, both retention and transfer problems were investigated, as well as a measure of cognitive load. A further difference was that the previously noted studies did not measure prior knowledge in this domain, which we included to ensure that its overall impact would be investigated.

Mathematics, specifically algebra, was chosen as the learning topic, as it was expected to generate a worked-example effect, based on previous research (see [17,18,47]), and importantly has high element interactivity [11], with an expected medium level effect size (see the meta-analysis of Bichler et al. [28]).

3.3. Study Hypotheses

A number of measures were collected in the study. Participants completed a pre-test to assess *prior-knowledge*, a post-test to measure *retention* of the information, and a *transfer* test to see if the knowledge gained could be applied to more complex tasks. A difficulty-rating measure was collected to gain an assessment (index) of cognitive load (see [48–50]). Participants were also measured for their MAGO using a three-item scale based on Elliot and Murayama's survey [51].

Based on the literature outlined above, a number of hypotheses were formulated. Firstly, the extensive research into worked examples indicates that for novice learners in particular, worked examples lead to greater retention of information with reduced cognitive load compared to problem solving. Regarding transfer, even though there are studies that have shown that problem-solving strategies such as invention (see [9]) can lead to transfer success, the majority of studies into CLT have found that worked examples leads to greater transfer than problem solving (see [5]). Therefore, it was predicted that:

Hypothesis 1. *Worked examples compared to problem solving would lead to greater (a) retention, (b) transfer, and (c) lower cognitive load.*

Regarding a MAGO, the evidence reported above suggests that having a MAGO is correlated with successful learning [52], including transfer [9,10] as well as motivation [53]. Hence, it was predicted that:

Hypothesis 2. *A high compared to a low MAGO would lead to greater (a) retention, (b) transfer, and (c) lower cognitive load.*

Of great interest were the possible interactions between worked examples and a mastery goal orientation. Due to limited studies in this domain and conflicting findings, no specific hypotheses were made, but were investigated as an open research question:

To what extent is the worked example effect moderated by a mastery goal orientation?

4. Method

4.1. Participants

Ninety-eight grade-9 boys (mean age of 13.9 years) from a high school in Sydney participated in the experiment. These participants were randomly allocated into one of two experimental groups: worked examples ($n = 50$) or problem solving ($n = 48$).

4.2. Materials

Study phases. The experiment consisted of 4 phases: a MAGO survey, a prior knowledge test (pre-test), an acquisition phase, and a test phase (post-test and transfer).

The MAGO survey. At the start of the experiment, participants were given three items from the Achievement Goal Orientation-Revised (AGO-R) questionnaire to measure their mastery approach goal orientation [51]. It should be noted that it is quite common to use parts of a more comprehensive survey to examine specific variables such as a MAGO (see [54,55]). The 3 items were modified to focus specifically on mathematics. For example, the item "My aim is to completely master the material presented in this class" was modified to "My aim is to completely master mathematics". The three items were finalised as:

- (1) My aim is to completely master mathematics.
- (2) My goal was to learn mathematics as much as possible.
- (3) I am striving to understand the content of mathematics as thoroughly as possible.

Each item was scored on a 5-point Likert scale ranging from Strongly Disagree (−2) to Strongly Agree (2). The 3 ratings were added together to give a final score ranging from a possible minimum of −6 to a maximum of 6. A Cronbach alpha score was 0.70, indicating an acceptable degree of reliability [56].

Pre-test. To test for prior knowledge, a pre-test was conducted consisting of a set of six algebra problems of an advanced nature for the age group. Each problem contained one unknown variable that had to be calculated (e.g., see Figure 1). The solutions involved substitution and the re-arrangement of the formula, and other algebraic transformations. There were three types of problems. The first problem type was an algebra question that required the participants to understand the square root function applied to an algebraic function (e.g., $\sqrt{\frac{x}{5}}$). The second type of problem required an understanding of an algebraic cubic power (i.e., x^3). The final type of problem required an understanding of algebraic fractions (e.g., $\frac{x}{1-b}$). The pre-test consisted of 2 questions of each type, and was presented on A4 sheets of paper.

Study the follow problem	Solve the following problem
Find x , if $y = 4$ and $z = 3$	Find a , if $b = 4$ and $c = 5$
Where $y = z \sqrt{\frac{x}{5}}$	Where $b = c \sqrt{\frac{a}{4}}$
<u>Working</u>	
$y = z \sqrt{\frac{x}{5}}$	
$4 = 3 \sqrt{\frac{x}{5}}$	
$\frac{4}{3} = \sqrt{\frac{x}{5}}$	
$\frac{16}{9} = \frac{x}{5}$	
$\frac{80}{9} = x$	
$\therefore x = \frac{80}{9}$	

Figure 1. A worked example problem pair.

Using a pre-test enabled us to screen for students with high prior knowledge, which might have led to expertise-reversal effects by reducing the effectiveness of worked examples [6] The low scores on this test (see the Section 5) indicated that, overall, students had low prior knowledge in this domain, and no individual was excluded.

Acquisition phase. All participants were required to learn about the algebra topic either using worked examples or by being left to their own problem-solving strategies. The worked-example instrument consisted of 6 problem pairs, consisting of 6 step-by-step solutions and 6 similar problems to solve. The use of problem pairs in this fashion has been shown to be a highly effective way of using worked examples (see [5]). Using this format, participants were required to study a worked example and then solve a similar problem immediately. Figure 1 provides an example of a problem pair that was presented on a single sheet of A4 paper. The only difference between each of the worked examples

and the corresponding problem to be solved were the variable terms (e.g., y or b) and the numbers to be substituted into the variables.

The problem-solving material consisted of the 12 identical problems used in the worked examples material, except the 6 step-by-step solutions that were provided in the worked examples material that were not provided in the problem-solving material. Therefore, these participants were required to solve all twelve algebra questions without any guidance. However, both groups were provided with final answers (but no solution steps) to all problems that participants were required to solve themselves without guidance. Hence, some feedback was provided, as the participants' answers could be checked.

Test phase. The test phase consisted of a *post-test*, which had six problems identical to the pre-test, and a *transfer* test consisting of six questions. The six transfer questions were also algebraic in nature, but more complex than the post-test questions, involving more variables to be manipulated.

Cognitive load measure. To measure cognitive load following the acquisition phase, a single-item difficulty scale was constructed. Participants were asked to indicate the level of perceived difficulty experienced during this learning phase using a 9-point Likert scale, consisting of: "extremely easy (1), very easy (2), fairly easy (3), slightly easy (4), neither easy or difficult (5), slightly difficult (6), fairly difficult (7), very difficult (8), and extremely difficult (9)". This scale was a modified version of the original Paas scale (see [49]), which used a 9-point scale based on mental effort. Both difficulty and mental effort have been used extensively in cognitive load theory research to obtain a reliable index of cognitive load (see [48,50]).

4.3. Procedure

After an introduction explaining the purpose and format of the study, participants were given the MAGO survey to complete on a single sheet of paper (3 min). Secondly, they were given the prior knowledge pre-test for a maximum of 10 min. Participants were then given 30 min for the acquisition phase. Once the participants completed the acquisition materials, they were instructed to indicate the level of perceived difficulty experienced. Finally, they were given 20 min for the test phase. All phases were completed in a single session of about 70 min. After each phase, all materials and responses were collected.

4.4. Scoring of the Tests

The pre-test and post-test were identical, consisting of 6 problems to be solved. For each correct answer, 2 marks were awarded, giving a maximum score of 12 marks for these tests. Similarly, for the 6 transfer questions, a maximum score of 2 was awarded for each problem, resulting in a maximum score of 12. A number of partial marks were awarded for final incorrect answers on all tests that contained some correct steps, such as successful substitution and algebraic transformation, according to a strict marking rubric. Two experienced researchers graded all student answers, and when differences arose, they were resolved in order to reach full agreement. Cronbach alpha scores on the pre-test, post-test, and transfer test were 0.79, 0.90, and 0.87, respectively, indicating a high degree of reliability.

5. Results

5.1. Improvements in Learning from Pre-Test to Post-Test

To investigate whether learning occurred from pre-test to post-test, which were identical, t -tests were conducted on the combined worked examples and problem-solving groupings. A significant improvement was found in the worked examples group from the pre-test ($M = 2.21$, $SD = 2.13$) to the post-test ($M = 4.90$, $SD = 3.90$): $t(50) = 56.45$, $p \leq 0.01$, $d = 0.90$. The problem-solving group also improved from pre-test ($M = 1.94$, $SD = 1.81$) to post-test ($M = 2.89$, $SD = 2.38$), $t(47) = 12.17$, $p < 0.01$, $d = 0.45$. Thus, both strategies resulted in learning gains.

5.2. Regression Analyses

To test the hypotheses, stepwise linear regression was used to analyse the data for each dependent variable (retention test, transfer, and cognitive load measures). Potential predictors included prior knowledge (pre-test), mastery level, learning strategy (worked example vs. problem solving), and all interactions. Examination of correlational data and initial regression analysis revealed a number of significant predictors across the different measures, as reported below. In the following analyses, R^2 represents the explained variance by Step 1, ΔR^2 represents the change in variance on subsequent steps, and β represents the standardized coefficients.

Retention test. For the retention post-test, a two-factor model was identified where prior knowledge ($\beta = 0.68, p < 0.001, R^2 = 0.49$) and the mastery–worked examples interaction ($\beta = 0.25, p < 0.001, \Delta R^2 = 0.06$) were found to be significant (Table 1). To investigate this interaction, analyses were conducted by splitting participants into two groups based on the median mastery scores. Scores for the Low MAGO group ranged from -1 to 3 , and the High MAGO group ranged from 4.0 to 6.5 .

Table 1. Descriptive statistics and intercorrelations among variables of retention scores.

Variable	<i>n</i>	<i>M</i>	<i>SD</i>	Variable		
				1	2	3
Retention test	98	3.86	3.34	-		
Prior knowledge	98	2.032	1.94	0.000 **	-	
Mastery Approach Goal orientation	98	3.28	1.89	0.006	0.002	-
Mastery Approach $\times$ Worked Examples	98	4.97	1.66	0.002	0.320	0.214 **

Note: ** $p < 0.001$.

Clearly, prior knowledge impacted on each condition, and therefore it was used as a covariate to control for its impact. Hence, a 2 (mastery levels) $\times$ 2 (Worked examples vs. problem solving) ANCOVA was conducted: $F(1, 57) = 7.16, p = 0.01, \eta_p^2 = 0.11$. Simple effects tests revealed a significant difference in the higher MAGO group, $F(1, 25) = 8.60, p = 0.01, \eta_p^2 = 0.26$, where the worked examples group ($M = 4.69, SE = 3.68$) outperformed the problem solving group ($M = 1.50, SE = 1.49$). There was no significant difference in the low MAGO group, $F < 1$, whereas the worked-example group ($M = 2.06, SE = 1.78$) scored less than the problem-solving group ($M = 2.59, SE = 1.39$), but not significantly.

Cognitive load measure. For the cognitive load measures, a three-factor model was identified where prior knowledge, ($\beta = -0.36, p < 0.001, R^2 = 0.21$), mastery–worked examples interaction ($\beta = -0.37, p < 0.001, \Delta R^2 = 0.14$), and mastery levels ($\beta = -0.25, p < 0.01, \Delta R^2 = 0.07$) were found to be significant factors (Table 2). The negative β -values indicated that high prior knowledge and a high mastery level led to reduced cognitive load.

Table 2. Descriptive statistics and intercorrelations among variables of cognitive load measure.

Variable	<i>n</i>	<i>M</i>	<i>SD</i>	Variable		
				1	2	3
Cognitive Load	98	7.51	2.00	-		
Prior knowledge	98	2.032	1.94	0.000 **	-	
Mastery Approach Goal orientation	98	3.28	1.89	0.000	0.002 *	-
Mastery Approach $\times$ Worked Examples	98	4.97	1.66	0.000	0.320	0.214 **

Note: * $p < 0.01$, ** $p < 0.001$.

The mastery–worked examples interaction was investigated further with an ANCOVA, using the same format as applied to the Retention analysis. For the high MAGO, there was a significant effect: $F(1, 45) = 12.90, p < 0.01, \eta_p^2 = 0.22$, with the combined problem-solving groups reporting higher levels of difficulty ($M = 8.00, SE = 0.42$) than the worked example groups ($M = 5.86, SE = 0.41$). A significant main effect was also found for the low MAGO group, $F(1, 47) = 11.78, p < 0.01, \eta_p^2 = 0.20$, with the problem-solving group ($M = 8.60, SE = 0.20$) reporting higher levels of difficulty than the worked-example group ($M = 7.62, SE = 0.20$). For both mastery groups, problem solving led to greater cognitive load, as measured through the difficulty scale, compared to worked examples. However, this interaction was caused by a larger worked-example advantage (lower cognitive load) for the high MAGO group, as shown by the differences in the group means reported above.

Transfer test. For the transfer test, a one-factor model was identified, with only prior knowledge as a significant predictor ($\beta = 0.43, p < 0.001, R^2 = 0.65$).

Relationship between MAGO scores and prior knowledge. A significant correlation was found between MAGO and prior knowledge ($r = 0.29, p < 0.01$), which is consistent with previous research showing this relationship (see [57]).

5.3. Testing the Hypotheses

Hypothesis 1a predicted a worked-example effect, in that worked examples would be superior to problem solving for retention problems. This was partially confirmed in that the worked-example effect was moderated by MAGO levels. A worked-example effect was found only for learners with the higher level of MAGO, but not the lower level. Hypotheses 1c was confirmed, as, overall, worked examples reduced cognitive load, as measured through the difficulty scale, although the effect was more pronounced at the higher MAGO level. Hypothesis 1b was not supported, as the only influence on the transfer problems was prior knowledge. Hypothesis 2a and 2b predicted a higher MAGO would lead to greater retention and transfer, but both of these predictions were not supported. However, participants with a higher MAGO experienced less cognitive load during instruction.

6. General Discussion

An aim of the study was to use learning content that would lead to a worked example effect. This aim was partially achieved, as participants with high levels of mastery whom learned from worked examples scored higher on retention problems than those who followed their own problem-solving strategies during the acquisition phase. This result is consistent with the worked-examples effect (see [4]). However, the interaction between mastery levels and the use of worked examples is the most significant and novel finding in this study. Learners with high MAGO levels benefited most from worked examples compared to problem solving than learners with lower MAGO levels. It may be that high MAGO learners had a greater desire to engage in worked examples consistent with a desire to master the materials (see [37,38]), and may view worked examples as the superior learning strategy. Furthermore, the worked-examples strategy of “study a solution to a problem, then solve a similar one”, originally devised by Sweller and Cooper [18] for motivational purposes (even though they never tested this assumption), may have a further positive effect on MAGO learners, whereas other formats of worked examples have been successfully used from a learning perspective (see [58]), but they have not been compared from a motivational perspective. Hence, a research avenue could further investigate the link between motivation and a worked-example format.

This interaction is also interesting because it does not mirror the usual results relating to expertise. The expertise reversal effect dictates that as expertise in the domain increases, worked examples lose their effectiveness [6]. However, the high MAGO learners were found to have higher levels of prior knowledge, as measured in the pre-test, than those with lower mastery levels, but no expertise reversal effect was found. In fact, an opposite effect of a reversal was found. It was notable that, in this study, prior knowledge of all participants combined was very low (mean = 2.0 from a maximum of 12 on the pre-test), suggesting

that perhaps a minimum level of prior knowledge is required for worked examples to have the most impact.

The study found that worked examples did not lead to a direct advantage on the transfer test. The only factor impacting on transfer in this study was prior knowledge, suggesting that the interventions did not increase prior knowledge sufficiently to make a significant impact. Previous research [9,10,46] into mastery orientations and worked examples has not included prior-knowledge measures, and therefore the present study extends their research further by including this factor, but also demonstrates how critical it is to include it. The early research into worked examples (see [17]) found that a significant amount of exposure to worked examples was needed to achieve the required automation of fundamental knowledge and skills that lead to transfer effects (see also [58,59]). Hence, a greater intervention period in this study may have been needed to promote transfer.

There are also a number of other potential explanations into why worked examples did not promote transfer. Belenky and Nokes-Malach [9] suggested that worked examples may fail to promote transfer because learners may pay less attention to conceptual understanding due to a focus on repetition of the non-transfer tasks. Although learners who initially had to learn through problem solving also did not show transfer advantages, some previous studies had found that problem-solving strategies, such as invention (see [9]), can lead to transfer. However, facilitating the development of invention through specific activities, even though it is a form of problem solving, is very different to the totally non-directed problem solving used here. Furthermore, some of the best results for invention activities were obtained in conjunction with a worked example (see [9]). Hence, more directed problem solving strategies (see [60]), as well as greater time [59], may be required to bridge the prior knowledge gap for transfer improvements in this domain.

Furthermore, Schwartz and colleagues [61,62] suggest two views of transfer: one as the direct application of previously learned knowledge to a problem question, and the other as the preparation for future learning. In our study, the direct application of previously learned procedural and conceptual knowledge was needed to answer the transfer question. Similarly to our study, several studies have also reported no transfer or mixed results (see [48]). Clearly, further research is required to test these possible explanations into a lack of transfer effects.

Regarding mastery effects, having a high MAGO did not lead to greater test scores than those with lower MAGO levels. Although a high MAGO has been frequently associated with higher achievement (see [52]), some research (see [40]) has also failed to find a positive relationship between MAGO and academic performance. However, a high MAGO did lower cognitive load during instruction. This may be due to high MAGO learners having a greater working memory capacity, as proposed by [41], as well as a desire to be more cognitively engaged in learning (see [38]).

Study Limitations

Firstly, only males were used in the study. Males who have extrinsic motivation tend to perform poorly with a lack of self-regulation (see [63]) and are less likely than girls to hold mastery goal orientations [64]. Therefore, further studies need to include females to see if the results are replicated.

Secondly, it is worth noting that we used the exact same pre-test and post-test. Hence, each group had more experience of problem solving during the pre-test, which could have led to pre-instruction learning. However, there is no evidence that this impacted on the overall results, as the predicted worked-example effect occurred. Future research might consider using a different test.

Thirdly, we focused only on a mastery goal orientation. Achievement goal orientation research also includes a performance goal orientation, as well as individual desires to 'approach' the desired state or 'avoid' the undesired state ([36]). Hence, future research could also include scales to gain more insights into the cognitive processes at work in similar instructional environments.

Fourthly, we used a single measure of cognitive load (difficulty scale) based on the original Paas scale [49]. Whereas this has been used successfully in multiple CLT studies [48,50], it is a perceived level of difficulty and corresponds to an index of total cognitive load only. Hence, future research in this domain could measure individual cognitive loads aligning with intrinsic, germane, and extraneous loads (see [65]), which could provide more insights.

7. Conclusions

There were some novel directions taken in this study. Little research has been conducted that combines worked examples with goal orientation theory. The present study used a learning environment that combined classical worked-example designs that feature the alternation of a worked example with a similar problem to solve (see [4]). As expected, compared to problem solving, worked examples showed a number of advantages, including superior retention of information and lowered cognitive load during instruction. However, the retention learning advantage was confined to learners with high mastery levels only. This finding that only MAGO learners were motivated with a desire to improve and master skills within the worked-example framework is, as far as we know, a unique finding, demonstrating that investigating the impact of goal orientations on worked examples and problem-solving strategies has great potential and is worthy of further research. Judging by these initial findings, such research could potentially identify a new CLT effect: *A mastery goal worked-example effect*. Furthermore, this research has extended a recent trend to combine CLT research with motivational aspects (see [53,55,66,67]), which opens up new avenues for research that can help in understanding how to design materials to enhance learning and instruction. Interactions between CLT and motivation have been rarely studied until more recent times. The aim of the present study was to continue this recent trend by investigating how learners with a mastery achievement goal orientation (see [8]) respond to learning from worked examples and problem solving [17]. Furthermore, the study also demonstrated the importance of prior knowledge to new learning scenarios, as it had the largest impact on all the measures collected, especially transfer.

Author Contributions: Conceptualization, H.M.L. and P.A.; methodology, H.M.L. and P.A.; formal analysis, H.M.L. and P.A.; investigation, H.M.L. and P.A.; data curation, H.M.L.; writing—original draft preparation, H.M.L. and P.A.; writing—review and editing, H.M.L. and P.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of University of New South Wales (13,067, July 2013).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data could be shared by contacting with the first author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sweller, J.; van Merriënboer, J.J.G.; Paas, F. Cognitive architecture and instructional design: 20 years later. *Educ. Psychol. Rev.* **2019**, *31*, 261–292. [CrossRef]
2. Tricot, A.; Sweller, J. Domain-specific knowledge and why teaching generic skills does not work. *Educ. Psychol. Rev.* **2014**, *26*, 265–283. [CrossRef]
3. Atkinson, R.K.; Derry, S.J.; Renkl, A.; Wortham, D. Learning from examples: Instructional principles from the worked examples research. *Rev. Educ. Res.* **2000**, *70*, 181–214. [CrossRef]
4. Sweller, J.; Ayres, P. Worked examples. In *The Oxford Handbook of Educational Psychology*; O'Donnell, A., Barnes, N.C., Reeve, J.M., Eds.; Oxford University Press: Oxford, UK, 2022. [CrossRef]
5. Sweller, J.; Ayres, P.; Kalyuga, S. *Cognitive Load Theory*; Springer: Berlin/Heidelberg, Germany, 2011. [CrossRef]
6. Kalyuga, S.; Ayres, P.; Chandler, P.; Sweller, J. The expertise reversal effect. *Educ. Psychol.* **2003**, *38*, 23–31. [CrossRef]

7. Tuovinen, J.E.; Sweller, J. A comparison of cognitive load associated with discovery learning and worked examples. *J. Educ. Psychol.* **1999**, *91*, 334–341. [CrossRef]
8. Dweck, C. Motivational processes affecting learning. *Am. Psychol.* **1986**, *41*, 1040–1104. [CrossRef]
9. Belenky, D.M.; Nokes-Malach, T.J. Motivation and transfer: The role of mastery-approach goals in preparation for future learning. *J. Learn. Sci.* **2012**, *21*, 399–432. [CrossRef]
10. Belenky, D.M.; Nokes-Malach, T.J. Mastery-approach goals and knowledge transfer: An investigation into the effects of task structure and framing instructions. *Learn. Individ. Differ.* **2013**, *25*, 21–34. [CrossRef]
11. Chen, O.; Paas, F.; Sweller, J. A cognitive load theory approach to defining and measuring task complexity through element interactivity. *Educ. Psychol. Rev.* **2023**, *35*, 63. [CrossRef]
12. Greenberg, K.; Zheng, R. Revisiting the debate on germane cognitive load versus germane resources. *J. Cogn. Psychol.* **2023**, *35*, 295–314. [CrossRef]
13. Paas, F.; van Merriënboer, J.J. Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks. *Curr. Dir. Psychol. Sci.* **2020**, *29*, 394–398. [CrossRef]
14. Sweller, J.; van Merriënboer, J.J.G.; Paas, F. Cognitive architecture and instructional design. *Educ. Psychol. Rev.* **1998**, *10*, 251–296. [CrossRef]
15. Sweller, J. Cognitive load during problem solving: Effects on learning. *Cogn. Sci.* **1988**, *12*, 257–285.
16. Renkl, A. The worked example principle in multimedia learning. In *The Cambridge Handbook of Multimedia Learning*, 3rd ed.; Mayer, R.E., Fiorella, L., Eds.; Cambridge University Press: Cambridge, UK, 2022; pp. 231–242.
17. Cooper, G.; Sweller, J. The effects of schema acquisition and rule automation on mathematical problem-solving transfer. *J. Educ. Psychol.* **1987**, *79*, 347–362. [CrossRef]
18. Sweller, J.; Cooper, G. The use of worked examples as a substitute for problem solving in learning algebra. *Cogn. Instr.* **1985**, *2*, 59–89. [CrossRef]
19. Paas, F.G.W.C.; van Merriënboer, J.J.G. The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Hum. Factors* **1993**, *35*, 737–743. [CrossRef]
20. Retnowati, E.; Ayres, P.; Sweller, J. Can collaborative learning improve the effectiveness of worked examples in learning mathematics? *J. Educ. Psychol.* **2017**, *109*, 666–679. [CrossRef]
21. Retnowati, E.; Ayres, P.; Sweller, J. Collaborative learning effects when students have complete or incomplete knowledge. *Appl. Cogn. Psychol.* **2018**, *32*, 681–692. [CrossRef]
22. Renkl, A.; Atkinson, R.K.; Große, C.S. How fading worked solution steps works—A cognitive load perspective. *Instr. Sci.* **2004**, *32*, 59–82. [CrossRef]
23. Bokosmaty, S.; Sweller, J.; Kalyuga, S. Learning geometry problem solving by studying worked examples: Effects of learner guidance and expertise. *Am. Educ. Res. J.* **2015**, *52*, 307–333. [CrossRef]
24. Ayres, P. Can the isolated-elements strategy be improved by targeting points of high cognitive load for additional practice? *Learn. Instr.* **2013**, *23*, 115–124. [CrossRef]
25. Chen, O.; Retnowati, E.; Kalyuga, S. Effects of worked examples on step performance in solving complex problems. *Educ. Psychol.* **2019**, *39*, 188–202. [CrossRef]
26. Grosse, C.S.; Renkl, A. Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learn. Instr.* **2007**, *17*, 612–634. [CrossRef]
27. Kalyuga, S.; Sweller, J. The redundancy principle in multimedia learning. In *The Cambridge Handbook of Multimedia Learning*, 3rd ed.; Mayer, R.E., Fiorella, L., Eds.; Cambridge University Press: Cambridge, UK, 2022; pp. 212–220. [CrossRef]
28. Bichler, S.; Schwaighofer, M.; Stadler, M.; Bühner, M.; Greiff, S.; Fischer, F. How working memory capacity and shifting matter for learning with worked examples—A replication study. *J. Educ. Psychol.* **2020**, *112*, 1320–1337. [CrossRef]
29. Barbieri, C.A.; Miller-Cotto, D.; Clerjuste, S.N.; Chawla, K. A Meta-analysis of the Worked Examples Effect on Mathematics Performance. *Educ. Psychol. Rev.* **2023**, *35*, 11. [CrossRef]
30. Covington, M.V. Goal theory, motivation, and school achievement: An integrative review. *Annu. Rev. Psychol.* **2000**, *51*, 171–200. [CrossRef]
31. Maehr, M.L.; Midgley, C. Enhancing student motivation: A schoolwide approach. *Educ. Psychol.* **1991**, *26*, 399–427. [CrossRef]
32. Senko, C. When do mastery and performance goals facilitate academic achievement? *Contemp. Educ. Psychol.* **2019**, *59*, 101795. [CrossRef]
33. Ames, C. Classrooms: Goals, structures, and student motivation. *J. Educ. Psychol.* **1992**, *84*, 261–271. [CrossRef]
34. Murayama, K.; Elliot, A.J. Achievement motivation and memory: Achievement goals differentially influence immediate and delayed remember-know recognition memory. *Personal. Soc. Psychol. Bull.* **2011**, *37*, 1339–1348. [CrossRef]
35. Elliot, A.; McGregor, H.; Gable, S. Achievement goals, study strategies, and exam performance: A mediational analysis. *J. Educ. Psychol.* **1999**, *91*, 549–563. [CrossRef]
36. Elliot, A.; McGregor, H. A 2 × 2 achievement goal framework. *J. Personal. Soc. Psychol.* **2001**, *80*, 501–519. [CrossRef]
37. Hulleman, C.S.; Schrager, S.M.; Bodmann, S.M.; Harackiewicz, J.M. A meta-analytic review of achievement goal measures: Different labels for the same constructs or different constructs with similar labels? *Psychol. Bull.* **2010**, *136*, 422–449. [CrossRef]

38. Pintrich, P.R.; Schrauben, B. Students' motivational beliefs and their cognitive engagement in classroom academic tasks. In *Student Perceptions in the Classroom*; Schunk, D.H., Meece, J.L., Eds.; Lawrence Erlbaum Associates Inc.: Mahwah, NJ, USA, 1992; pp. 149–183.
39. Lee, F.K.; Ning, F.; Goh, H.C. Interaction between cognitive and noncognitive factors: The influences of academic goal orientation and working memory on mathematical performance. *Educ. Psychol.* **2014**, *34*, 73–91. [CrossRef]
40. Neroni, J.; Meijs, C.; Leontjevas, R.; Kirschner, P.A.; De Groot, R.H.M. Goal orientation and academic performance in adult distance education. *Int. Rev. Res. Open Distrib. Learn. (IRRODL)* **2018**, *19*, 192–208. [CrossRef]
41. Lee, F.K.; Sheldon, K.M.; Turban, D.B. Personality and the goal-striving process: The influence of achievement goal patterns, goal level and mental focus on performance and enjoyment. *J. Appl. Psychol.* **2003**, *88*, 256–265. [CrossRef]
42. Pintrich, P.R. Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *J. Educ. Psychol.* **2000**, *92*, 544–555. [CrossRef]
43. Steele-Johnson, D.; Beauregard, R.S.; Hoover, P.B.; Schmidt, A.M. Goal orientation and task demand effects on motivation, affect, and performance. *J. Appl. Psychol.* **2000**, *85*, 724–738. [CrossRef]
44. Cook, D.A.; Castillo, R.M.; Gas, B.; Artino, A.R., Jr. Measuring achievement goal motivation, mindsets and cognitive load: Validation of three instruments' scores. *Med. Educ.* **2017**, *51*, 1061–1074. [CrossRef]
45. Crippen, K.J.; Biesinger, K.D.; Muis, K.R.; Orgill, M. The role of goal orientation and self-efficacy in learning from web-based worked examples. *J. Interact. Learn. Res.* **2009**, *20*, 385–403.
46. Belenky, D.; Nokes, T. Motivation and transfer: The role of achievement goals in preparation for future learning. *Proc. Annu. Meet. Cogn. Sci. Soc.* **2009**, *31*, 31. Available online: <https://escholarship.org/uc/item/9mt6w2b8> (accessed on 20 January 2024).
47. Manson, E.; Ayres, P. Investigating how errors should be flagged and worked examples structured when providing feedback to novice learners of mathematics. *Educ. Psychol.* **2021**, *41*, 153–171. [CrossRef]
48. Ayres, P.; Lee, J.Y.; Paas, F.; van Merriënboer, J.J. The validity of physiological measures to identify differences in intrinsic cognitive load. *Front. Psychol.* **2021**, *12*, 702538. [CrossRef]
49. Paas, F.G. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *J. Educ. Psychol.* **1992**, *84*, 429–434. [CrossRef]
50. Ayres, P. Subjective measures of cognitive load: What can they reliability measure? In *Cognitive Load Measurement and Application: A Theoretical Framework for Meaningful Research and Practice*; Zheng, R.Z., Ed.; Routledge: New York, NY, USA, 2018; pp. 9–28. [CrossRef]
51. Elliot, A.J.; Murayama, K. On the measurement of achievement goals: Critique, illustration, and application. *J. Educ. Psychol.* **2008**, *100*, 613–628. [CrossRef]
52. Payne, S.C.; Youngcourt, S.S.; Beaubien, J.M. A meta-analytic examination of the nomological net. *J. Appl. Psychol.* **2007**, *92*, 128–150. [CrossRef]
53. Gupta, U.; Zheng, R.Z. Cognitive load in solving mathematics problems: Validating the role of motivation and the interaction among prior knowledge, worked examples, and task difficulty. *Eur. J. STEM Educ.* **2020**, *5*, 5. [CrossRef]
54. Guo, J.; Hu, X.; Elliot, A.J.; Marsh, H.W.; Murayama, K.; Basarkod, G.; Parker, P.D.; Dicke, T. Mastery-approach goals: A large-scale cross-cultural analysis of antecedents and consequences. *J. Personal. Soc. Psychol.* **2023**, *125*, 397–420. [CrossRef]
55. Xu, K.M.; Koorn, P.; de Koning, B.; Skuballa, I.T.; Lin, L.; Henderikx, M.; Marsh, H.W.; Sweller, J.; Paas, F. A growth mindset lowers perceived cognitive load and improves learning: Integrating motivation to cognitive load. *J. Educ. Psychol.* **2020**, *113*, 1177–1191. [CrossRef]
56. George, D.; Mallery, P. *SPSS for Windows Step by Step: A Simple Guide and Reference. 11.0 Update*, 4th ed.; Allyn & Bacon: Boston, MA, USA, 2003.
57. Bell, B.S.; Kozlowski, W.J. Goal orientation and ability: Interactive effects on self-efficacy, performance, and knowledge. *J. Appl. Psychol.* **2002**, *87*, 497–505. [CrossRef]
58. Detterman, D.K. The case for the prosecution: Transfer as an epiphenomenon. In *Transfer on Trial: Intelligence, Cognition, and Instruction*; Detterman, D.K., Sternberg, R.J., Eds.; Ablex Publishing: Norwood, NJ, USA, 1993; pp. 1–24.
59. Schwaighofer, M.; Fischer, F.; Bühner, M. Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educ. Psychol.* **2015**, *50*, 138–166. [CrossRef]
60. Youssef-Shalala, A.; Ayres, P.; Schubert, C.; Sweller, J. Using a general problem-solving strategy to promote transfer. *J. Exp. Psychol. Appl.* **2014**, *20*, 215–231. [CrossRef]
61. Bransford, J.D.; Schwartz, D.L. Chapter 3: Rethinking transfer: A simple proposal with multiple implications. *Rev. Res. Educ.* **1999**, *24*, 61–100.
62. Schwartz, D.L.; Bransford, J.D.; Sears, D. Efficiency and innovation in transfer. *Transf. Learn. A Mod. Multidiscip. Perspect.* **2005**, *3*, 1–51. [CrossRef]
63. Patrick, H.; Ryan, A.M.; Pintrich, P.R. The differential impact of extrinsic and mastery goal orientations on males' and females' self-regulated learning. *Learn. Individ. Differ.* **1999**, *11*, 153–171.
64. Kenney-Benson, G.A.; Pomerantz, E.M.; Ryan, A.M.; Patrick, H. Sex differences in math performance: The role of children's approach to schoolwork. *Dev. Psychol.* **2006**, *42*, 11–26. [CrossRef]
65. Leppink, J.; Paas, F.; van der Vleuten, C.P.M.; van Gog, T.; van Merriënboer, J.J.G. Development of an instrument for measuring different types of cognitive load. *Behav. Res. Methods* **2013**, *45*, 1058–1072. [CrossRef]

66. Feldon, D.F.; Franco, J.; Chao, J.; Peugh, J.; Maahs-Fladung, C. Self-efficacy change associated with a cognitive load-based intervention in an undergraduate biology course. *Learn. Instr.* **2018**, *56*, 64–72. [CrossRef]
67. Martin, A.J.; Evans, P. Load reduction instruction: Sequencing explicit instruction and guided discovery to enhance students' motivation, engagement, learning, and achievement. In *Advances in Cognitive Load Theory: Rethinking Teaching*; Ford, S.T., Agostinho, S., Sweller, J., Eds.; Routledge: London, UK, 2019. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

The Role of Cognitive Learner Prerequisites for Cognitive Load and Learning Outcomes in AR-Supported Lab Work

Kristin Altmeyer ^{1,*}, Roland Brünken ¹, Jochen Kuhn ² and Sarah Malone ¹¹ Department of Education, Saarland University, 66123 Saarbrücken, Germany² Faculty of Physics, Ludwig-Maximilians-Universität München (LMU Munich), 80539 Munich, Germany

* Correspondence: kristin.altmeyer@uni-saarland.de

Abstract: Augmented Reality (AR) can enhance student-centered lab work by bridging the spatial and temporal split between virtual information and observed real-world phenomena. While the Cognitive Theory of Multimedia Learning and the Cognitive Load Theory suggest that AR can reduce extraneous cognitive load (ECL) and foster learning, the empirical results remain inconsistent. This re-analysis of three related studies with different target groups and AR devices explores whether learners' spatial abilities and verbal working memory capacity moderate the effect of AR support in lab work settings on ECL and conceptual knowledge gains. Although these moderators could not be confirmed consistently, the results indicate that tablet-based AR holds the potential to support learners with low spatial abilities. Moreover, low verbal working memory learners were demonstrated to be particularly vulnerable to the spatial contiguity failure that can be caused by smartglasses AR. Moderation effects were only observed for ECL but not for conceptual knowledge gains. The findings highlight that the benefit of AR support can depend on learners' cognitive prerequisites and additional contextual factors, such as the AR device used and the age of the target group. The design and implementation of AR-supported lab work environments should account for these factors to optimize the learning outcomes.

Keywords: augmented reality; science education; lab work; cognitive load; individual differences; spatial ability; verbal working memory

Citation: Altmeyer, K.; Brünken, R.; Kuhn, J.; Malone, S. The Role of Cognitive Learner Prerequisites for Cognitive Load and Learning Outcomes in AR-Supported Lab Work. *Educ. Sci.* **2024**, *14*, 1161. <https://doi.org/10.3390/educsci14111161>

Academic Editor: Mike Joy

Received: 1 July 2024

Revised: 11 October 2024

Accepted: 11 October 2024

Published: 25 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Considered a key feature of scientific work, hands-on experiments are an essential part of science education. Augmented Reality (AR) can support this instructional approach by integrating virtual learning information into real-world experimental setups. The resulting real-time spatial integration of real phenomena and virtual information is assumed to reduce split attention, potentially reducing learning-irrelevant cognitive load and contributing to improved learning outcomes in science experiments (e.g., [1]). In the course of an interdisciplinary research project on the use of AR to enhance physics lab work, we empirically investigated this theoretical assumption in three studies. While the three studies were based on similar AR-supported learning environments, they varied in the AR technology employed and the sample characteristics. Contrary to theoretical assumptions about the effects of AR, we found mixed, negligible, or even contradictory results on cognitive load across the three studies. This aligns with a recent research synthesis also reporting contradictory effects of AR on cognitive load levels [2]. Moreover, in contrast to the numerous meta-analyses that have already demonstrated the potential of AR for enhancing learning (e.g., [3,4]), our studies showed ambiguous effects of AR on learning outcomes. Our findings led to the conclusion that the potential of AR to improve learning processes may depend on a variety of factors, including technical and pedagogical implementation but also learner prerequisites. Prior studies have indicated that spatial abilities and verbal working memory capacity are critical in processing information from various

sources (e.g., [5]). Although these cognitive learner prerequisites seem to be promising factors that could contribute to the explanation of the ambiguous effects of AR in educational settings, their impact on learning within AR environments remains largely unexplored. In this current research, we re-analyzed our three studies on AR-supported physics lab work to shed light on how individual learner differences in spatial and verbal processing influence the effect of AR on cognitive load and learning outcomes across different samples and AR technologies.

1.1. Augmented Reality to Support Learning in Lab Work Settings

AR refers to settings where real-world environments are supplemented with virtual information [6,7]. It thus ranges on a continuum between purely real and purely virtual environments and is assigned to the upper term Mixed Reality [8]. AR can be implemented using different devices such as head-mounted displays (HMDs; e.g., smartglasses) or monitor-based interfaces (e.g., tablet PCs). There is broad meta-analytic evidence showing that AR is a beneficial tool in education that can support learning processes (e.g., [3,9]). Moreover, AR seems to be particularly common and helpful in science-related disciplines [4,10].

Experimentation in lab work settings is considered a fundamental aspect of science education that aims to provide students with opportunities to acquire essential conceptual knowledge and procedural skills [11,12]. Lab work has recently emerged as a particularly promising field for educational AR applications [13]. For instance, Akçayır and Akçayır [14] demonstrated that AR not only enhanced students' laboratory skills in a university physics course but also fostered positive attitudes towards physics labs. Further, AR can facilitate the creation of hybrid multiple external representations (MERs) by spatially and temporally integrating virtual learning representations—such as single measurement values or scientific models—into a real-world experimental setup. This integration is particularly expedient as conventional lab work settings often suffer from a spatial and temporal disconnect within multiple sources of essential information: Additional information typically has to be presented on worksheets or separate displays that are not spatially linked to the experimental observations or components, do not adapt in real time, and are often presented before or after the actual lab work.

There are numerous examples of applications that have made use of AR-based hybrid MERs in lab work settings, considering different target groups and devices. For example, Lauer et al. [15] described an AR application, applicable for both AR tablets and smartglasses, that provides real-time adaptive visualizations of electrical circuit schematics. These virtual visualizations are spatially attached to real circuit components (e.g., resistors and cables) and aim to support elementary school children in mentally linking the components and corresponding schematics. Donhauser et al. [16] presented an AR-supported experimental learning environment on the topic of Lorentz force for secondary education. They used smartglasses to enrich physical observations with virtual visualizations of magnetic fields and vector triads. The studies to be re-analyzed in this current work investigated physics lab work scenarios on electricity education. In all studies, AR was used to spatially integrate time-adaptive measurement values into the physical experimentation environment by means of AR smartglasses [17] or tablets [1,18].

1.1.1. Theoretical Background

Instructional psychology provides several complementary theoretical approaches that can be applied to explain the positive effects of the hybrid MERs created by AR for lab work. The *Cognitive Theory of Multimedia Learning* (CTML; [19]) is grounded in the assumption of the dual coding of information [20] in separate verbal and pictorial working memory channels with limited processing capacity. According to the CTML, learners process information more effectively when it is presented through a combination of both verbal and pictorial representations (multimedia effect). After selecting and organizing the relevant elements of single verbal and pictorial representations, learners have to perform

a complex integration process to mentally link the representations with each other and with prior knowledge. According to the temporal and spatial contiguity principles derived from the CTML [21], presenting representations close to each other in time and space enhances the mental integration processes by relieving the working memory. Thus, these principles underscore the importance of aligning the relevant learning information to promote learning processes.

The *Framework of Coherence Formation* [22,23] draws a distinction between two consecutive mental integration mechanisms needed for the successful processing of MERs: Selection and organization processes to map relevant elements within single representations are subsumed under the term *local coherence formation*. Subsequent *global coherence formation* describes mental mapping processes across mental models of different representations and results in an integrated mental representation. In line with the CTLM, the framework states that the spatial and temporal integration of information facilitates global coherence formation and fosters learning.

Both theories explain the positive effect of information presented in a spatially and temporally integrated manner (e.g., hybrid MERs created through AR) by reducing the cognitive costs of limited working memory resources. Increasing the distance between representations is assumed to increase the working memory load by necessitating learners to retain larger units of information between gazes, which can manifest in fewer gaze shifts and longer information processing intervals [24]. The composition and interplay of the cognitive costs attributed to learning processes is described by *Cognitive Load Theory* (CLT; [25,26]). CLT distinguishes three categories of cognitive load pertinent to learning: Intrinsic cognitive load (ICL) is determined by the level of the task-inherent complexity of information. Germane cognitive load (GCL) is assumed to result from the active processing of learning information. Extraneous cognitive load (ECL) depicts the task-irrelevant processing caused by the design of instructional material and should be minimized to optimize the cognitive capacity available for effective learning. The most common method to measure the three types of cognitive load in a differentiated way is by using retrospective subjective rating scales. In recent years, a variety of cognitive load scales have been developed that have individual focal points and have been shown to cover different aspects of cognitive load (e.g., [27–29]). From a CLT perspective, AR is assumed to reduce ECL in lab work settings by avoiding split attention [30] between experimental phenomena and the corresponding information necessary to understand the underlying mechanisms and concepts. Consequently, AR-enhanced learning setups are expected to result in higher learning gains compared to conventional split-source experimental settings (e.g., [1]). The systematic review of Buchner et al. [2], however, presents mixed findings regarding the influence of AR on cognitive load, indicating that the effect of AR may depend on a variety of moderating factors, the relevance of which may, in turn, depend on the specific learning domain.

One factor that might impact cognitive load in AR-supported learning environments is the AR device used. Against the background of theoretical and practical considerations, Thees et al. [17] discuss the advantages and disadvantages of the different AR devices used in lab work settings. Compared to smartglasses, most teachers and learners are more familiar with screen-based augmentation by means of easily accessible and affordable handheld tablet PCs or smartphone cameras. However, the redundant visualization of the experimental environment by these devices (i.e., students see objects twice: on the tablet display and as a physical object behind the display) could induce additional ECL (see also [31]). In contrast, HMD devices such as smartglasses provide learners with a more immersive experience through a stronger sense of spatial and temporal blending between virtual and real representations. Moreover, they enable hands-free interaction during experimentation. A recent meta-analysis on AR in science learning [4] found no significant influence of the particular AR device used on academic achievement. However, the limited number of included studies using HMD devices considerably limits the conclusions.

1.1.2. Re-Analyzed Studies

The three studies re-analyzed in this paper originate from the interdisciplinary collaborative research project GeAR (“Gelingensbedingungen und Grundsatzfragen von Augmented Reality in experimentellen Lehr-Lernszenarien entlang der schulischen Bildungsbiographie” which translates to “Success factors and fundamental questions of using Augmented Reality in experimental teaching and learning scenarios across the educational pathway”). This project addressed various success factors of AR in student-centered laboratory work through several empirical studies.

Based on the theoretical assumptions described above, we conducted three related studies to investigate how AR can enhance physics lab work scenarios on electricity education. All studies compared an AR-supported condition with a separate display condition. The AR condition created spatial and temporal contiguity by spatially integrating time-adaptive measurement values into the physical experimentation environment by means of AR smartglasses [17] or tablets [1,18]. The separate display setup induced split attention between the real experimental setup and the measured values presented as a matrix on an external tablet display. All studies hypothesized an advantage for the integrated AR presentation format of the measurement values, which was expected to manifest in lower ECL levels and higher conceptual knowledge gains. However, the effects of the integrated AR presentation were mixed across the studies.

The first study under investigation [1] examined a sample of university students and found no differences between the conditions in either of two different ECL scales employed. Nevertheless, in line with the hypotheses, the AR-supported condition showed moderately higher learning gains.

The second study [17] used smartglasses-based AR in an experimental setting for university students, which was very similar to the first study [1]. Contrary to expectations, the separate tablet display condition caused less ECL and was superior in terms of the conceptual knowledge test items closely related to the instruction. However, differences in ECL were only present in one of two employed ECL scales. The results highlight the potential drawbacks of smartglasses and provide evidence for the so-called spatial contiguity failure [32]. The integration of measured values into the experimental environment broke up coherent structures between the measurement values. Because of technical limitations regarding the field of view of smartglasses, subjects in the AR condition were not able to observe all measured values at once. This spatial fragmentation was assumed to impede referential connections and coherence formation within measurement values, which were necessary for conceptual understanding.

The third study [18] adapted the electricity lab work setting for a target group of elementary school children. The AR visualizations were realized by tablets placed in fixed positions. The results showed that AR had no significant impact either on ECL or the conceptual knowledge test performance.

To sum up, the effects of AR across our three studies that investigated different AR devices and samples were mixed and often did not confirm the theoretical assumptions. Following evidence on learning with MERs (e.g., [33]), so far unconsidered individual differences in learners’ prerequisites may have acted as moderators that concealed expected effects and should therefore be addressed as explanatory factors.

1.2. The Impact of Learner Prerequisites in AR-Based Learning Environments

Several meta-analyses have sought to explain the heterogeneous effects of AR studies by means of moderator analyses. However, only few of these included moderators related to learners’ prerequisites. For example, some meta-analyses explored the impact of learners’ individual educational level on the effectiveness of AR. The results are not consistent. While AR was found to be equally beneficial for different levels of education in some meta-analyses [4,34], others suggest that the learners’ educational level can significantly impact the effect of AR on academic success [35] and motivation [36]. These mixed results

may also reflect the varying effects of AR across different learning content. Consequently, some more content-related cognitive learner prerequisites should also be considered.

Throughout AR-supported lab work, learners encounter hybrid MERs consisting of verbal (e.g., the measurement values of current) or pictorial (e.g., experimental components) representations [19]. Against the background of capacity-limited dual channels for processing verbal and pictorial information [19,20], visuospatial abilities and verbal working memory capacity can thus be considered as particularly relevant learner prerequisites [5] for learning with hybrid MERs in lab work settings. Remarkably, despite the increasing prevalence of AR in educational research, these cognitive prerequisites have hardly been investigated in the context of learning with AR-based hybrid MERs [10,37].

1.2.1. The Role of Spatial Ability for Processing and Learning with (Hybrid) MERs

Spatial ability as a component of general cognitive ability is described as the capacity to “generate, retain, retrieve, and transform well-structured visual images” [38] (p. 98). It comprises five sub-factors [39], of which spatial visualization and spatial relations are the most investigated in the context of visuospatially enriched learning environments that require the understanding and manipulation of visual and spatial information. Since the factors of spatial ability are inextricably linked to visuospatial working memory [40,41], visuospatial working memory capacity can also be considered as a spatial ability measure [41,42].

Previous research has shown that spatial ability is positively related to learning outcomes in general [43] and to STEM learning outcomes in particular (e.g., [44]). Moreover, prior research revealed that especially high spatial ability learners benefit from classical multimedia learning settings providing combinations of text and pictures (e.g., in science education; [45]). They are assumed to process integrated verbal and pictorial learning information more effectively and thus to be able to use the resulting extra cognitive resources for the construction of referential connections between representations [5,46,47]. Consequently, high spatial ability learners are expected to profit from more elaborated mental models. In line with this, in a review, Cromley [48] found mean correlations from .20 to .40 for the relationship between spatial skills and different types of multimedia learning outcomes.

Further, the majority of research supports the assumption that spatial ability moderates learning in *spatially enriched* multimedia learning environments providing dynamic animations or 3D visualizations instead of static 2D visualizations. Nevertheless, the evidence supports opposing hypotheses on this moderation [43]: The ability-as-enhancer hypothesis [47] suggests that learners with high spatial ability profit most from spatial visualizations (e.g., [49,50]) as they are equipped with the necessary cognitive resources to manage the additional visuospatial load and successfully construct mental models. Learners with low spatial abilities, however, should experience particularly high cognitive load levels and have difficulties synthesizing the individual representations of MERs into a coherent mental model [51].

In contrast, the ability-as-compensator hypothesis [52] states that learners with high spatial abilities are able to compensate for lacking explicit spatial visualizations, while spatially weak learners struggle to construct a correct mental model themselves and therefore benefit most from explicit external visuospatial support (e.g., [53,54]).

In his meta-analytic review, Höffler [43] found evidence supporting the ability-as-compensator hypothesis. This hypothesis might be especially applicable for rather simple visuospatial enrichments that do not overload low spatial learners' working memory resources but spare their resources through simplifying integrated mental model construction. Kühl et al. [42] attempted to reconcile the opposing hypotheses, proposing that their applicability depends on the ability range of the learners under investigation. They suggested that medium abilities can enhance the benefits of spatial visualizations compared to low-ability learners who are unable to take advantage of spatial information. At the same time, very high abilities can compensate for lacking spatial information, resulting in medium ability learners gaining the most benefit from external spatial enrichment.

Spatiality is considered an important characteristic of AR experiences [6,55]. Virtual visualizations in AR can adapt dynamically to real-world changes and are perceived as spatially integrated into a 3D real world. Although AR learning setups can therefore be counted among spatially enriched learning environments, the moderating effect of spatial abilities on learning outcomes in AR has barely been investigated to date [34,37,56,57]. So far, empirical evidence has shown that AR can be a valuable tool to train spatial ability (e.g., [58–60]) and foster spatial thinking [61], which highlights the interdependency of AR and spatial ability. Moreover, Habig [37] investigated the influence of sex-related visuospatial differences for AR-supported chemistry learning. In their study, males and females performed a test on stereochemistry with half of the tasks using 2D figures to display chemical structures and the other half using tablet-based AR representations incorporating rotatable 3D models of chemical structures. They found that males were more likely to solve AR problems correctly, while females performed better in conventional 2D tasks. Their results support the ability-as-enhancer hypothesis, indicating that the required cognitive processes in the visuospatially enriched AR environment might have been hard to perform for students with lower spatial abilities, such as females. Krüger et al. [57] compared 3D and 2D AR representations of the human heart and found that 3D AR visualizations were superior in promoting GCL and spatial knowledge acquisition. Their study also supports the ability-as-enhancer hypothesis, suggesting that learners with higher spatial abilities profited most from 3D visualizations. Given the limited amount of research and the apparent relationship between AR and spatial abilities, Cheng and Tsai [56], in their review on suggestions for future research on AR in science learning, emphasized the need for investigating the impact of individual differences in mental spatial processing on the learning experience, process, and performance in AR.

1.2.2. The Role of Verbal Working Memory for Processing and Learning with (Hybrid) MERs

Verbal working memory refers to the mental maintenance of verbal (i.e., textual or numerical) information, and its capacity can affect the processing of verbal information. In contrast to findings on spatial abilities, the empirical evidence concerning the role of verbal working memory for learning with MERs is rather sparse. Nevertheless, verbal and numerical working memory capacity is also assumed to affect multimedia learning outcomes ([5,62]; for a review, see [63]). For example, Pazzaglia et al. [64] investigated a hypermedia learning environment for children and found both visuospatial and verbal working memory influencing the processing of multimedia material. In particular, they observed verbal memory capacity to be positively related to the acquisition of semantic knowledge. The results of Plass et al.'s [65] study on language learning with MERs underscore the relevance of visuospatial and verbal abilities, with verbal ability being assumed to be particularly helpful under high-load multimedia conditions. In their study, verbal ability was assessed with a vocabulary speed test, which has been shown to be highly correlated with verbal working memory capacity. Studies using a dual-task methodology further support the presumption that verbal working memory is involved in text processing during multimedia learning [66]. Moreover, prior research demonstrated that learners with low verbal working memory capacity suffered most from split attention between representations but also suggested that integrative instructional design can compensate for working memory deficits [67].

In the context of hybrid MERs in AR-supported lab work environments, the processing of verbal information is of crucial importance. Experimental setups often include verbal instructions, hints, or explanations that guide the learning process. Moreover, measurement values as numerical information that is also processed in the verbal working memory channel play a central role in science experiments. They must be linked to physical phenomena to make them quantifiable, analyzable, and understandable.

1.3. Present Study and Hypotheses

This present research aims to uncover the possible moderating effects of spatial abilities and verbal working memory capacity in three studies on AR-supported lab work environments on electricity education. Re-analyses to detect the potential moderators on the effects of AR support on ECL and conceptual knowledge gains may help to explain and reconcile the ambiguous results. The inconsistency of AR effects across the three studies may be attributed to aptitude-treatment interactions. The studies were collaboratively conceptualized by various disciplines, each integrating their specific research questions. As a result, each study typically investigated not only technological and pedagogical research questions but also methodological issues related to measurements. Consequently, they included multiple measures for assessing learning outcomes, cognitive load, and learner prerequisites.

As ECL was measured with two separate cognitive load scales in Study 1 [1] and Study 2 [17], convergent and discriminant moderator effects, depending on the scale, will be discussed. Further, conclusions will be drawn to expand the theoretical presumptions on learning with hybrid MERs in AR-supported lab work settings by addressing the impact of interindividual differences regarding spatial and verbal processing.

Hypotheses 1a and 1b explore how individual differences in spatial processing influence the effect of AR-induced hybrid MERs on ECL and learning outcomes in a lab work setup. Who benefits most from spatial enrichment is assumed to depend on the requirements of the learning material [43] and learners' spatial ability range [42]. Since integrated virtual measurement values provide a rather simple visuospatial enrichment, we do not expect them to cognitively overload but to relieve the ECL of low spatial learners.

Hypothesis 1a. Spatial ability moderates the effect of the presentation format (AR vs. separate display) on ECL. The instructional reduction of split attention is expected to show particularly strong diminishing effects on ECL in learners with low spatial abilities.

Learning outcomes, such as conceptual knowledge gains, are assumed to be closely related to self-rated cognitive load levels and can also indicate split attention effects. Moreover, they can capture aspects of load that cannot be assessed by retrospective subjective rating scales. Therefore, we additionally investigated the moderating role of spatial abilities for the effect of AR on conceptual knowledge gains.

Hypothesis 1b. Spatial ability moderates the effect of the presentation format on learning outcomes. The instructional reduction of split attention is expected to be particularly beneficial for learning outcomes in learners with low spatial abilities.

Hypotheses 2a and 2b address the moderating effect of verbal working memory capacity. It is assumed that learners with particularly high verbal working memory capacity are able to encode and store larger units of verbal information between glances [24]. However, for learners with low verbal working memory capacity, integrated MERs can facilitate resource-consuming mental search processes that are necessary for coherence formation. This should be reflected in lower ECL levels.

Hypothesis 2a. Verbal working memory capacity moderates the effect of the presentation format on ECL. The instructional reduction in split attention is expected to show particularly strong diminishing effects on ECL in learners with low verbal working memory capacity.

The instructional reduction of split attention effects and the facilitation of mental search processes should also free mental capacities for meaningful learning, which should lead to improved learning outcomes. Further, learning outcomes can reflect additional aspects beyond self-rated cognitive load. Consequently, we also investigated the moderating role of verbal working memory capacity for the effect of AR on conceptual knowledge gains.

Hypothesis 2b. Verbal working memory capacity moderates the effect of the presentation format on learning outcomes. The instructional reduction of split attention is expected to be particularly beneficial for learning outcomes in learners with low verbal working memory capacity.

2. Materials and Methods

2.1. Study Samples

All the three studies to be re-analyzed applied a pretest–posttest design to compare an AR-supported with a separate display lab work setting on electrical circuits regarding ECL levels and learning outcomes in terms of conceptual knowledge gains. In the first study under investigation [1], $N = 50$ (80% female) university students were assigned to a separate display condition or a tablet-based AR environment (age: $M = 25.98$; $SD = 4.67$). While the vast majority of students were accustomed to using smartphones and tablets regularly, most of them reported being unfamiliar with AR. The valid sample of the second study [17] consisted of $N = 107$ engineering students (14% female, age: $M = 19.05$; $SD = 5.20$) working with separate displays or smartglasses-based AR. In the third study examining tablet-based AR for elementary school students [18], $N = 114$ children took part (47% female, age: $M = 9.06$, $SD = 0.87$). In none of the studies did the participants have any domain-specific prior knowledge about electricity that went beyond what they had learned about it at school. Furthermore, in none of the studies were the participants known to have cognitive impairments. However, this cannot be completely ruled out as it was not explicitly asked about. It should further be noted that the samples may not be representative of the respective age groups due to selection bias. University students and children participating in scientific studies are likely to have a higher-than-average level of education. Due to time constraints in the three studies, the sample sizes for certain moderator variables were reduced. Some participants took longer than expected to complete the lab work phase and therefore were unable to complete some of the cognitive tests at the end of the investigation. An overview of the three studies, including the instruments used and corresponding sample sizes, can be found in Table 1.

Table 1. Summary of re-analyzed studies.

Study	Sample	Presentation Format (n)	Spatial Ability Tests (n)	Verbal Working Memory Tests (n)
Study 1 [1]	Mixed university students	Separate tablet display (25)	Paper folding test (25) Card rotation test (25) Visual pattern tests (16)	Visual digit span test (25)
		Handheld tablet AR (25)	Paper folding test (25) Card rotation test (25) Visual pattern test (20)	Visual digit span test (25)
Study 2 [17]	Engineering university students	Separate tablet display (58)	Paper folding test (55) Card rotation test (53) Visual pattern tests (55)	Visual digit span test (40)
		Smartglasses (49)	Paper folding test (39) Card rotation test (38) Visual pattern tests (43)	Visual digit span test (29)
Study 3 [18]	Elementary school students	Separate tablet display (56)	Spatial relations test (54)	Visual digit span test (47)
		Fixed tablet AR (58)	Spatial relations test (58)	Visual digit span test (49)

2.2. Procedures and Measures of Cognitive Load, Conceptual Knowledge, Spatial Ability, and Verbal Working Memory Capacity

The three studies followed the same workflow, which is depicted in Figure 1.

After an introduction to the topic of electricity, the subjects were provided with a pretest on their conceptual understanding of electrical circuits (adapted from [68,69]; see [1,17,18] for exemplary items). The subjects were then randomly assigned to the AR-supported or the separate display condition. Subsequently, the participants familiarized themselves with the technology and were then guided through a series of experiments dealing with serial and parallel electrical circuits. In every experiment, the subjects first built up an electrical circuit. While university students worked with resistors, elementary students used light bulbs instead. For every experiment, voltage was then applied on the circuit and manipulated at the power supply. The subjects observed and compared

measured values of voltage only [18] or of voltage and amperage [1,17] for different circuit components. During the experimentation, the participants answered questions on the relations of the measured values. The virtual measurement values were displayed above the corresponding components for the AR-supported conditions. The separate display condition was provided with a matrix of measurement values on a tablet (see Figure 2). For the study on elementary students [18], gaze data were recorded during lab work. After the lab work phase, cognitive load was assessed via subjective rating scales. Study 1 [1] and Study 2 [17] used adapted versions of Leppink et al.'s [29] scale translated to German, as well as an adapted version of Klepsch et al.'s [27] German scales to assess ICL, ECL, and GCL. The third study [18] provided a cognitive load questionnaire for ICL and ECL particularly tailored to the needs of elementary school children [70]. Afterwards, the students completed the test on conceptual knowledge for a second time, and, depending on the study, some more specific posttests.

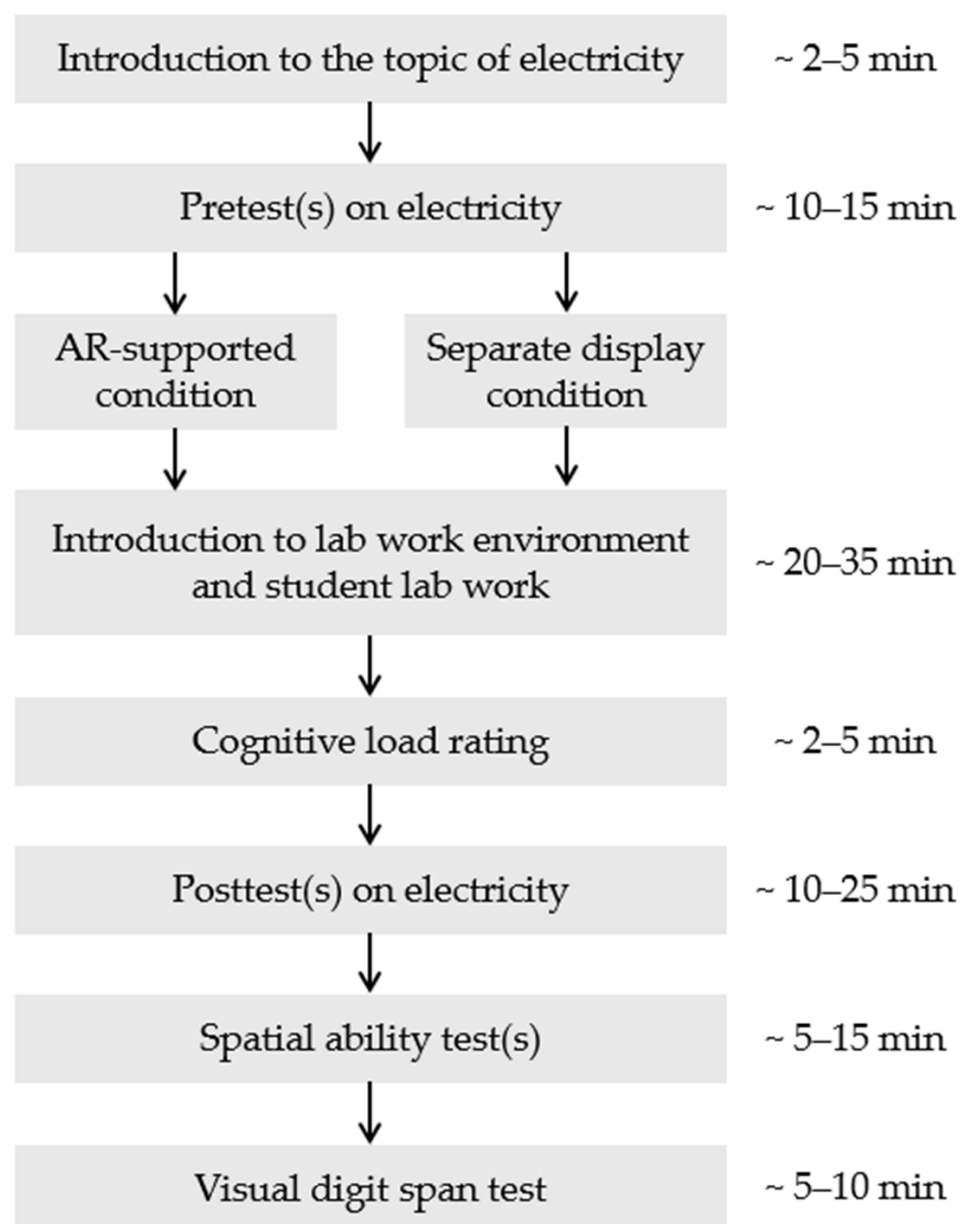


Figure 1. Workflow and time specifications of the three studies to be re-analyzed for moderating variables. Particularly for Study 3 investigating a sample of children, the time required for the single parts of the study varied significantly between the individuals.

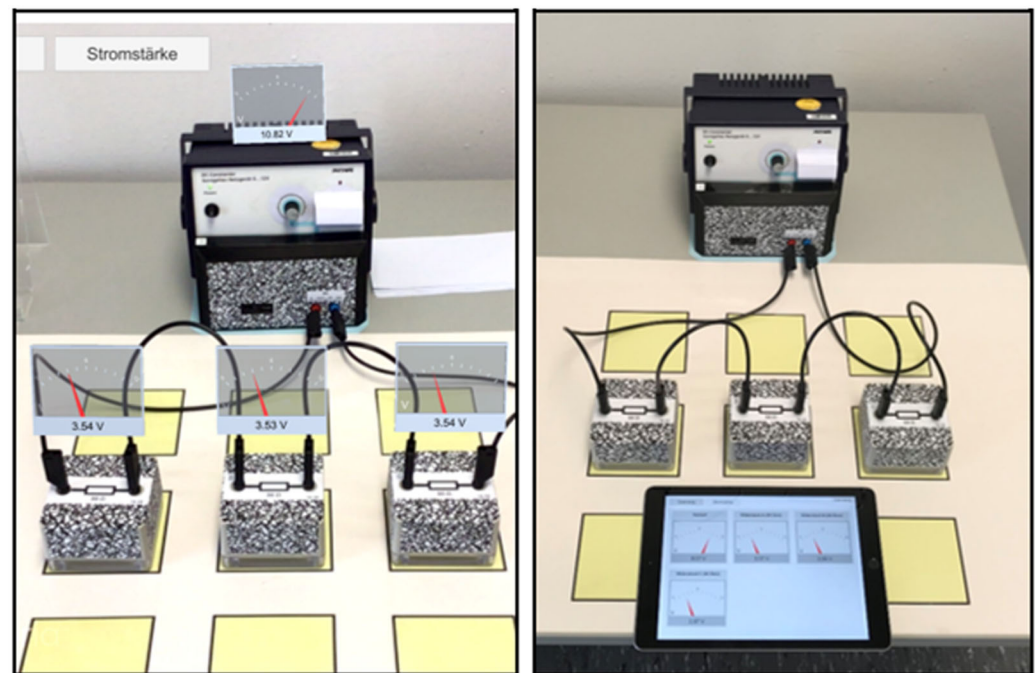


Figure 2. Left: AR-supported learning environment with measurement data virtually integrated into the experimental setup by smartglasses or tablet cameras. Right: separate display learning environment with measurement data provided grouped together on an external tablet.

Having completed the posttests on electricity, three spatial ability tests were performed for the samples of university students. The paper folding test measured the spatial visualization factor and the card rotation test measured the spatial relations factor [71]. Both tests capture the ability to mentally transform objects, e.g., by rotation or inversion [41]. However, the card rotation test is constructed of less complex 2D items and has to be completed faster [40]. The visual pattern test [72] was used to measure the visual component of visuospatial memory in the samples of university students. The test covers the recall performance of two-dimensional patterns arranged in matrices. In the study on elementary school children, spatial ability was assessed by means of the spatial relations subtest of the Primary Mental Abilities Tests Battery by Thurstone and Thurstone [73], which is a common instrument for children of elementary school age. Subsequently, verbal working memory capacity was captured by visual digit span tests. The university students had to recall sequences of four to thirteen digits [71]. The elementary school students completed an adaptive visual digit span test with tasks of increasing difficulty (a visual adaption of the “Repeating Numbers” test of the WISC-V; [74]). Finally, all the participants gave demographic information.

3. Results

3.1. Descriptive Statistics

Table 2 describes the score calculation for dependent and moderator variables. It takes into account the benefit of guessing in speed tests (concerning the paper folding and card rotation tests).

Table 2. Score calculations and maximum scores for dependent and moderator variables.

Variable	Score Calculation	Maximum
Conceptual knowledge gain	Difference between the number of correctly solved tasks in the posttest and the pretest	13 ¹ , 10 ² , 11 ³
ECL	Mean of ECL item responses	7 ^{1,2} , 5 ³

Table 2. *Cont.*

Variable	Score Calculation	Maximum
Paper folding test	Number of correctly solved tasks minus one-fifth the number of incorrect tasks in 3 min	10 ^{1,2}
Card rotation test	Number of correctly solved tasks minus the number of incorrect tasks in 3 min	80 ^{1,2}
Visual pattern test	Mean of the complexity indices of the last three correctly recalled patterns	15 ^{1,2}
Spatial relations test	Number of correctly solved tasks	16 ³
Visual digit span test	Mean of the last three correctly recalled digit sequence lengths	12.33 ^{1,2} , 9.66 ³

¹ Study 1 [1]; ² Study 2 [17]; ³ Study 3 [18].

The descriptive statistics for conceptual knowledge gain, ECL, and moderator variables are displayed in Tables 3 and 4, separately for each study.

Table 3. Means (*M*) and standard deviations (*SD*) of dependent and moderator variables for Study 1 [1] and Study 2 [17], separated by presentation format conditions.

	Study 1 [1]		Study 2 [17]	
	Separate Display	AR	Separate Display	AR
Conceptual knowledge gain	−0.28 (2.37)	0.84 (1.84)	0.09 (1.98)	0.00 (1.62)
ECL (Leppink et al. [29])	1.54 (0.51)	1.31 (0.49)	1.69 (0.46)	1.69 (0.56)
ECL (Klepsch et al. [27])	1.88 (0.94)	1.64 (0.87)	1.88 (0.82)	2.25 (0.99)
Paper folding test	5.82 (2.51)	5.62 (2.66)	5.87 (2.24)	6.14 (2.32)
Card rotation test	55.72 (15.99)	56.20 (13.00)	53.15 (21.06)	57.18 (21.66)
Visual pattern test	10.27 (1.38)	10.49 (1.99)	10.94 (1.64)	10.53 (1.75)
Visual digit span test	6.76 (1.00)	6.80 (1.32)	6.24 (1.49)	6.51 (0.84)

Table 4. Means (*M*) and standard deviations (*SD*) of dependent and moderator variables, for Study 3 [18], separated by presentation format conditions.

	Separate Display	AR
Conceptual knowledge gain	4.14 (2.56)	3.78 (2.36)
ECL	1.48 (0.50)	1.51 (0.53)
Spatial relations test	13.41 (1.85)	13.72 (1.83)
Visual digit span test	4.19 (0.67)	4.01 (0.78)

3.2. Assumptions and Preliminary Considerations for the Following Analyses

To re-analyze our three studies with the aim of identifying explanatory influences on the effects of AR, the PROCESS macro by Hayes [75] was used to perform several moderator analyses. Alongside the standard regression analyses yielding *p*-values, 5000 bootstraps were calculated for each analysis to generate and test for bias-corrected and accelerated 90% bootstrap confidence intervals (BCa 90% CI; one-tailed). A heteroscedasticity consistent standard error and covariance matrix estimator was selected (HC3, Davidson-McKinnon). Bootstrapping is a robust approach that generally makes no assumptions in terms of distribution characteristics, which is particularly relevant for rather small samples. It is still crucial to acknowledge that the limited sample sizes might diminish the statistical power of the analyses. Therefore, the following results must be interpreted with caution and the understanding that they only suggest potential trends.

3.3. Hypothesis 1a: The Moderating Role of Spatial Ability for the Effect of Presentation Format on ECL

Hypothesis 1 concerns the moderator variables of spatial ability. With regard to Hypothesis 1a, performance in the paper folding test acted as a marginally significant

moderator (although not indicated by the BCa 90% CI) and performance in the card rotation test was a significant moderator (see Figure 3) for the effect of presentation format on ECL measured with Klepsch et al.'s [27] scale in Study 1 [1] (see Table 5). Subjects with low spatial abilities showed less ECL in the AR-supported condition compared to the separate display presentation format.

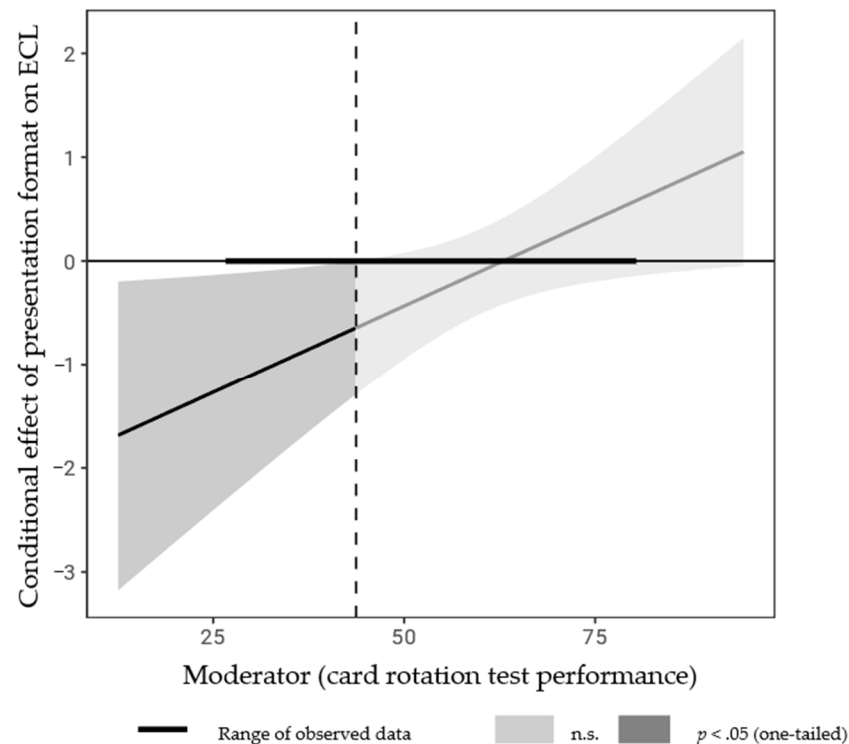


Figure 3. Johnson–Neyman plot illustrating the significant interaction between the presentation format (AR was coded as 1 and the separate display condition was coded as 0) and the moderator card rotation test performance for the dependent variable ECL [27] in Study 1.

Table 5. Moderator analyses for the effect of presentation format on ECL measured by adapted versions of Leppink et al.'s [29] and Klepsch et al.'s [27] scales in Study 1 [1].

Moderator	ECL (Leppink et al. [29])					ECL (Klepsch et al. [27])				
	<i>df</i>	<i>F</i>	<i>p</i>	BCa 90% CI	ΔR^2	<i>df</i>	<i>F</i>	<i>p</i>	BCa 90% CI	ΔR^2
Paper folding test	1/46	0.16	.344	[−0.071; 0.116]	.003	1/46	2.47	.061 [†]	[−0.012; 0.007]	.035
Card rotation test	1/46	0.01	.462	[−0.016; 0.018]	.000	1/46	4.49	.034 [*]	[0.003; 0.063]	.068
Visual pattern test	1/36	0.14	.358	[−0.263; 0.169]	.004	1/36	0.13	.358	[−0.260; 0.404]	.003
Visual digit span test	1/46	0.93	.169	[−0.386; 0.104]	.024	1/46	0.12	.365	[−0.079; 0.055]	.003

[†] $p < .10$, and ^{*} $p < .05$.

Neither in Study 2 [17] (see Table 6) nor in Study 3 [18] (see Table 7) were the spatial ability test performances identified as significant moderators with regard to the effect of the presentation format on ECL.

Table 6. Moderator analyses for the effect of presentation format on ECL, measured by adapted versions of Leppink et al.'s [29] and Klepsch et al.'s [27] scales in Study 2 [17].

Moderator	ECL (Leppink et al. [29])					ECL (Klepsch et al. [27])				
	<i>df</i>	<i>F</i>	<i>p</i>	BCa 90% CI	ΔR^2	<i>df</i>	<i>F</i>	<i>p</i>	BCa 90% CI	ΔR^2
Paper folding test	1/90	0.46	.249	[−0.002; 0.002]	.004	1/90	0.97	.163	[−0.217; 0.055]	.011
Card rotation test	1/87	0.49	.245	[−0.013; 0.005]	.007	1/87	0.91	.171	[−0.026; 0.007]	.013
Visual pattern test	1/94	0.04	.422	[−0.097; 0.123]	.001	1/94	0.00	.479	[−0.218; 0.205]	.000
Visual digit span test	1/65	0.36	.276	[−0.336; 0.159]	.007	1/65	3.29	.037 *	[−0.726; −0.030]	.045

* $p < .05$.**Table 7.** Moderator analyses for the effect of presentation format on ECL measured by an adapted version of Altmeyer et al.'s [70] scale for elementary school children in Study 3 [18].

Moderator	<i>df</i>	<i>F</i>	<i>p</i>	BCa 90% CI	ΔR^2
Spatial relations test	1/108	0.08	.392	[−0.075; 0.105]	.001
Visual digit span test	1/92	1.70	.098 †	[−0.055; 0.457]	.020

† $p < .10$.

3.4. Hypothesis 1b: The Moderating Role of Spatial Ability for the Effect of Presentation Format on Conceptual Knowledge Gain

The results for Hypothesis 1b showed no significant moderation effect of spatial ability for the effect of presentation format on conceptual knowledge gain for all three studies. Tables 8–10 illustrate the statistics on the moderator effects.

Table 8. Moderator analyses for the effect of presentation format on conceptual knowledge gain for Study 1 [1].

Moderator	<i>df</i>	<i>F</i>	<i>p</i>	BCa 90% CI	ΔR^2
Paper folding test	1/46	1.18	.141	[−0.498; 0.107]	.013
Card rotation test	1/46	0.30	.293	[−0.045; 0.090]	.005
Visual pattern test	1/36	0.08	.388	[−0.467; 0.256]	.003
Visual digit span test	1/46	0.01	.462	[−1.209; 1.078]	.000

Table 9. Moderator analyses for the effect of presentation format on conceptual knowledge gain for Study 2 [17].

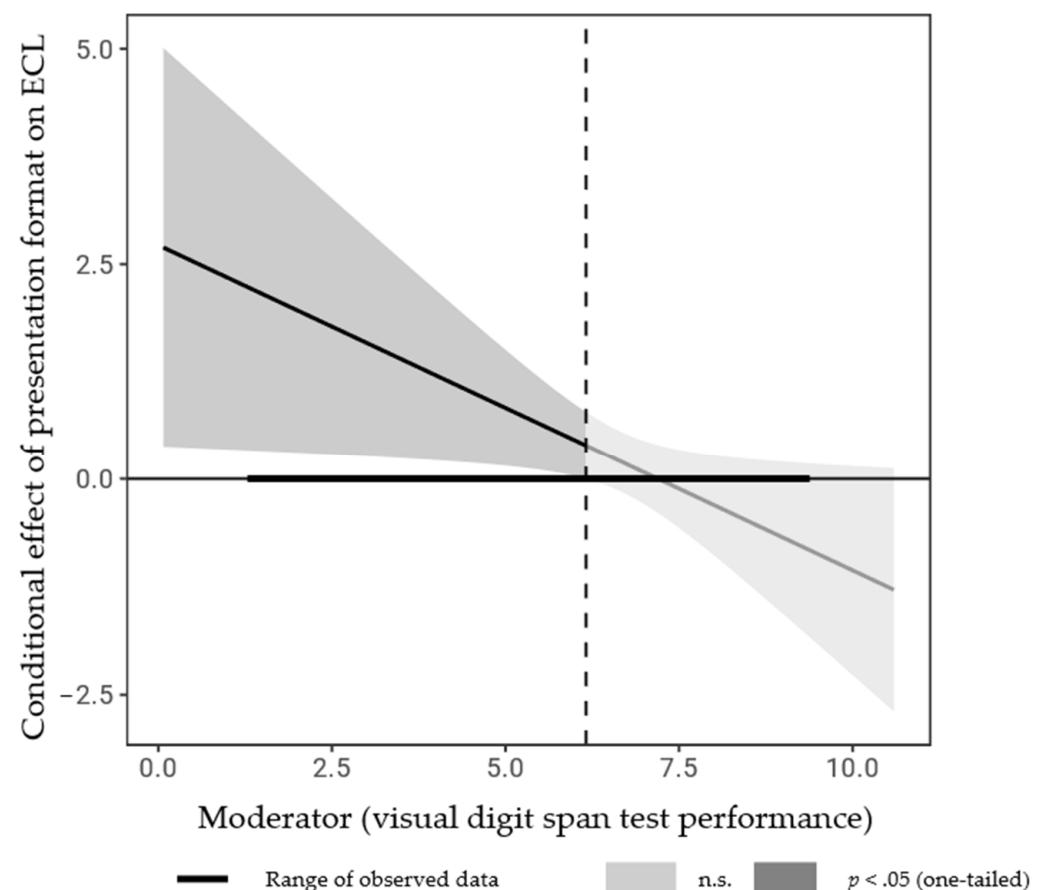
Moderator	<i>df</i>	<i>F</i>	<i>p</i>	BCa 90% CI	ΔR^2
Paper folding test	1/90	0.17	.342	[−0.200; 0.330]	.002
Card rotation test	1/87	0.19	.331	[−0.033; 0.019]	.002
Visual pattern test	1/94	1.13	.146	[−0.613; 0.135]	.012
Visual digit span test	1/65	0.60	.221	[−0.489; 1.331]	.013

Table 10. Moderator analyses for the effect of presentation format on conceptual knowledge gain for data of Study 3 [18].

Moderator	<i>df</i>	<i>F</i>	<i>p</i>	BCa 90% CI	ΔR^2
Spatial relations test	1/108	0.00	.492	[−0.410; 0.400]	.000
Visual digit span test	1/92	0.44	.254	[−1.667; 0.714]	.005

3.5. Hypothesis 2a: The Moderating Role of Verbal Working Memory Capacity for the Effect of Presentation Format on ECL

Hypothesis 2 investigates the role of verbal working memory. The moderator analyses for Hypothesis 2a revealed no significant moderation effect of verbal working memory in Study 1 [1] (see Table 5). In Study 3 [18], there was a marginally significant moderator (which was not indicated by the BCa 90% CI), indicating that children scoring high in verbal working memory report less ECL in the separate display condition (see Table 7). In Study 2 [17], performance in the digit span test acted as a significant moderator for the effect of the presentation format on ECL, measured with Klepsch et al.'s [27] scale (see Table 6). In this study, subjects scoring low in verbal working memory reported less ECL for the separate display compared to the smartglasses AR-supported experimentation environment (see Figure 4).

**Figure 4.** Johnson–Neyman plot illustrating the significant interaction between the presentation format (AR was coded as 1 and the separate display condition was coded as 0) and the moderator visual digit span test performance for the dependent variable ECL [27] in Study 2.

3.6. Hypothesis 2b: The Moderating Role of Verbal Working Memory Capacity for the Effect of Presentation Format on Conceptual Knowledge Gain

Concerning Hypothesis 3b, verbal working memory was no significant moderator for the effect of presentation format on conceptual knowledge gains in any study (see Tables 8–10).

3.7. Exploratory Results

The relationships between dependent variables and moderators across the participants of both conditions (separate display and AR-supported) were investigated in an explorative manner for each study (see Tables 11–13).

Table 11. Bivariate correlations (two-tailed) between variables of Study 1 [1].

	CKG		ECL [29]		ECL [27]		PFT		CRT		VPT	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Conceptual knowledge gain (CKG)												
ECL (Leppink et al. [29])	−.10	.510										
ECL (Klepsch et al. [27])	−.06	.705	.48	<.001 **								
Paper folding test (PFT)	.16	.261	.09	.543	−.13	.368						
Card rotation test (CRT)	−.07	.656	−.10	.486	−.23	.113	.25	.085 †				
Visual pattern test (VPT)	.25	.114	−.14	.395	−.32	.045 *	.43	.005 *	.13	.434		
Visual digit span test	.04	.777	−.11	.464	−.33	.018 *	−.02	.882	.16	.274	.32	.047 *

† $p < .10$; * $p < .05$; and ** $p < .01$.

Table 12. Bivariate correlations (two-tailed) between variables of Study 2 [17].

	CKG		ECL [29]		ECL [27]		PFT		CRT		VPT	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Conceptual knowledge gain (CKG)												
ECL (Leppink et al. [29])	−.11	.260										
ECL (Klepsch et al. [27])	−.05	.592	.55	<.001 **								
Paper folding test (PFT)	.17	.105	.10	.320	−.04	.673						
Card rotation test (CRT)	.02	.880	−.07	.504	−.16	.133	.33	.001 **				
Visual pattern test (VPT)	.21	.034 *	−.10	.351	−.20	.045 *	.27	.009 **	−.00	.982		
Visual digit span test	.04	.749	.07	.551	−.09	.472	.17	.172	.22	.066 †	.16	.197

† $p < .10$; * $p < .05$; and ** $p < .01$.

Table 13. Bivariate correlations (two-tailed) between variables of Study 3 [18].

	CKG		ECL		SRT	
	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>
Conceptual knowledge gain (CKG)						
ECL	−.05	.592				
Spatial relations test (SRT)	.01	.907	−.09	.350		
Visual digit span test	.21	.039 *	−.28	.006 **	.16	.132

* $p < .05$; and ** $p < .01$.

For Study 3 [18], correlations between learner prerequisites and mean number of transitions between different representations that were captured by eye tracking during lab work were exploratively investigated. The statistics are displayed in Table 14.

Table 14. Bivariate correlations (two-tailed) between eye tracking variables and moderator variables of Study 3 [18].

		Spatial Relations Test		Visual Digit Span Test	
		<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Mean number of transitions within single circuits	Circuit ↔ tablet	.03	.852	.19	.229
	Workbook ↔ tablet	−.29	.046 *	−.18	.256
	Workbook ↔ circuit	−.24	.094 †	−.03	.870
Mean number of transitions between two circuits	Tablet ↔ tablet	.30	.037 *	.06	.692
	Circuit ↔ circuit	−.12	.428	−.01	.956

† $p < .10$; and * $p < .05$.

4. Discussion

This present research re-analyzed three recent studies to investigate the moderating role of learners' spatial ability and verbal working memory capacity for the effect of AR support on ECL and conceptual knowledge gains in lab work environments. The three related studies under investigation shared similar lab work setups and workflows but differed in terms of the sample characteristics and AR devices used. While the expected moderators could not be confirmed across all studies, the re-analyses revealed that tablet-based AR reduced ECL compared to a separate display condition among university students with particularly low spatial abilities [1]. In contrast, smartglasses-based AR induced more ECL than the separate display condition in university students with low verbal memory capacity [17]. No significant moderators were found for the effect of AR support on conceptual knowledge. In summary, the findings of the re-analyses point out that the mental processing benefits gained from AR support during lab work may depend on learners' individual cognitive prerequisites. The varying relevance of spatial and verbal learner prerequisites across the three studies underscores the interdependence of these learner-centered moderator variables with further contextual factors, such as the particular AR device used and age of the target group.

4.1. The Moderating Role of Spatial Ability

Hypothesis 1a could be confirmed for Study 1 [1] as the performance in the paper folding and card rotation spatial ability tests acted as (marginally) significant moderators for the effect of presentation format on self-rated ECL. In a sample of university students, learners with low spatial abilities reported less ECL (measured with Klepsch et al.'s [27] scale) in the tablet-based AR condition than in the separate display condition. While the original research only showed significant differences between the conditions in conceptual knowledge acquisition, the moderator analysis thus suggests that AR may also impact ECL in certain subjects. The results are in line with the ability-as-compensator hypothesis [52,53], suggesting that the spatial enrichment through AR relieved the working memory of spatially weak learners. While prior studies on spatial enhancement in or through AR [37,57] found evidence for the ability-as-enhancer hypothesis, the current support for the ability-as-compensator hypothesis may be explained by the rather simple visuospatially integrated measurement values that did not overload low spatial learners' working memory resources but supported their mental integration processes [22]. The ability range of the participants further contributes to the manifestation of the ability-as-compensator hypothesis [42]. Given the selection bias in the sample of university students, even those learners considered as spatially weaker may still possess sufficient visuospatial abilities to benefit from the external spatial enrichment provided. Tablet-based AR may have served as a "cognitive prosthetic" [76] for low spatial learners. The spatial and temporal contiguity [21] created by AR probably supported low spatial learners in the formation of mental interrelations between the experimental setup and measurement data, thereby

reducing ECL. In the other two studies, spatial ability was not identified as a significant moderator. With regard to Study 2 [17] on smartglasses AR-supported lab work, the spatial contiguity failure [32] caused by the technically limited field of view of smartglasses AR could have offset the positive effects of AR, regardless of the individual spatial prerequisites. The ECL of elementary school students using tablet-based AR in Study 3 [18] was also not significantly affected by spatial learner prerequisites. The lack of significant effects in Study 3 [18] could be due to the fact that individual spatial abilities are still developing at elementary school age and may therefore have less explanatory power in this age group.

Contrary to Hypothesis 1b, spatial ability did not significantly influence the effect of the presentation format on conceptual knowledge gains in any study. Although prior research shows that ECL and learning outcomes are related, ECL seemed to be the more sensitive for spatial moderator effects in Study 1 [1].

4.2. The Moderating Role of Verbal Working Memory Capacity

With regard to Hypothesis 2a, the re-analysis of Study 2 [17] implies that, for low verbal working memory learners, using smartglasses AR induced more ECL than using a separate display to observe measurement values. This is in line with Thees et al.'s [17] assumption that the limited field of view of smartglasses broke up coherent structures between measurement values and thereby disrupted coherence formation processes [32]. Thus, the referential connections between measured values augmented by means of smartglasses required a lot of working memory resources. While high verbal working memory learners seemed to be able to keep larger units of measurement digits in their working memory and could thereby compensate for the spatial fragmentation, particularly high ECL was caused in learners with low verbal working memory capacity. The current re-analysis confirms the findings of Study 2 [17] that smartglasses AR can indeed impede certain cognitive matching processes. However, it extends and refines these conclusions by taking into account the dependence of the effects on individual learner prerequisites: The segmentation of measurement values by smartglasses may not be generally obstructive but seems to specifically challenge learners with lower working memory capacity. In contrast, the tablet-based AR used in Study 1 [1] and Study 3 [18] allowed students to view all measurement values simultaneously. This was also the case for the separate display conditions that saw the measurement values grouped together in a grid. Thus, memorizing and comparing measurement values was equally demanding in both conditions. This might explain why verbal working memory capacity showed no significant moderating effect in any of the tablet-based AR studies. Nevertheless, a marginal significant moderating effect for Study 3 [18] (though not reflected in the 90% BCa CI and requiring cautious interpretation) suggests that, for learners with high verbal working memory capacity, AR might have induced higher levels of ECL compared to the separate display condition. According to the ability-as-compensator hypothesis, high working memory learners may not benefit as much from AR, since they can manage split-source material without integrative AR support. However, AR can pose additional technological challenges, particularly for younger participants. For instance, adjusting the tablet to maintain the visibility of the markers and keeping the measurement values in view can be difficult. As a result, AR may have acted as a distraction rather than a support for children. Especially since children's inhibitory control is still developing, AR could have made it harder for them to focus on the learning task. These distractions likely contributed to higher ECL, making the separate display format more effective in reducing extraneous processing for children with sufficient verbal working memory capacity to manage split attention between information sources.

Similar to the results on spatial ability, the re-analyses did not find verbal working memory capacity to significantly moderate the effect of presentation format on conceptual knowledge gains in any study (Hypothesis 2b). ECL, as a measure closer to actual learning processes, seems to be a more sensible dependent variable for moderator effects of cognitive learner prerequisites than conceptual knowledge gains in Study 2 [17].

4.3. Further Findings and Considerations

The explorative results on the relationship between potential moderators and dependent variables across conditions revealed that spatial ability tests are related but nevertheless capture different aspects of spatial ability [40]. For every study, the performance in the digit span test was negatively related to ECL, which confirms that the ECL measurements are related to working memory load [25]. For Study 1 [1], the exploratory results additionally showed a negative relationship between visuospatial working memory and ECL. Conceptual knowledge gain was positively related to visuospatial working memory in Study 2 [17] and digit span performance in Study 3 [18]. This points to the assumed link between cognitive prerequisites and learning outcomes [43,63] and also indicates that both may be influenced by a general intelligence factor.

Study 1 [1] and Study 2 [17] both used two different scales to assess the dimensions of cognitive load. Remarkably, although the ECL measures showed highly significant correlations in both studies, the moderation effects and meaningful correlations only manifested in Klepsch et al.'s [27] scale. This clearly highlights the fact that cognitive load questionnaires vary in their assessed load aspects [28] and that some scales seem to be more sensitive to specific group differences and moderator effects than others. Future studies should therefore choose the load scale carefully, match it precisely to the objectives of the study, and combine different cognitive load measurement approaches to examine convergent validity aspects.

The exploratory results on transitions between representations captured in Study 3 [18] indicate that subjects with high spatial abilities needed to perform fewer transitions between workbook and tablet, as well as between workbook and circuit, within single circuits. In contrast, the higher the spatial ability performance, the more transitions between the tablets of two experimental setups were performed. These results underscore that spatial ability clearly influences the processing of MERs in lab work setups. To deepen our understanding of how individual learner differences impact information processing, promising process measures like gaze behavior should be further investigated.

4.4. Limitations and Future Research

The first limitation of the present re-analysis concerns the sample sizes and compositions of the samples. Regression-based approaches require large samples to ensure adequate statistical power. Consequently, as with many behavioral regression studies, the samples of the present re-analyses must be considered statistically underpowered [77,78]. The findings should therefore be interpreted as indicative of possible trends rather than as definitive conclusions. However, despite this limitation, this study provides valuable insights into the role of cognitive learner prerequisites for the effectiveness of AR support in lab work environments, particularly given the paucity of research in this area. The present results can further be included in future meta-regressions to achieve sufficient power and precision [77,79] and can stimulate future research.

Further, the homogeneity of the samples consisting of university students and selected children regarding general cognitive ability and thus spatial and verbal competencies might have impeded the investigation of the moderator variables. Kühl et al. [42] suggest that the specific manifestation of spatial ability as a moderator variable depends on the ability range of the learners under investigation. Against this background, it is all the more important to examine the full and representative range of cognitive abilities to draw reliable conclusions about their influences.

Moreover, the gender distribution in the re-analyzed studies was uneven, particularly among university students. While the sample of Study 1 [1] mainly consisted of female subjects (80%), men predominated in Study 2 [17] (86%). Research has shown that gender is related to spatial ability and even impacts multimedia outcomes when results are controlled for spatial ability [80]. Consequently, the differences between the studies cannot be related exclusively to differences in experimental design (e.g., the use of smartglasses vs. tablet-based AR), but may also reflect the impact of imbalances in sample distribution. Habig [37]

has identified sex differences in AR-supported chemistry learning tasks, theoretically attributing these disparities to variations in visuospatial abilities between genders. Future research should investigate and align both cognitive prerequisites and gender to better understand their interrelated influences.

A further limitation is that the learning tasks may have been too easy for the target groups, potentially reducing the effects of AR and masking the interaction effects. This limitation might explain why the cognitive variables did not moderate the effect of the presentation format on conceptual knowledge gains. Future research should explore the moderators of learning outcomes in AR-supported environments using diverse samples and more challenging tasks across various domains. This approach would also enhance the generalizability of the findings.

The retrospective self-assessment of cognitive load depicts another drawback of this present research. Retrospective cognitive load self-assessment is biased by a lack of objectivity and direct temporal connection to the learning process [70]. In future studies, objective, physiological, or behavioral online measurements (e.g., smartpen measures, [70]) could additionally be considered and combined with traditional questionnaires to achieve more reliable, multimodal cognitive load measurements. Moreover, the studies showed floor effects in cognitive load ratings. This constraint of variance may have covered potential moderation effects. Another cognitive load-related limitation is described in Schroeder and Cenkci's [81] review, which revealed missing evidence for differences in ECL between integrated and spatially distant designs. They argued that the positive effects of integrated learning environments might rather result from a beneficial allocation of germane resources. Consequently, future studies should also examine the moderating role of learner prerequisites for the effect of AR on GCL.

An additional constraint is that cognitive prerequisites were assessed after the intervention for all studies. Since the use of AR applications has been shown to have positive effects on a user's spatial ability [59–61], it is possible that the AR interventions in the re-analyzed studies may have influenced subjects' spatial abilities. However, this can be considered a minor constraint since the AR learning environments of the present studies did not meet the conditions of an AR-based spatial ability training (e.g., the possibility to interact with three-dimensional objects to perceive their spatial relations [58]). Nonetheless, future research on spatial ability moderators should take into account this potential interdependence between AR interventions and spatial ability measurements.

In addition to cognitive prerequisites, future research should also focus on affective learner characteristics, which have been found to be closely linked not only to AR environments [14,82] but also cognitive load and learning outcomes [83].

4.5. Practical Implications and Conclusions

This study is among the first to investigate how cognitive learner prerequisites impact the effectiveness of AR support in lab work learning environments. By re-analyzing the data from three studies involving different target groups and AR devices, it extends and refines the findings of the original research and helps to reconcile the ambiguous results. The findings suggest that AR can effectively support learners with low spatial abilities by relieving their working memory load. Moreover, learners with low verbal working memory appear to be particularly affected by the spatial contiguity failure that can be caused by smartglasses AR. However, AR smartglasses offer several advantages, especially for hands-free experimentation and a more immersive experience. As technology develops, their current limitations, such as the severely limited field of view, are likely to be addressed. These improvements could allow smartglasses to realize their full potential and increase their effectiveness in learning environments. From a practical perspective, the design and implementation of AR-supported learning environments should carefully consider the interplay of learning objectives, individual cognitive strengths and weaknesses of the target learners, and contextual factors such as accessible AR devices to create the most effective learning experiences.

Author Contributions: Conceptualization, K.A., R.B., J.K. and S.M.; methodology, K.A., R.B. and S.M.; formal analysis, K.A.; investigation, K.A.; resources, R.B. and J.K.; data curation, K.A. and S.M.; writing—original draft preparation, K.A. and S.M.; writing—review and editing, S.M., R.B. and J.K.; supervision, R.B. and S.M.; project administration, R.B., J.K. and S.M.; funding acquisition, J.K. and R.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Federal Ministry of Education and Research, Germany (BMBF; Project GeAR, Grant number: 01JD1811) and supported by LMUexcellent, funded by the Federal Ministry of Education and Research (BMBF) and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the Länder.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Faculty of Empirical Human Sciences and Economics at Saarland University on 13 October 2021 for studies involving humans.

Informed Consent Statement: Informed consent was obtained from all the subjects involved in this study.

Data Availability Statement: This article involves the re-analysis of data from three previously described studies; no new data were generated. Details of the original studies are provided in [1,17,18]. For further information or access to the data used in this re-analysis, please contact the corresponding authors. Data will be made available upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Altmeyer, K.; Kapp, S.; Thees, M.; Malone, S.; Kuhn, J.; Brünken, R. The Use of Augmented Reality to Foster Conceptual Knowledge Acquisition in STEM Laboratory Courses—Theoretical Background and Empirical Results. *Br. J. Educ. Technol.* **2020**, *51*, 611–628. [CrossRef]
2. Buchner, J.; Buntins, K.; Kerres, M. The Impact of Augmented Reality on Cognitive Load and Performance: A Systematic Review. *J. Comput. Assist. Learn.* **2022**, *38*, 285–303. [CrossRef]
3. Garzón, J.; Acevedo, J. Meta-Analysis of the Impact of Augmented Reality on Students' Learning Gains. *Educ. Res. Rev.* **2019**, *27*, 244–260. [CrossRef]
4. Xu, W.W.; Su, C.Y.; Hu, Y.; Chen, C.H. Exploring the Effectiveness and Moderators of Augmented Reality on Science Learning: A Meta-Analysis. *J. Sci. Educ. Technol.* **2022**, *31*, 621–637. [CrossRef]
5. Gyselinck, V.; Cornoldi, C.; Dubois, V.; De Beni, R.; Ehrlich, M.F. Visuospatial Memory and Phonological Loop in Learning from Multimedia. *Appl. Cogn. Psychol. Off. J. Soc. Appl. Res. Mem. Cogn.* **2002**, *16*, 665–685. [CrossRef]
6. Azuma, R.T. A Survey of Augmented Reality. *Presence-Virtual Augment. Real.* **1997**, *6*, 355–385. [CrossRef]
7. Radu, J. Augmented Reality in Education: A Meta-Review and Cross-Media Analysis. *Pers. Ubiquitous Comput.* **2014**, *18*, 1533–1543. [CrossRef]
8. Milgram, P.; Kishino, F. A Taxonomy of Mixed Reality Visual-Displays. *IEICE Trans. Inf. Syst.* **1994**, *E77D*, 1321–1329.
9. Malone, S.; Garzón, J.; Kuhn, J. Three Decades of Augmented Reality in Education: A Second-Order Meta-Analysis and Research Synthesis. *OSF Prepr.* **2023**. [CrossRef]
10. Ibáñez, M.B.; Delgado-Kloos, C. Augmented Reality for STEM Learning: A Systematic Review. *Comput. Educ.* **2018**, *123*, 109–123. [CrossRef]
11. Hofstein, A.; Lunetta, V.N. The Laboratory in Science Education: Foundations for the Twenty-First Century. *Sci. Educ.* **2004**, *88*, 28–54. [CrossRef]
12. Bybee, R.W. Scientific Inquiry And Science Teaching. In *Scientific Inquiry and Nature of Science*; Flick, L.B., Lederman, N.G., Eds.; Science & Technology Education Library; Springer: Dordrecht, The Netherlands, 2004; Volume 25, pp. 1–14. ISBN 978-1-4020-2671-3.
13. Jiang, S.Y.; Tatar, C.; Huang, X.D.; Sung, S.H.; Xie, C. Augmented Reality in Science Laboratories: Investigating High School Students' Navigation Patterns and Their Effects on Learning Performance. *J. Educ. Comput. Res.* **2022**, *60*, 777–803. [CrossRef]
14. Akçayır, M.; Akçayır, G. Advantages and Challenges Associated with Augmented Reality for Education: A Systematic Review of the Literature. *Educ. Res. Rev.* **2017**, *20*, 1–11. [CrossRef]
15. Lauer, L.; Peschel, M.; Malone, S.; Altmeyer, K.; Brunken, R.; Javaheri, H.; Amiraslanov, O.; Grünerbl, A.; Lukowicz, P. Real-Time Visualization of Electrical Circuit Schematics: An Augmented Reality Experiment Setup to Foster Representational Knowledge in Introductory Physics Education. *Phys. Teach.* **2020**, *58*, 518–519. [CrossRef]
16. Donhauser, A.; Küchemann, S.; Kuhn, J.; Rau, M.; Malone, S.; Edelsbrunner, P.; Lichtenberger, A. Making the Invisible Visible: Visualization of the Connection Between Magnetic Field, Electric Current, and Lorentz Force with the Help of Augmented Reality. In *Smartphones as Mobile Minilabs in Physics*; Kuhn, J., Vogt, P., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 373–377. ISBN 978-3-030-94043-0.

17. Thees, M.; Altmeyer, K.; Kapp, S.; Rexig, E.; Beil, F.; Klein, P.; Malone, S.; Brünken, R.; Kuhn, J. Augmented Reality for Presenting Real-Time Data During Students' Laboratory Work: Comparing a Head-Mounted Display With a Separate Display. *Front. Psychol.* **2022**, *13*, 804742. [CrossRef]
18. Altmeyer, K.; Malone, S.; Barz, M.; Lauer, L.; Kuhn, J.; Brünken, R. Coherence Processes in Augmented Reality Lab Work. 2024; *in preparation*.
19. Mayer, R.E. Cognitive Theory of Multimedia Learning. In *The Cambridge Handbook of Multimedia Learning*; Mayer, R.E., Fiorella, L., Eds.; Cambridge University Press: Cambridge, UK, 2021; pp. 57–72. ISBN 978-1-108-89433-3.
20. Paivio, A. *Mental Representations: A Dual Coding Approach*; Oxford University Press: Oxford, UK, 1990.
21. Fiorella, L.; Mayer, R.E. Principles for Reducing Extraneous Processing in Multimedia Learning: Coherence, Signaling, Redundancy, Spatial Contiguity, and Temporal Contiguity Principles. In *The Cambridge Handbook of Multimedia Learning*; Mayer, R.E., Fiorella, L., Eds.; Cambridge University Press: Cambridge, UK, 2021; pp. 185–198. ISBN 978-1-108-89433-3.
22. Seufert, T.; Brünken, R. Supporting Coherence Formation in Multimedia Learning. In *Instructional Design for Effective and Enjoyable Computer-Supported Learning, Proceedings of the First Joint Meeting of the EARLI SIGs Instructional Design and Learning and Instruction with Computers [CD-ROM]*; Gerjets, P., Kirschner, P.A., Elen, J., Joiner, R., Eds.; Knowledge Media Research Center: Tübingen, Germany, 2004; pp. 138–147.
23. Seufert, T.; Brünken, R. Cognitive Load and the Format of Instructional Aids for Coherence Formation. *Appl. Cogn. Psychol.* **2006**, *20*, 321–331. [CrossRef]
24. Bauhoff, V.; Huff, M.; Schwan, S. Distance Matters: Spatial Contiguity Effects as Trade-Off between Gaze Switches and Memory Load. *Appl. Cogn. Psychol.* **2012**, *26*, 863–871. [CrossRef]
25. Sweller, J.; van Merriënboer, J.J.; Paas, F.G.W.C. Cognitive Architecture and Instructional Design. *Educ. Psychol. Rev.* **1998**, *10*, 251–296. [CrossRef]
26. Sweller, J.; van Merriënboer, J.J.G.; Paas, F. Cognitive Architecture and Instructional Design: 20 Years Later. *Educ. Psychol. Rev.* **2019**, *31*, 261–292. [CrossRef]
27. Klepsch, M.; Schmitz, F.; Seufert, T. Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. *Front. Psychol.* **2017**, *8*, 1997. [CrossRef] [PubMed]
28. Thees, M.; Kapp, S.; Altmeyer, K.; Malone, S.; Brünken, R.; Kuhn, J. Comparing Two Subjective Rating Scales Assessing Cognitive Load During Technology-Enhanced STEM Laboratory Courses. *Front. Educ.* **2021**, *6*, 705551. [CrossRef]
29. Leppink, J.; Paas, F.; Van Der Vleuten, C.P.M.; Van Gog, T.; Van Merriënboer, J.J.G. Development of an Instrument for Measuring Different Types of Cognitive Load. *Behav. Res. Methods* **2013**, *45*, 1058–1072. [CrossRef]
30. Sweller, J.; Chandler, P. Why Some Material Is Difficult to Learn. *Cogn. Instr.* **1994**, *12*, 185–233. [CrossRef]
31. Makransky, G.; Terkildsen, T.S.; Mayer, R.E. Adding Immersive Virtual Reality to a Science Lab Simulation Causes More Presence but Less Learning. *Learn. Instr.* **2019**, *60*, 225–236. [CrossRef]
32. Beege, M.; Wirzberger, M.; Nebel, S.; Schneider, S.; Schmidt, N.; Rey, G.D. Spatial Continuity Effect vs. Spatial Contiguity Failure. Revising the Effects of Spatial Proximity between Related and Unrelated Representations. *Front. Educ.* **2019**, *4*, 86. [CrossRef]
33. Korbach, A.; Brünken, R.; Park, B. Learner Characteristics and Information Processing in Multimedia Learning: A Moderated Mediation of the Seductive Details Effect. *Learn. Individ. Differ.* **2016**, *51*, 59–68. [CrossRef]
34. Garzón, J.; Pavón, J.; Baldiris, S. Systematic Review and Meta-Analysis of Augmented Reality in Educational Settings. *Virtual Real.* **2019**, *23*, 447–459. [CrossRef]
35. Batdi, V.; Talan, T. Augmented Reality Applications: A Meta-Analysis and Thematic Analysis. *Turk. J. Educ.* **2019**, *8*, 276–297. [CrossRef]
36. Cai, Y.; Pan, Z.L.; Liu, M. Augmented Reality Technology in Language Learning: A Meta-Analysis. *J. Comput. Assist. Learn.* **2022**, *38*, 929–945. [CrossRef]
37. Habig, S. Who Can Benefit from Augmented Reality in Chemistry? Sex Differences in Solving Stereochemistry Problems Using Augmented Reality. *Br. J. Educ. Technol.* **2019**, *51*, 629–644. [CrossRef]
38. Lohman, D.F. Spatial Ability and g. In *Human Abilities: Their nature and measurement*; Dennis, I., Tapsfield, P., Eds.; Erlbaum: Hillsdale, NJ, USA, 1996; pp. 97–116. ISBN 978-0-203-77400-7.
39. Carroll, J.B. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*; Cambridge University Press: Cambridge, UK, 1993; ISBN 978-0-521-38712-5.
40. Miyake, A.; Friedman, N.P.; Rettinger, D.A.; Shah, P.; Hegarty, M. How Are Visuospatial Working Memory, Executive Functioning, and Spatial Abilities Related? A Latent-Variable Analysis. *J. Exp. Psychol. Gen.* **2001**, *130*, 621–640. [CrossRef] [PubMed]
41. Hegarty, M.; Waller, D.A. Individual Differences in Spatial Abilities. In *The Cambridge Handbook of Visuospatial Thinking*; Shah, P., Miyake, A., Eds.; Cambridge University Press: Cambridge, UK, 2005; pp. 121–169.
42. Köhl, T.; Fehrer, B.C.O.F.; Münzer, S. Unifying the Ability-as-Compensator and Ability-as-Enhancer Hypotheses. *Educ. Psychol. Rev.* **2022**, *34*, 1063–1095. [CrossRef]
43. Höfler, T.N. Spatial Ability: Its Influence on Learning with Visualizations—A Meta-Analytic Review. *Educ. Psychol. Rev.* **2010**, *22*, 245–269. [CrossRef]
44. Buckley, J.; Seery, N.; Canty, D. A Heuristic Framework of Spatial Ability: A Review and Synthesis of Spatial Factor Literature to Support Its Translation into STEM Education. *Educ. Psychol. Rev.* **2018**, *30*, 947–972. [CrossRef]

45. Opfermann, M.; Schmeck, A.; Fischer, H.E. Multiple Representations in Physics and Science Education—Why Should We Use Them? In *Multiple Representations in Physics Education*; Treagust, D., Duit, R., Fischer, H., Eds.; Models and Modeling in Science Education; Springer: Cham, Switzerland, 2017.
46. Gyselinck, V.; Ehrlich, M.F.; Cornoldi, C.; De Beni, R.; Dubois, V. Visuospatial Working Memory in Learning from Multimedia Systems. *J. Comput. Assist. Learn.* **2000**, *16*, 166–176. [CrossRef]
47. Mayer, R.E.; Sims, V.K. For Whom Is a Picture Worth a Thousand Words? Extensions of a Dualcoding Theory of Multimedia Learning. *J. Educ. Psychol.* **1994**, *86*, 389–401. [CrossRef]
48. Cromley, J.G. Learning from Multiple Representations: Roles of Task Interventions and Individual Differences. In *Handbook of Learning from Multiple Representations and Perspectives*; van Meter, P., List, A., Lombardi, D., Kendeou, P., Eds.; Routledge: New York, NY, USA, 2020; pp. 62–75. ISBN 978-0-429-44396-1.
49. Diaz, D.D.; Sims, V.K. Augmenting Virtual Environments: The Influence of Spatial Ability on Learning from Integrated Displays. *High Abil. Stud.* **2003**, *14*, 191–212. [CrossRef]
50. Huk, T. Who Benefits from Learning with 3D Models? The Case of Spatial Ability. *J. Comput. Assist. Learn.* **2006**, *2006*, 392–404. [CrossRef]
51. van Bruggen, J.M.; Kirschner, P.A.; Jochems, W. External Representation of Argumentation in CSCL and the Management of Cognitive Load. *Learn. Instr.* **2002**, *12*, 121–138. [CrossRef]
52. Hays, T.A. Spatial Abilities and the Effects of Computer Animation on Short-Term and Long-Term Comprehension. *J. Educ. Comput. Res.* **1996**, *14*, 139–155. [CrossRef]
53. Höffler, T.N.; Leutner, D. The Role of Spatial Ability in Learning from Instructional Animations—Evidence for an Ability-as-Compensator Hypothesis. *Comput. Hum. Behav.* **2011**, *27*, 209–216. [CrossRef]
54. Köhl, T.; Stebner, F.; Navratil, S.C.; Fehrer, B.C.O.F.; Münzer, S. Text Information and Spatial Abilities in Learning with Different Visualizations Formats. *J. Educ. Psychol.* **2018**, *110*, 561–577. [CrossRef]
55. Krüger, J.M.; Buchholz, A.; Bodemer, D. Augmented Reality in Education: Three Unique Characteristics from a User's Perspective. In Proceedings of the 27th International Conference on Computers in Education, Taiwan, China, 2–6 December 2019; pp. 412–422.
56. Cheng, K.H.; Tsai, C.C. Affordances of Augmented Reality in Science Learning: Suggestions for Future Research. *J. Sci. Educ. Technol.* **2013**, *22*, 449–462. [CrossRef]
57. Krüger, J.M.; Palzer, K.; Bodemer, D. Learning with Augmented Reality: Impact of Dimensionality and Spatial Abilities. *Comput. Educ. Open* **2022**, *3*, 100065. [CrossRef]
58. Dünser, A.; Steinbügl, K.; Kaufmann, H.; Glück, J. Virtual and Augmented Reality as Spatial Ability Training Tools. In Proceedings of the 7th ACM SIGCHI New Zealand Chapter's International Conference on Computer-Human Interaction Design Centered HCI—CHINZ '06, Christchurch, New Zealand, 6–7 July 2006; pp. 125–132.
59. Gün, E.T.; Atasoy, B. The Effects of Augmented Reality on Elementary School Students' Spatial Ability and Academic Achievement. *Egit. Ve Bilim-Educ. Sci.* **2017**, *42*, 31–51. [CrossRef]
60. Hoe, Z.Y.; Lee, I.J.; Chen, C.H.; Chang, K.P. Using an Augmented Reality-Base Training System to Promote Spatial Visualization Ability for the Elderly. *Univers. Access Inf. Soc.* **2019**, *18*, 327–342. [CrossRef]
61. Keller, S.; Habig, S. Supporting Spatial Thinking in Organic Chemistry Through Augmented Reality—An Explorative Interview Study. In *Student Reasoning in Organic Chemistry*; Graulich, N., Shultz, G., Eds.; Advances in Chemistry Education Research; Social Society of Chemistry: Cambridge, UK, 2022; pp. 19–35. ISBN 978-1-83916-491-0.
62. Seufert, T.; Schütz, M.; Brünken, R. Memory Characteristics and Modality in Multimedia Learning: An Aptitude-Treatment-Interaction Study. *Learn. Instr.* **2009**, *19*, 28–42. [CrossRef]
63. Schüler, A.; Scheiter, K.; van Genuchten, E. The Role of Working Memory in Multimedia Instruction: Is Working Memory Working during Learning from Text and Pictures? *Educ. Psychol. Rev.* **2011**, *23*, 389–411. [CrossRef]
64. Pazzaglia, F.; Toso, C.; Cacciamani, S. The Specific Involvement of Verbal and Visuospatial Working Memory in Hypermedia Learning. *Br. J. Educ. Technol.* **2008**, *39*, 110–124. [CrossRef]
65. Plass, J.L.; Chun, D.M.; Mayer, R.E.; Leutner, D. Cognitive Load in Reading a Foreign Language Text with Multimedia Aids and the Influence of Verbal and Spatial Abilities. *Comput. Hum. Behav.* **2003**, *19*, 221–243. [CrossRef]
66. Kruley, P.; Sciamia, S.C.; Glenberg, A.M. Online Processing of Textual Illustrations—In the Visuospatial Sketchpad - Evidence for Dual-Task Studies. *Mem. Cognit.* **1994**, *22*, 262–272. [CrossRef] [PubMed]
67. Fenesi, B.; Kramer, E.; Kim, A. Split-Attention and Coherence Principles in Multimedia Instruction Can Rescue Performance for Learners with Lower Working Memory Capacity. *Appl. Cogn. Psychol.* **2016**, *30*, 691–699. [CrossRef]
68. Urban-Woldron, H.; Hopf, M. Entwicklung Eines Testinstruments Zum Verständnis in Der Elektrizitätslehre [Development of a Diagnostic Instrument for Testing Student Understanding of Basic Electricity Concepts]. *Z. Für Didakt. Naturwiss.* **2012**, *18*, 201–227.
69. Burde, J.P. *Konzeption Und Evaluation Eines Unterrichtskonzepts Zu Einfachen Stromkreisen Auf Basis Des Elektronengasmodells*; Logos Verlag: Berlin, Germany, 2018; ISBN 978-3-8325-4726-4.
70. Altmeyer, K.; Barz, M.; Lauer, L.; Peschel, M.; Sonntag, D.; Brünken, R.; Malone, S. Digital Ink and Differentiated Subjective Ratings for Cognitive Load Measurement in Middle Childhood. *Br. J. Educ. Psychol.* **2023**, *93*, 368–385. [CrossRef]
71. Ekstrom, R.B.; French, J.W.; Harman, H.H.; Dermen, D. *Manual for Kit of Factor-Referenced Cognitive Tests*; Educational Testing Service: Princeton, NJ, USA, 1976.

72. Della Sala, S.; Gray, C.; Baddeley, A.; Wilson, L. *Visual Patterns Test: A Test of Short-Term Visual Recall*; Thames Valley Test Company: Suffolk, UK, 1997; ISBN 978-1-874261-16-2.
73. Thurstone, T.G.; Thurstone, L.L. *Primary Mental Abilities Tests*; Science Research Associates: Chicago, IL, USA, 1962.
74. Wechsler, D. *Wechsler Intelligence Scale for Children*, 5th ed.; Pearson: Bloomington, MN, USA, 2014.
75. Hayes, A.F. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*, 3rd ed.; The Guilford Press: New York, NY, USA, 2022; ISBN 978-1-4625-4903-0.
76. Hegarty, M.; Kriz, S. Effects of Knowledge and Spatial Ability on Learning from Animation. In *Learning in Animation: Research Implications for Design*; Lowe, R., Schnotz, W., Eds.; Cambridge University Press: Cambridge, UK, 2008; pp. 3–29.
77. Maxwell, S.E. Sample Size and Multiple Regression Analysis. *Psychol. Methods* **2000**, *5*, 434–458. [CrossRef]
78. Green, S.B. How Many Subjects Does It Take To Do A Regression Analysis. *Multivar. Behav. Res.* **1991**, *26*, 499–510. [CrossRef]
79. Schmidt, F.L. What Do Data Really Mean? Research Findings, Meta-Analysis, and Cumulative Knowledge in Psychology. *Am. Psychol.* **1992**, *47*, 1173–1181. [CrossRef]
80. Heo, M.; Toomey, N. Learning with Multimedia: The Effects of Gender, Type of Multimedia Learning Resources, and Spatial Ability. *Comput. Educ.* **2020**, *146*, 103747. [CrossRef]
81. Schroeder, N.L.; Cenkci, A.T. Do Measures of Cognitive Load Explain the Spatial Split-Attention Principle in Multimedia Learning Environments? A Systematic Review. *J. Educ. Psychol.* **2020**, *112*, 254–270. [CrossRef]
82. Khan, T.; Johnston, K.; Ophoff, J. The Impact of an Augmented Reality Application on Learning Motivation of Students. *Adv. Hum.-Comput. Interact.* **2019**, *2019*, 7208494. [CrossRef]
83. Plass, J.L.; Kalyuga, S. Four Ways of Considering Emotion in Cognitive Load Theory. *Educ. Psychol. Rev.* **2019**, *31*, 339–359. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Comparing Real and Imitative Practice with No Practice during Observational Learning of Hand Motor Skills from Animations

Maliha Naushad Mian ¹, Hannah Beder ¹, Nadine Marcus ^{1,*} and Paul Ayres ²¹ School of Computer Science & Engineering, University of New South Wales, Sydney, NSW 2052, Australia; m.mian@unsw.edu.au (M.N.M.); hannah.beder@gmail.com (H.B.)² School of Education, University of New South Wales, Sydney, NSW 2052, Australia; p.ayres@unsw.edu.au

* Correspondence: nadinem@unsw.edu.au

Abstract: In two experiments, we compared the effects of practice (real and imitative) with no practice on the observational learning of hand motor skills from animated videos. Experiment 1 investigated learning to play a series of piano clips of varying complexity. Results demonstrated improved learning efficiency with imitative practice compared to no practice. Experiment 2 featured a paper-folding task, and results indicated that real practice led to significantly greater learning than no practice. Furthermore, a significant interaction was found with gender and practice, where females learned best with both real and imitative practice, but males did not benefit from these interventions. However, males outperformed females in the no practice condition. Overall, we found benefits of practice versus no practice for both tasks. However, the most effective type of practice was dependent upon the task: imitative practice for piano playing, and real practice for paper folding. Task complexity and gender were also found to be moderating factors.

Keywords: cognitive load theory; embodied cognition; observational learning; practice; gender; instructional animations

Citation: Mian, M.N.; Beder, H.; Marcus, N.; Ayres, P. Comparing Real and Imitative Practice with No Practice during Observational Learning of Hand Motor Skills from Animations. *Educ. Sci.* **2024**, *14*, 949. <https://doi.org/10.3390/educsci14090949>

Academic Editor: Dorothe Kienhues

Received: 22 July 2024

Revised: 23 August 2024

Accepted: 26 August 2024

Published: 28 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In educational research, there is a need to design online learning content that is effective for learning across all students and that is relevant to multiple disciplines. Learning using instructional animations [1–3] and incorporating practice [4,5] while learning are both areas that have been studied by many researchers, but mostly independently of each other. The present study addresses this gap by combining both domains to explore when and how it is best to use practice to support learning from instructional animations. By doing so, the study aims to provide a more comprehensive understanding of the interaction between practice and instructional-based animation learning, leading to more effective educational designs. The purpose of this study is to compare the effects of real and imitative practice with no practice for learning two different hand motor-based skills from instructional animations. By investigating these effects, the research can offer practical insights into how educators can better tailor their instructional designs based on the learning domain and learner characteristics, with the end goal of improving learning outcomes across different contexts.

1.1. Instructional Animations

Instructional animations have been used across many domains and have been highly successful in teaching hand motor tasks [1–3,6]. We use the term animations to refer to instructional animations, including both video-based animations and animated graphics, which is consistent with how many researchers have used this term (e.g., [1,2,6–8]). A meta-analysis conducted by Hoffler and Leutner [9] suggests that animations lead to better learning compared to equivalent statics for movement-based tasks. Many studies

have found advantages for animations in tasks like tying knots [10], conducting first aid procedures [11], building Lego shapes [12] and creating origami shapes [8,13].

Cognitive load theory [14] has been used to help explain why instructional animations may be preferred for learning motor tasks based on the mirror neuron system (see [10,13]). A mirror neuron is a type of neuron that has been found to be involved in the learning of motor tasks—it fires when an individual is performing an action or while observing another person performing the same action [15,16]. This has been found to support our ability to engage in imitative learning, although the activation patterns during imitation are not quite as pronounced or the same as during physical practice [17]. Thus, instructional animations of motor tasks may activate these neurons. Additionally, Paas and Sweller [18] argue that movement-based learning can be acquired with less effort, since it is a form of biologically primary knowledge. Biologically primary knowledge refers to what we have evolved to acquire fairly easily without much conscious effort. It involves minimal taxing of our working memory, as it has an innate component to it. Some examples include learning our first language, face recognition, basic human movements, and the ability to interact socially [19–21]. Therefore, arguably, animations displaying human movements may be easier to learn from than those representing mechanical non-human movements.

Some considerations are important to understand when designing instructional animations. The transient nature of animations [22] has been found to cause a burden on our limited working memory, especially in certain contexts, such as dealing with complex information that is high in element interactivity. Information that is high in element interactivity involves multiple interacting elements that need to be simultaneously processed to be understood [23]. However, animations have proven to be effective and superior to statics when the impact of transient information is reduced by dividing the animations into shorter sections with pauses [8]. Additionally, the pace of an animation has the potential to interfere with the quality of the learner's perception of the task and thus inhibit their ability to reproduce the learning task, particularly if the animation speed is high [24]. Moreover, when several interacting elements are required to be processed simultaneously, the increase in intrinsic load may overburden the learner's working memory [23]. Transience, pace, and levels of element interactivity are all factors that need to be considered when developing instructional animations [8,23]. However, considerations of embodied cognition can help alleviate these factors.

1.2. Embodied Cognition

The embodied theory of cognition refers to the linking of the mind with the body and environment [25]. Since action, perception, and cognition are interrelated, it allows for the exploration of how learning can become a bodily experience using human movement. An example of this was demonstrated by Mavilidi et al. [26], who investigated the effect of integrating physical activities into a science lesson for preschool children's learning. Integrating physical activity was found to lead to significantly higher learning outcomes and perceived enjoyment of learning compared to a sedentary style of teaching. Problem-solving for real mechanical systems including pulleys have also been found to improve learning when there were hands-on interactions rather than only referring to diagrams [27]. More recently Chettaoui et al. [28] found benefits to involving body movements with user interfaces such as manipulation of objects, compared to just learning from the interface, for the long-term learning of the basics of human anatomy. Hence, embodied cognition has been found to have benefits when it comes to learning.

In contrast to physical interactions, gesturing has also been found to facilitate learning, since it reduces cognitive load due to embodied cognition related effects. Goldin-Meadow et al. [29] found that both adults and children were able to remember a significantly more extensive list of letters and words as they gestured while solving a math problem compared to the groups that did not gesture. Additionally, Valenzano et al. [30] found that preschool children who studied an instructional animation on symmetry containing both narration and observation of teachers' gestures outperformed the children in the narration-only

group. Cook et al. [31] asked children to view a computer avatar providing instruction for mathematics either with or without gestures. Children who viewed the gesturing avatar were found to learn and solve problems quicker. Hence, there is evidence that learning performance can be improved with the use of gestures, as it is an embodied cognition strategy that has successfully been used to support learning.

Benefits associated with mimicking have also been found in diverse domains, including dance [32], guitar playing [17], and vocabulary learning [3,33,34]. Learning to play musical instruments can involve making meaningful imitative gestures relevant to the task at hand. Mierowsky et al. [7] explored the effect of mimicking gestures on learning to play musical scores on the piano. For novices, it was found that mimicking gestures led to greater performance while playing the easy and medium musical pieces, whereas for more experienced learners, gesturing was beneficial for the more difficult pieces (since gesturing was found to be redundant for the easy and medium pieces). Hence, mimicking gestures of how to perform a task can be used to enhance the learning process due to embodied cognition-related benefits.

1.3. Practice

Practice during learning can help support the recall of new information and the automation of new skills, which can free up cognitive resources to deal with more demanding tasks [35]. Practice plays a major role in how well we learn and acquire expertise [36]. It does not simply involve repeating a task but also focuses on attention, rehearsal, and repetition, which can lead to new knowledge acquisition that can be a starting point for more complex learning. Practice has been found to have positive effects on performance in various fields that involve both motor and cognitive skill acquisition, such as chess and music [37–39]. Practice can be a form of embodied cognition when it involves direct tactile interaction with the learning task at hand, e.g., playing the piano or creating origami while learning how to complete the tasks, or when it involves imitating or mimicking the task at hand, e.g., writing a character in the air.

Various practice types have been incorporated to assess their impact on learning. Of interest to the present study is the comparison of real and imitative practice. McBride and Rothstein [5] explored the benefits of real physical versus mental practice and a combination of both when learning to use a paddle to hit a golf ball to a target. They found that all forms of practice led to improved skill accuracy, with combined practice being best, followed by physical practice and mental practice, respectively. Their study supports the benefits of all types of practice when learning motor skills, with real benefits associated with real physical practice. De Stefani et al. [4] found benefits associated with observational imitative learning (where participants observed performance by an expert athlete) for young children when learning new sports actions when compared to descriptive learning that was integrated with static images. Hence, the studies described above demonstrate that physical, mental, and observational forms of practice have all been found to be effective for supporting the learning of motor skills.

1.4. Present Study

The present study investigates the impact of practice type on learning from instructional animations. The study comprises two experiments, with each focusing on a different hand motor-based task—piano playing (Experiment 1) and paper folding (Experiment 2). The overarching research question of the study is: How do real and imitative practices compare to no practice in terms of learning performance, efficiency, and cognitive load in the observational learning of these two tasks? This research question has been broken down into the following sub-topics:

- Learning Performance—How do real practice and imitative practice differ in terms of learning outcomes compared to no practice?
- Cognitive Load—How do real practice and imitative practice differ in terms of cognitive load compared to no practice?

- Learning Efficiency—How do real practice and imitative practice differ in terms of learning efficiency compared to no practice?
- Impact of Learning Task—Does the type of task, e.g., varying the sensory modality involved, influence the effectiveness of different practice types?

The study, consisting of two experiments, explores different types of practice while learning from instructional animations. The first experiment involves the hand motor task of piano playing, and the second uses a different hand motor task of paper folding. The impact of physically interacting with an object (real practice) versus using mimicking gestures (imitative practice) or no practice (observational only) while learning is explored in both experiments. Experiment 1 involves pressing keys on the keyboard for real practice (with no sound) and air piano playing for imitative practice. Experiment 2 involves folding the origami paper for real practice and simulating the paper folding hand actions for imitative practice. We explore tasks that included the mediums of audio and tactile (piano playing) and just tactile (origami) to observe whether similar or different effects were observed. Building on Experiment 1, Experiment 2 additionally considers the impact of gender and spatial ability on the different interventions.

As discussed earlier, instructional animations, when incorporating practice, have demonstrated the potential to enhance learning outcomes. We test a series of hypotheses across the two experiments, with the expectation that practice will support learning. Since real practice involves direct physical engagement with the task, it is hypothesized to lead to better learning performance with greater activation of the mirror neuron system [17]. By becoming more familiar with the tasks, learners are expected to have a reduced cognitive load, which is further expected to translate into greater learning efficiency, meaning the amount of learning achieved relative to invested cognitive load [40]. Accordingly, the following hypotheses were developed:

Hypothesis 1a. *Real practice will lead to better learning performance than no practice.*

Hypothesis 1b. *Real practice will lead to less cognitive load than no practice.*

Hypothesis 1c. *Real practice will lead to greater learning efficiency than no practice.*

The impact of imitative practice compared to no practice is also explored. Since imitative practice involves mentally rehearsing the task by mimicking actions, it is expected to also provide benefits in terms of learning performance and efficiency, with reduced cognitive load.

Hypothesis 2a. *Imitative practice will lead to better learning performance than no practice.*

Hypothesis 2b. *Imitative practice will lead to less cognitive load than no practice.*

Hypothesis 2c. *Imitative practice will lead to greater learning efficiency than no practice.*

We are also interested in the open research question of whether there are any learning differences between real and imitative practices for different learning content and under different learning conditions. Real practice, which involves physical engagement with the task, is expected to activate the mirror neuron system more effectively than imitative practice, leading to better learning performance.

2. Experiment 1

The first experiment used the same instructional materials as Mierowsky et al. [7], where they explored the effect of gesturing while learning different piano scores. Mierowsky et al. [7] found differences in benefits of mimicking gestures to support learning; the gestures were most useful when the instructional content was of a moderate level of difficulty, as per

level of expertise. For the experts, for the easier tasks, the gestures were redundant, but for the more complex tasks, they were beneficial. In contrast, for novices, for the easier tasks, the gestures supported learning, while for the very complex tasks, they potentially became overwhelming. In Experiment 1, all three groups observed video recordings. However, one group was observational learning only (no practice), the second group engaged in imitative practice (gesturing in the air), and the third group participated in real practice (pressing on the keyboard with the sound off).

Since this first experiment involved tasks with differing levels of difficulty, we expected to see potential interactions among the three groups consistent with Mierowsky et al. [7].

2.1. Method

2.1.1. Participants

Sixty university students (38 females, 22 males) between the ages of 18–44 were recruited to participate. A questionnaire was given to participants before the instructional phase to determine their experience level in playing the piano. At least one year of musical experience was required to be included in the study, including some piano playing experience. The level of musical expertise self-reported by participants was used to allocate participants equally across the three groups (imitative practice, real practice, and no practice), with 20 in each group.

2.1.2. Materials

2.1.2.1. Learning Materials

The instructional content for this experiment consisted of a series of piano-playing clips of varying difficulty. Six short video clips of hands playing the piano, accompanied by audio, were included. Two of these clips were practice clips provided to familiarize participants with the format of the videos.

The video clips were designed by Mierowsky et al. [7], with a focus on the clips being both a manageable length and tailored for a range of musical expertise levels. Since the information was transitory and there was no user control when watching the video clips, it was essential to ensure that the amount of transitory visual and auditory information presented was digestible. The difficulty of the pieces was set to avoid a floor or a ceiling effect (i.e., making it too easy or hard). Lengths and difficulty levels were decided based on the results of testing conducted by Mierowsky et al. [7].

The clips varied in difficulty, with Clip 1 being easy (consisting of a single hand playing for 10 s), Clip 2 (easy–medium), Clip 3 (medium–difficult), and Clip 4 (difficult), which consisted of one hand playing for 15 s. The clips were developed by Mierowsky et al. [7] using a Canon EOS-M digital camera on a tripod and a Casio Privia PX-330 electric piano.

Musical tenets such as phrasing, intervals, chords, and cadences enable musicians to remember, recall, or predict musical information accurately. Clips 2 and 3 take advantage of these elements, with some notes being played simultaneously in recognizable chords and intervals. Clip 4, however, does not rely on these musical characteristics and instead presents a series of individual notes that make little musical sense. Although Clip 4 is played with only one hand and might appear to be less challenging than Clips 2 and 3, it was found to be more difficult when considered through a musical lens.

All participants were presented with the same material in the same order using a Gigabyte P34G v2 14-inch laptop.

2.1.2.2. Cognitive Load Measures

To obtain a measure of cognitive load, subjective difficulty rating scale measures were used, adapted from Paas [41]. The Likert scale questions included “Please indicate how easy or difficult it was for you to understand the learning material” and “Please indicate how easy or difficult it was for you to play back the musical phrase”. The questions were followed by a 9-point Likert scale, with ratings ranging from “1—Extremely easy” to “9—Extremely difficult”.

2.1.3. Procedure

Each participant was welcomed, signed the consent form, and was briefed about the study's requirements. They were asked to watch each clip four times with a 3 s break between each viewing. Participants in the no practice group were instructed to place their hands flat on the table in front of them while watching the clips to prevent them from gesturing. Participants in the practice groups were asked to place their hands flat on the table for the first two viewings and then imitate the hand movements seen for their third and fourth viewing. Those in the imitative practice group gestured in the air as though playing the piano along with the clips, without tactile feedback. Participants in the real practice group pressed on the keyboard with the sound switched off, mimicking playing along with the clips. After viewing each clip four times, participants were asked to rate the learning difficulty on the Likert scale. They then entered the testing phase, where they were asked to play the musical piece they learned on a keyboard provided, with the audio turned on. After playing the piece, participants were asked to rate the testing difficulty on the Likert scale. Participants were thanked, and a \$20 gift voucher was given upon completion of the experiment.

2.1.4. Scoring Participants

The Levenshtein distance D was used to measure the difference between the correct music sequence and the music sequence performed by the participant [42]. Additionally, idle time, T , i.e., time taken to think and process the musical phrase, was accounted for. The performance score was calculated by deducting the sum of idle time T and Levenshtein distance D from the maximum score M as follows:

$$\text{Performance} = \max(M - (D + \min(T, 4)), 0). \quad (1)$$

The efficiency metric combined both cognitive load and task performance to consider the cognitive costs of learning [40]. Standardized scores for performance and cognitive load, Z , were considered to calculate the relative efficiency of the instructional conditions as follows:

$$\text{Efficiency} = \frac{Z_{\text{Performance}} - Z_{\text{Cognitive Load}}}{\sqrt{2}}. \quad (2)$$

2.2. Results

2.2.1. Performance Scores

A one-way repeated measures ANOVA was conducted on performance scores for the four different piano clips (mean scores shown in Table 1).

Table 1. Mean performance scores and standard deviations for different practice conditions and clips.

Condition	Clip 1	Clip 2	Clip 3	Clip 4	Combined Clips
Real practice	72.8 (33.6)	70.2 (29.1)	65.9 (34.5)	69.0 (20.0)	69.5 (29.3)
Imitative practice	83.3 (23.0)	85.0 (19.5)	74.7 (26.9)	70.0 (19.7)	78.3 (22.3)
No practice	70.2 (31.1)	70.1 (29.9)	54.4 (33.8)	54.7 (23.8)	62.4 (29.7)
Combined conditions	75.4 (29.2)	75.1 (26.2)	65.0 (31.7)	64.6 (21.2)	70.0 (27.1)

Within-subjects effects: There was a significant within-subjects effect [$F(3, 171) = 5.4$, $p = 0.001$, $\eta_p^2 = 0.088$], indicating that performance varied over the musical clips, where participants scored lower on the last 2 clips.

Between-subjects effects: The between-subjects effects just failed to reach significance at the $p = 0.05$ level [$F(2, 57) = 2.76$, $p = 0.07$, $\eta_p^2 = 0.088$].

Interaction effects: There was no practice $\times$ musical clip interaction [$F < 1$].

2.2.2. Cognitive Load Measures

Within-subjects effects: There was a significant within-subjects effect for cognitive load measures [$F(3, 171) = 14.6, p < 0.001, \eta_p^2 = 0.20$], indicating variations over the 4 musical clips. As can be seen from Table 2, participants found the later musical clips to be more difficult than the earlier ones.

Table 2. Mean of cognitive load (max = 9) and standard deviations for all clips.

Condition	Clip 1	Clip 2	Clip 3	Clip 4	Combined Clips
Real practice	4.60 (1.67)	5.25 (1.74)	5.25 (1.99)	6.60 (1.47)	5.42 (1.72)
Imitative practice	3.90 (1.74)	5.05 (1.64)	5.10 (2.05)	6.15 (1.84)	5.05 (1.82)
No practice	4.90 (1.59)	5.45 (1.57)	6.10 (1.41)	7.30 (1.26)	5.94 (1.46)
Combined conditions	4.47 (1.70)	5.25 (1.63)	5.48 (1.86)	6.68 (1.52)	5.47 (1.67)

Between-subjects effects: The between-subjects effects for cognitive load just failed to reach significance at the $p = 0.05$ level [$F(2, 57) = 2.830, p = 0.067, \eta_p^2 = 0.090$]. Notably, the no practice condition encountered the greatest level of difficulty (see Table 2), consistent with the performance means (see Table 1).

Interaction effects: There was no practice $\times$ musical clip interaction for the cognitive load measure [$F < 1$].

2.2.3. Efficiency Measures

Mean efficiency scores are presented in Table 3.

Within-subjects effects: There were no significant differences in efficiency scores within subjects [$F < 1$].

Between-subjects effects: There was a significant difference between conditions for efficiency scores [$F(2, 57) = 3.92, p = 0.025, \eta_p^2 = 0.12$]. Follow-up Bonferroni post hoc tests indicated that the imitative practice group had significantly higher efficiency scores than the no practice group ($p = 0.021$). No other pair-wise comparisons were significant.

Interaction effects: There was no practice $\times$ musical clip interaction for the efficiency measure [$F < 1$].

Table 3. Mean efficiency scores and standard errors across all clips.

Condition	Clip 1	Clip 2	Clip 3	Clip 4	Combined Clips
Real practice	−0.118 (1.23)	−0.128 (1.24)	0.109 (1.22)	0.179 (0.998)	0.010 (0.200)
Imitative practice	0.424 (1.23)	0.345 (1.24)	0.356 (1.22)	0.412 (0.998)	0.380 (0.200)
No practice	−0.306 (1.23)	−0.217 (1.24)	−0.465 (1.22)	−0.591 (0.998)	−0.400 (0.200)

2.2.4. Exploratory Analysis

In the above analysis, only efficiency scores produced a significant difference between practice conditions. However, for both performance and cognitive load scores, there were near misses (both $p < 0.07$). Examination of Tables 1 and 2 suggests that there may be differences according to individual clips, even though interactions were not detected. Therefore, some exploration analyses (ANOVA and post hoc procedures) were conducted on individual clips. For Clips 1, 2, and 3, no significant differences were found. However, for Clip 4, significant differences were found between imitative practice and no practice, showing that imitative practice had greater performance and efficiency scores with less cognitive load than the no practice group. Furthermore, for this clip, the real practice group also had significantly higher performance scores than the no practice group.

Although this analysis is exploratory in nature, some weak evidence emerged that the imitative practice condition had the most benefits for the musical clip with the highest difficulty to perform.

2.3. Discussion

Both hypotheses were tested in this experiment. While the type of practice had no significant effect on performance or cognitive load measures overall, some significant differences were observed as the video clips became more difficult. Participants who used imitative practice in the third (medium–difficult) and fourth (difficult) clips experienced greater performance scores than those who did not practice. Imitative practice in the fourth clip also resulted in lower cognitive load scores compared to no practice. Additionally, real practice in the fourth clip resulted in higher performance scores than no practice. Significant results were also observed in the efficiency scores across all clips. Participants who used imitative practice during the learning process experienced greater efficiency than no practice.

The overall ANOVAs across all four musical clips revealed that both real practice and imitative practice did not lead to greater learning or lower cognitive load than no practice, and therefore Hypotheses 1a, 1b, 2a, 2b could not be supported over the whole problem set. Hypothesis 2c was supported, as imitative practice was more efficient than no practice, but real practice was not more efficient than no practice, and therefore Hypothesis 1c was not supported. However, the exploratory analysis looking at individual musical clips suggests that the complexity of the clips may have moderated the effectiveness of the practice groups. This evidence suggests that, on the most difficult clip (fourth), imitative practice led to greater performance and efficiency, as well as lower cognitive load, than the no practice group. In addition, real practice had higher performance scores than the no practice group, suggesting that both practice forms were helpful.

Regarding the open question investigating differences between the two practice conditions, the weak exploratory evidence for Clip 4 suggests that, for this learning topic, imitative practice was superior to no practice only, suggesting an advantage for imitation.

3. Experiment 2

In Experiment 1, some significant differences were found in favor of imitative practice over no practice, but not for real practice. Hence, it remained to be seen whether the effects of real and imitative practice would differ in a different task requiring different sensory mediums. Thus, Experiment 2 was completed using a different movement-based learning task of paper folding. In the present study, spatial ability and gender have also been considered, as both factors have been found to influence performance in spatial-based motor tasks [1,2,43]. The spatial ability of an individual indicates their capacity to interact with and manipulate visual input [43]. Past research suggests that gender differences in spatial ability and learning exist [12,44–46] and are affected by various factors, including the type of visualizations used to measure learning and spatial ability. In particular, spatial ability has been found to be higher in males compared to females [47–49], with a meta-analysis conducted by Linn and Petersen [45] suggesting that gender differences arise on some types of spatial ability but not others, with the largest difference being found on measures of mental rotation and smaller differences found on measures of spatial perception. Moreover, research into the neuroanatomy of the human mirror neuron system supports potential gender differences, since females make use of mirror neurons (which fire when an action is observed and then recreated) to a greater degree compared to males [50–55].

3.1. Method

3.1.1. Participants

Sixty university students (30 females, 30 males) between the ages of 18 and 35 were recruited as participants. An online screening test was used to assess paper folding experience and collect demographic characteristics such as age and gender. Only participants with no experience or very limited paper folding experience were deemed eligible to participate. The participants were equally divided into three groups: imitative practice, where they were required to gesture hand movements in the air; real practice, where they completed

the actual paper folding task; and no practice. There were 20 participants in each group (F = 10, M = 10).

3.1.2. Materials

3.1.2.1. Learning Materials

Pilot trials were conducted on Wong et al.'s [8] animation materials to fine-tune the difficulty level of the paper folding tasks to avoid both ceiling and floor effects. Based on the findings of the pilot trials, the pace of the video was increased to 120%, speeding it up by approximately 40 s to prevent ceiling effects.

3.1.2.2. Cognitive Load Measures

Mental effort is an aspect of cognitive load that focuses on the cognitive capacity allocated by the learner to cater to the learning demands of a task and hence can be used to measure cognitive load [56]. As in the previous experiment, scales were adapted from Paas [18] and used to assess participants' perceived cognitive load levels after learning and completing the task. Unlike the previous experiment, the term 'mental effort' was used in the survey instead of 'difficulty' to attempt to more accurately reflect the cognitive resources invested by learners, and perhaps to then find significant differences in subjective ratings. De Leeuw and Mayer [57] found that mental effort is related to germane load, which relates to the cognitive resources contributing to learning, while task difficulty is more associated with intrinsic load, which is associated with the complexity of the learning material. This revised choice of wording was further validated with pilot testing.

3.1.2.3. Spatial Ability Test

A common measure of spatial ability is the Punched Holes Test [58]. It has been demonstrated that it is psychometrically reliable and significantly correlated with measures of executive functioning and working memory limitations [59,60]. Hence, the Punched Holes Spatial Ability Test was administered as a measure of spatial ability. It was administered as a two-part test, with 10 multiple choice questions in each part. For each question, participants were presented with an image of a paper being folded with a hole punched through it. They were given five options to correctly identify what the paper would look like when unfolded.

3.1.3. Procedure

Each participant was welcomed, signed the consent form, and was briefed about the study's requirements. They were then asked to complete the Punched Holes Test, which consisted of two parts, with three minutes given for each part, after receiving a verbal explanation of the task along with a sample question. Participants viewed the paper folding animation twice. In the no practice group, participants were asked to keep their hands on the table in front of them during both viewings to avoid any hand movements. For both practice groups, participants were asked to place their hands flat on the table in front of them during the first viewing. The real practice group was asked to complete the paper folding task using provided origami paper while watching the animation for the second time. The imitative practice group was briefed with four examples of meaningful hand gestures and then asked to mimic the hand movements of the task using mimicking gestures while watching the animation for the second time. Afterward, participants completed a post learning questionnaire to assess their mental effort during learning. They then entered the testing phase, where they were asked to complete the paper folding task in five minutes. Their hand movements were recorded using an iPhone camera placed on a desk-level tripod. Following the completion of the task, participants were asked to complete a post test questionnaire to assess their mental effort during the task. Participants were thanked and given a \$20 gift voucher upon completion of the experiment.

3.1.4. Scoring Participants

Wong et al.'s [8] marking criteria were revised to score the performance of participants. One mark was awarded for the correct completion of each of the 27 steps in the criteria, with a maximum possible score of 27/27. Since most participants were able to complete the task within the time limit, except for 10 participants, the scoring method aimed to differentiate participants who found it easier to complete the task compared to those who found it more difficult. Half a mark was deducted if a particular step was completed differently, was completed in the wrong direction, or if an extra step was added.

As with Experiment 1, efficiency was calculated to consider the cognitive costs of completing the task as follows:

$$\text{Efficiency} = \frac{Z_{\text{Performance}} - Z_{\text{Cognitive Load}}}{\sqrt{2}} \quad (3)$$

3.2. Results

3.2.1. Spatial Ability Scores

There was no significant difference between males ($M = 59.5$) and females ($F = 56.8$), $t(58) = 0.68$, $p = 0.50$.

3.2.2. Performance Measures

The mean performance scores (covariate-adjusted) and standard errors are reported in Table 4. Because of equal gender distributions, a 2-way ANCOVA with a factorial design of 3 (practice condition: real practice, imitative practice, no practice) $\times$ 2 (gender: female, male) was conducted, with the Punched Holes Test scores used as a covariate to control for spatial ability.

Table 4. Adjusted mean of performance (max = 100) and standard errors for females and males.

Condition	Overall	Females	Males
Real practice	65.2 (3.21)	70.3 (4.2)	58.8 (4.16)
Imitative practice	62.4 (3.21)	70.0 (4.2)	56.1 (4.17)
No practice	52.0 (3.19)	43.3 (4.9)	60.8 (4.94)

The 2-way ANCOVA found a main effect for performance scores [$F(2, 53) = 4.72$, $p = 0.013$]. Follow-up Bonferroni post hoc tests showed significantly higher performance in real practice compared to no practice ($p = 0.016$), but no significant difference was found between imitative and no practice ($p = 0.08$) or between real and imitative practice ($p > 0.10$).

There was no gender main effect ($F < 1$), but there was a significant interaction between practice conditions and gender [$F(2, 53) = 8.015$, $p = 0.001$]. To investigate this interaction (see Figure 1), follow-up analyses were completed for each gender. A one-way ANCOVA revealed a significant main effect on practice types for the performance of females [$F(2, 26) = 16.8$, $p < 0.01$]. Bonferroni post hoc tests demonstrated significantly higher performance in real practice compared to no practice ($p < 0.01$) and for imitative practice compared to no practice ($p < 0.001$), but no significant difference between real practice and imitative practice. There was no significant main effect across conditions for males ($F < 1$). However, it is interesting to further note that males scored greater than females ($p < 0.05$) when no practice was involved, but less than females for imitative practice ($p < 0.05$) (see Figure 1).

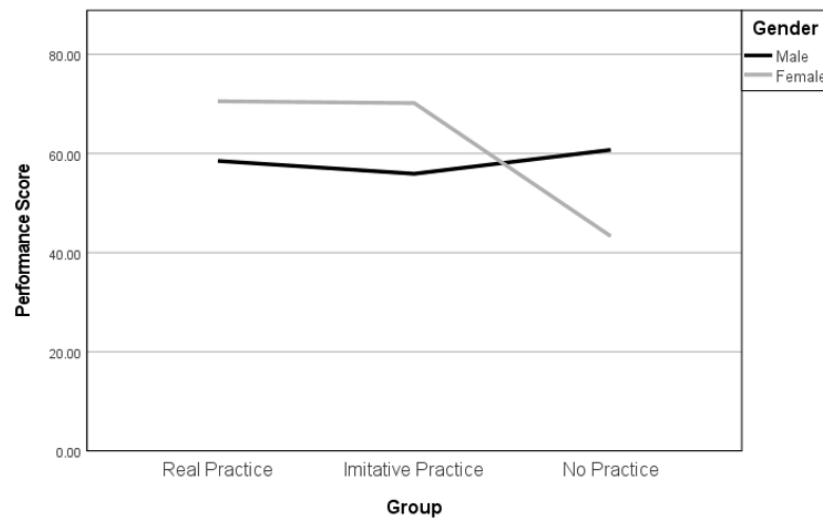


Figure 1. Performance mean scores (out of 100) across conditions for gender.

3.2.3. Cognitive Load Measures

The cognitive load mean scores (and SD) were 5.4 (1.4), 5.6 (1.8), and 5.4 (1.3) for real practice, imitative practice, and no practice, respectively. No significant main effects were found for practice condition and gender, nor was there a significant interaction (all $F < 1$).

3.2.4. Efficiency Measures

The mean efficiency-adjusted scores are presented in Table 5 and Figure 2. The ANCOVA test did not demonstrate a significant effect for practice conditions [$F(2, 53) = 0.771$, $p = 0.467$] or gender [$F(1, 53) = 0.146$, $p = 0.704$]; however, a significant interaction was found between practice conditions and gender [$F(2, 53) = 3.56$, $p = 0.035$]. To investigate this interaction, a one-way ANCOVA revealed a significant main effect for females across the practice types [$F(2, 26) = 4.92$, $p = 0.015$]. Bonferroni post hoc tests demonstrated real practice to be significantly more efficient than no practice ($p = 0.033$) and imitative practice to be significantly more efficient than no practice ($p = 0.041$). No significant main effect was found for the efficiency scores of males among the conditions using the ANCOVA test [$F < 1$]. Furthermore, when comparing males and females, no significant differences in efficiency were found across the practice conditions.

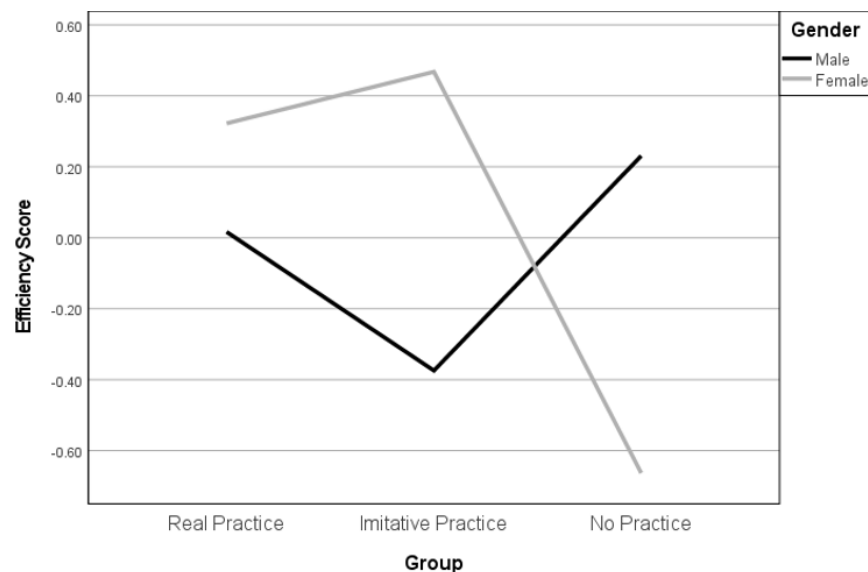


Figure 2. Efficiency mean scores across conditions for gender.

Table 5. Adjusted efficiency scores and standard errors.

Condition	Overall	Females	Males
Real practice	0.204 (0.240)	0.401 (0.343)	0.006 (0.338)
Imitative practice	0.012 (0.240)	0.426 (0.339)	−0.401 (0.338)
No practice	−0.216 (0.239)	−0.688 (0.338)	0.236 (0.338)

3.3. Discussion

Both broad hypotheses were tested in this experiment. Hypothesis 1a was supported, as real practice was superior to no practice for test performance, but not Hypothesis 2a, as imitative practice did not show any overall advantage over no practice (although it was supported for females). Hypotheses 1b and 2b were not supported, as no significant differences were found on the cognitive load measures. Hypotheses 1c and 2c were supported, but only for female learners, who demonstrated that both real and imitative practice were more efficient than no practice. For males, no significant differences were found for efficiency across conditions. Interestingly, no differences were found between males and females regarding the aspect of spatial ability measured, suggesting that spatial ability was not a factor in the gender interaction.

Controlling for spatial ability and gender allowed for an investigation of gender, where we found some significant gender effects, where females were found to learn best with both real and imitative practice, and no difference was found among practice conditions for males. Moreover, females learned more efficiently with both real and imitative practice than no practice, and there were no differences for males. In addition, the performance of females was significantly disadvantaged when learning with no practice in comparison to males (males performed significantly better without practice).

Regarding the open question, comparisons between the two practice conditions slightly favored real practice, as it demonstrated a clear test performance advantage over no practice for the whole sample, a benefit not shown by the imitative condition. However, when gender was considered, imitative practice was more advantageous for females than for males.

4. General Discussion

The two experiments aimed to compare different types of practice and their relative effectiveness at improving performance and learning efficiency while reducing cognitive load. The tasks in both experiments were from different disciplines and were human movement-based, with Experiment 1 focusing on piano playing and Experiment 2 on paper folding.

Experiment 1 demonstrated imitative practice to be superior, whereas real practice was superior for Experiment 2. In Experiment 1, imitative practice led to significantly increased learning efficiency compared to no practice, with the most difficult clip also having significantly improved performance and reduced cognitive load. In Experiment 2, real practice significantly enhanced learning performance and learning efficiency compared to no practice, although it did not reduce cognitive load. Females learned best from both real and imitative practice, and males performed better than females without practice.

Furthermore, gender differences were found in Experiment 2, where females were found to learn better and more efficiently using both real and imitative practice, with no differences for males. When comparing gender within practice types, females performed significantly better than males with imitative practice, and males performed significantly better than females with no practice.

Regarding the open question investigating differences between the two practice conditions, mixed results were found, with real practice leading to better learning for the paper folding task overall. Imitative practice led to better learning efficiency for piano playing and better performance for females compared to males for the paper folding task.

It appears that the type of practice that leads to better learning and efficiency is moderated by task complexity and gender-related effects.

While the interaction between culture and cognition has evolved significantly [61,62], making it a worthwhile dimension to consider when conducting this type of research, it was not the focus of these studies. Rather, we collected our data in a large western university. However, it is worth noting that our study did include participants of various nationalities, including some from Asia, Europe, and North America, which may reduce the potential cultural biases of the data.

4.1. Theoretical and Instructional Implications

4.1.1. Embodied Gesturing and Practice

Consistent with past findings relating to instructional animations and embodied cognition-related effects [3,29,63], the current study found that practice with instructional animations contributes to better learning. Evidence emerged that participants in the real practice group (Experiment 2) and the imitative practice group (Experiment 1) demonstrated an enhanced ability to replicate the learned tasks compared to participants in the no practice condition, who solely relied on observational learning. However, in both experiments, overall performance scores for real practice were not found to differ significantly from imitative practice. Additionally, evidence emerged that the effectiveness of the type of practice depended on the learning task.

In Experiment 1, audio feedback during real practice was omitted to prevent interference with the animated video audio channel, potentially impacting the degree of task replication for piano playing. In fact, it may partially explain why imitative practice was superior to real practice for the difficult tasks, where the tactile feedback may have been redundant [64]. Moreover, for less experienced learners, the tactile feedback perhaps induced split attention, similar to Cerpa et al. [65], who found that, when learning a computer application, the introduction of physical practice too soon in a complex task can cause a split attention-type of effect, where the focus is shifted to the physical apparatus and thus reduces cognitive capacity for more cognitive aspects of the task [64]. In Experiment 2, which involved no audio channel, participants likely perceived real practice to be more meaningful than imitative practice because it replicated the task exactly as explained in the video. Overall, the findings align with past research, demonstrating the effectiveness of practice in enhancing learning across diverse domains, including sports and music [37,38].

4.1.2. Spatial Ability and Gender

In Experiment 2, females were found to perform better with practice than with no practice (with a significant improvement in the performance and efficiency scores for both real and imitative practice), while there were no differences among practice types for males. When comparing genders for each practice type, males were found to perform significantly better with no practice, whereas females were found to perform significantly better with imitative practice. Wong et al. [12] found a similar gender interaction effect in a Lego-based learning task that incorporated practice during learning through instructional animations. Based on these results, it can be suggested that the absence of embodied practice may hinder the learning of females, while the inclusion of practice benefits females more than males. This implies that there may be underlying mechanisms that contribute to differences in how males and females learn, potentially related to greater mirror neuron activity in females [51–55,66].

Wong et al. [12] also tested for the correlation between spatial ability and gender; however, they were not able to attribute the significant gender interactions to spatial ability, since a significant correlation effect was not found. Halpern and Collaer [67] also found that paper folding tests of spatial ability did not demonstrate a link with gender differences. In these studies, it is possible that the relevant spatial ability construct may not have been captured accurately, using the most appropriate measurements. The Mental Rotations Test [68] is a common validated measure of spatial ability [46]. It can be explored in

combination with the Punched Holes Test and other validated measures of spatial ability. Given that this is the case and spatial ability is involved, a superior but unmeasured aspect of spatial ability could still be contributing to consistent performance for male participants across all three groups. The lack of some aspect of spatial ability in females may be introducing significant differences in the performance of females among the groups. It could also explain why females are disadvantaged when learning with observational learning only and may explain why embodied practice may help make up for the lack of spatial processing that may be required when only observing an animated video. The inclusion of new and different measures of spatial ability in future studies could thus be relevant, in order to try and ascertain whether different and yet unmeasured aspects of spatial ability may be contributing to gender differences.

Learning how to complete a movement-based task with the incorporation of practice may involve a variety of mechanisms that contribute to gender differences [69]. In particular, females have been found to be significantly disadvantaged when it comes to visuospatial memory [70]. Lower visuospatial memory is compensated for by the use of more gestures [71,72]. Hence, it is likely that females engaged in more embodied practice, leading to better results.

4.2. Limitations and Recommendations for Future Research

Experiment 1 was conducted on university-aged students, and it would be useful to explore if results can be replicated within other age groups, including children. Future studies could test different age groups using similar materials to examine whether similar results are seen across different age groups. While the focus was on different levels of complexity of instructional content and gender was not considered, it would also be useful to explore gender differences in these types of materials as well.

In Experiment 2, it would also be useful to explore if these results persist with different age groups. Moreover, it is possible that the spatial ability construct may have not been fully captured through the Punched Holes Test. The Mental Rotation Test is a spatial ability test that has been found to have the largest difference between scores for females and males [45,46]. Utilizing a range of spatial ability tests, including the Mental Rotation Test, may be needed to reflect the spatial ability differences more accurately between females and males. Piloting the correlation between the scores on a range of spatial ability tests and performance may be helpful in determining the type of spatial ability test to utilize in follow-up studies.

It is interesting to consider whether the advantages of embodied practice while learning human motor tasks from animations may persist over time. Cook et al. [73] found similar results from gesturing and non-gesturing groups, with improved performance only emerging post test. The present work could be expanded to include a delayed post test to assess whether the same effects persist long-term.

Although some interesting trends have been observed about gender when it comes to learning through instructional animations [12], there is not a lot of research to understand the reasons for these gender differences. Gupta et al. [69] found differences in instructional support needed for learning from instructional animations across genders, with females needing the extra support and males learning best from observational learning only, as well as finding that including spatial ability as a covariate impacted results. The second experiment has been successful in suggesting a significant interaction between gender and learning content design, as well as potential impacts from spatial ability. It is important to note that the gender effects found were based on a smaller sample size and should be validated in future studies. This paves the way for more work to be done in various domains to further investigate these interactions and the reasons for these differences. It is possible that, in some contexts, females need more instructional support when learning from animations than males, and this is worth exploring in the future with different learning tasks.

Our study was conducted in a large western university with many international students and, while it may be problematic to generalize these results to all other populations, we note that our student population was multi-cultural. However, like most research conducted within one country, it may not always generalize to other cultures. Future studies could more specifically address differences in cultural and socioeconomic contexts impacting cognition and learning in non-western environments, by collecting data from a different student population, thereby better considering possible cultural influences on learning [62]. Collecting data from different socioeconomic and cultural contexts would thus further enhance the applicability and generalizability of these findings [74,75].

4.3. Conclusions

The efficacy of animations as an instructional format is often undermined by the transience of information presented, as well as the limitations of working memory. However, some of these limitations are reduced due to our innate ability to learn by observing human movement effects. The increasing accessibility of animated instructional resources in an e-learning setting necessitates the need for more research into techniques to increase their effectiveness. The present study investigated the impact of embodied practice on learning from video-based instructional animations in human movement-related domains through two different experiments and found some clear benefits of embodied practice in different learning domains and contexts. Results suggest that practice can support learning, but whether it needs to be real or imitative will depend on both the type of instructional content as well as the knowledge of the learners, with possible gender effects in support of females benefiting most from practice. The results of this study may be of value to educators, as they provide guidance on when and how to incorporate practice into learning from instructional animations. The results suggest that the type of practice selected should be based on both the learning domain and learner characteristics. Educators can thus use these insights to make informed decisions that facilitate learning outcomes.

Author Contributions: Conceptualization: H.B. (Experiment 1), M.N.M. (Experiment 2), N.M., and P.A.; Methodology: H.B. and M.N.M.; Formal analysis: H.B., M.N.M., N.M., and P.A.; Investigation: H.B. and M.N.M.; Writing—original draft preparation: M.N.M.; Writing—review and editing: N.M. and P.A.; Supervision: N.M. and P.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of University of New South Wales (HC16138, March 2016 to March 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data can be obtained by contacting the first and second authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. de Koning, B.B.; Marcus, N.; Brucker, B.; Ayres, P. Does observing hand actions in animations and static graphics differentially affect learning of hand-manipulative tasks? *Comput. Educ.* **2019**, *141*, 103636. [CrossRef]
2. Koning, B.B.; Mok, K.; Marcus, N.; Ayres, P. Investigating the role of hand perspective in learning from procedural animations. *Br. J. Educ. Psychol.* **2022**, *93*, 251–269. [CrossRef] [PubMed]
3. Lajevardi, N.; Narang, N.S.; Marcus, N.; Ayres, P. Can mimicking gestures facilitate learning from instructional animations and static graphics? *Comput. Educ.* **2017**, *110*, 64–76. [CrossRef]
4. de Stefani, E.; Rodà, F.; Volta, E.; Pincolini, V.; Farnese, A.; Rossetti, S.; Pedretti, F.; Ferrari, P.F. Learning new sport actions: Pilot study to investigate the imitative and the verbal instructive teaching methods in motor education. *PLoS ONE* **2020**, *15*, e0237697. [CrossRef] [PubMed]
5. McBride, E.R.; Rothstein, A.L. Mental and Physical Practice and the Learning and Retention of Open and Closed Skills. *Percept. Mot. Ski.* **1979**, *49*, 359–365. [CrossRef]

6. Garland, T.B.; Sanchez, C.A. Rotational perspective and learning procedural tasks from dynamic media. *Comput. Educ.* **2013**, *69*, 31–37. [CrossRef]
7. Mierowsky, R.; Marcus, N.; Ayres, P. Using mimicking gestures to improve observational learning from instructional videos. *Educ. Psychol.* **2019**, *40*, 550–569. [CrossRef]
8. Wong, A.; Leahy, W.; Marcus, N.; Sweller, J. Cognitive load theory, the transient information effect and e-learning. *Learn. Instr.* **2012**, *22*, 449–457. [CrossRef]
9. Höffler, T.N.; Leutner, D. Instructional animation versus static pictures: A meta-analysis. *Learn. Instr.* **2007**, *17*, 722–738. [CrossRef]
10. Ayres, P.; Marcus, N.; Chan, C.; Qian, N. Learning hand manipulative tasks: When instructional animations are superior to equivalent static representations. *Comput. Hum. Behav.* **2009**, *25*, 348–353. [CrossRef]
11. Arguel, A.; Jamet, E. Using video and static pictures to improve learning of procedural contents. *Comput. Hum. Behav.* **2009**, *25*, 354–359. [CrossRef]
12. Wong, M.; Castro-Alonso, J.C.; Ayres, P.; Paas, F. Gender Effects When Learning Manipulative Tasks From Instructional Animations and Static Presentations. *Educ. Technol. Soc.* **2015**, *18*, 1176–1647.
13. Wong, A.; Marcus, N.; Ayres, P.; Smith, L.; Cooper, G.A.; Paas, F.; Sweller, J. Instructional animations can be superior to statics when learning human motor skills. *Comput. Hum. Behav.* **2009**, *25*, 339–347. [CrossRef]
14. Kalyuga, S. Cognitive Load Theory: How Many Types of Load Does It Really Need? *Educ. Psychol. Rev.* **2011**, *23*, 1–19. [CrossRef]
15. Iacoboni, M.; Woods, R.P.; Brass, M.; Bekkering, H.; Mazziotta, J.C.; Rizzolatti, G. Cortical mechanisms of human imitation. *Science* **1999**, *286*, 2526–2528. [CrossRef] [PubMed]
16. Rizzolatti, G.; Craighero, L. The mirror-neuron system. *Annu. Rev. Neurosci.* **2004**, *27*, 169–192. [CrossRef]
17. Higuchi, S.; Holle, H.; Roberts, N.; Eickhoff, S.B.; Vogt, S. Imitation and observational learning of hand actions: Prefrontal involvement and connectivity. *NeuroImage* **2012**, *59*, 1668–1683. [CrossRef]
18. Paas, F.; Sweller, J. An Evolutionary Upgrade of Cognitive Load Theory: Using the Human Motor System and Collaboration to Support the Learning of Complex Cognitive Tasks. *Educ. Psychol. Rev.* **2012**, *24*, 27–45. [CrossRef]
19. Chomsky, N. On the Biological Basis of Language Capacities. In *The Neuropsychology of Language*; Springer: Boston, MA, USA, 1976; pp. 1–24. [CrossRef]
20. Geary, D.C. Principles of evolutionary educational psychology. *Learn. Individ. Differ.* **2002**, *12*, 317–345. [CrossRef]
21. Giese, M.A.; Poggio, T. Neural mechanisms for the recognition of biological movements. *Nat. Rev. Neurosci.* **2003**, *4*, 179–192. [CrossRef]
22. Hasler, B.S.; Kersten, B.; Sweller, J. Learner Control, Cognitive Load and Instructional Animation. *Appl. Cognit. Psychol.* **2007**, *21*, 713–729. [CrossRef]
23. Sweller, J. Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educ. Psychol. Rev.* **2010**, *22*, 123–138. [CrossRef]
24. Mayer, R.E.; Chandler, P. When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *J. Educ. Psychol.* **2001**, *93*, 390–397. [CrossRef]
25. Barsalou, L.W. Grounded Cognition: Past, Present, and Future. *Top. Cogn. Sci.* **2010**, *2*, 716–724. [CrossRef] [PubMed]
26. Mavilidi, M.F.; Okely, A.D.; Chandler, P.; Paas, F. Effects of Integrating Physical Activities Into a Science Lesson on Preschool Children’s Learning and Enjoyment. *Appl. Cogn. Psychol.* **2017**, *31*, 281–290. [CrossRef]
27. Ferguson, E.L.; Hegarty, M. Learning with Real Machines or Diagrams: Application of Knowledge to Real-World Problems. *Cogn. Instr.* **1995**, *13*, 129–160. [CrossRef]
28. Chettaoui, N.; Atia, A.; Bouhlel, M.S. Examining the Effects of Embodied Interaction Modalities on Students’ Retention Skills in a Real Classroom Context. *J. Comput. Educ.* **2022**, *9*, 549–569. [CrossRef]
29. Goldin-Meadow, S.; Nusbaum, H.; Kelly, S.D.; Wagner, S. Explaining math: Gesturing lightens the load. *Psychol. Sci.* **2001**, *12*, 516–522. [CrossRef]
30. Valenzano, L.; Alibali, M.W.; Klatzky, R. Teachers’ gestures facilitate students’ learning: A lesson in symmetry. *Contemp. Educ. Psychol.* **2003**, *28*, 187–204. [CrossRef]
31. Cook, S.W.; Friedman, H.S.; Duggan, K.A.; Cui, J.; Popescu, V. Hand Gesture and Mathematics Learning: Lessons From an Avatar. *Cogn. Sci.* **2017**, *41*, 518–535. [CrossRef]
32. Darden, G.F. Demonstrating Motor Skills-Rethinking that Expert Demonstration. *J. Phys. Educ. Recreat. Danc.* **1997**, *68*, 31–35. [CrossRef]
33. De Nooijer, J.A.; Van Gog, T.; Paas, F.; Zwaan, R.A. Effects of imitating gestures during encoding or during retrieval of novel verbs on children’s test performance. *Acta Psychol.* **2013**, *144*, 173–179. [CrossRef] [PubMed]
34. LeBarton, E.S.; Goldin-Meadow, S.; Raudenbush, S. Experimentally-induced Increases in Early Gesture Lead to Increases in Spoken Vocabulary. *J. Cogn. Dev. Off. J. Cogn. Dev. Soc.* **2015**, *16*, 199. [CrossRef] [PubMed]
35. Brown, S.W.; Bennett, E.D. The role of practice and automaticity in temporal and nontemporal dual-task performance. *Psychol. Res.* **2002**, *66*, 80–89. [CrossRef]
36. Campitelli, G.; Gobet, F. Deliberate Practice. *Curr. Dir. Psychol. Sci.* **2011**, *20*, 280–285. [CrossRef]
37. De Bruin, A.B.; Smits, N.; Rikers, R.M.; Schmidt, H.G. Deliberate practice predicts performance over time in adolescent chess players and drop-outs: A linear mixed models analysis. *Br. J. Psychol.* **2008**, *99*, 473–497. [CrossRef]

38. Keith, N.; Ericsson, K.A. A Deliberate Practice Account of Typing Proficiency in Everyday Typists. *J. Exp. Psychol. Appl.* **2007**, *13*, 135–145. [CrossRef]
39. Platz, F.; Kopiez, R.; Lehmann, A.C.; Wolf, A. The influence of deliberate practice on musical achievement: A meta-analysis. *Front. Psychol.* **2014**, *5*, 646. [CrossRef]
40. Paas, F.G.; Van Merriënboer, J.J. The Efficiency of Instructional Conditions: An Approach to Combine Mental Effort and Performance Measures. *Hum. Factors* **1993**, *35*, 737–743. [CrossRef]
41. Paas, F.G. Training Strategies for Attaining Transfer of Problem-Solving Skill in Statistics: A Cognitive-Load Approach. *J. Educ. Psychol.* **1992**, *84*, 429–434. [CrossRef]
42. Post, O.; Toussaint, G. The Edit Distance as a Measure of Perceived Rhythmic Similarity. *Empir. Musicol. Rev.* **2011**, *6*, 164–179. [CrossRef]
43. Höffler, T.N. Spatial ability: Its influence on learning with visualizations—A meta-analytic review. *Educ. Psychol. Rev.* **2010**, *22*, 245–269. [CrossRef]
44. Castro-Alonso, J.C.; Wong, M.; Adesope, O.O.; Ayres, P.; Paas, F. Gender Imbalance in Instructional Dynamic Versus Static Visualizations: A Meta-analysis. *Educ. Psychol. Rev.* **2019**, *31*, 361–387. [CrossRef]
45. Linn, M.C.; Petersen, A.C. Emergence and Characterization of Sex Differences in Spatial Ability: A Meta-Analysis. *Child Dev.* **1985**, *56*, 1479. [CrossRef]
46. Zell, E.; Krizan, Z.; Teeter, S.R. Evaluating gender similarities and differences using metasynthesis. *Am. Psychol.* **2015**, *70*, 10–20. [CrossRef]
47. Voyer, D.; Voyer, S.; Bryden, M.P. Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychol. Bull.* **1995**, *117*, 250–270. [CrossRef]
48. Uttal, D.H.; Meadow, N.G.; Tipton, E.; Hand, L.L.; Alden, A.R.; Warren, C.; Newcombe, N.S. The malleability of spatial skills: A meta-analysis of training studies. *Psychol. Bull.* **2013**, *139*, 352–402. [CrossRef]
49. Geer, E.A.; Ganley, C. Sex differences in social and spatial perspective taking: A replication and extension of Tarampi et al. (2016). *Q. J. Exp. Psychol.* **2023**, *76*, 93–108. [CrossRef]
50. Acharya, S.; Shukla, S. Mirror neurons: Enigma of the metaphysical modular brain. *J. Nat. Sci. Biol. Med.* **2012**, *3*, 118–124. [CrossRef]
51. Cheng, Y.W.; Tzeng, O.J.; Decety, J.; Imada, T.; Hsieh, J.C. Gender differences in the human mirror system: A magnetoencephalography study. *Neuroreport* **2006**, *17*, 1115–1119. [CrossRef]
52. Cheng, Y.; Chou, K.H.; Decety, J.; Chen, I.Y.; Hung, D.; Tzeng, O.J.; Lin, C.P. Sex differences in the neuroanatomy of human mirror-neuron system: A voxel-based morphometric investigation. *Neuroscience* **2009**, *158*, 713–720. [CrossRef] [PubMed]
53. Grabowska, A. Sex on the brain: Are gender-dependent structural and functional differences associated with behavior? *J. Neurosci. Res.* **2016**, *95*, 200–212. [CrossRef]
54. Hobson, H.M.; Bishop, D.V. The interpretation of mu suppression as an index of mirror neuron activity: Past, present and future. *R. Soc. Open Sci.* **2017**, *4*. [CrossRef]
55. Schulte-Rüther, M.; Markowitsch, H.J.; Fink, G.R.; Piefke, M. Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: A functional magnetic resonance imaging approach to empathy. *J. Cogn. Neurosci.* **2007**, *19*, 1354–1372. [CrossRef]
56. Kirschner, P.A. Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learn. Instr.* **2002**, *12*, 1–10. [CrossRef]
57. DeLeeuw, K.E.; Mayer, R.E. A Comparison of Three Measures of Cognitive Load: Evidence for Separable Measures of Intrinsic, Extraneous, and Germane Load. *J. Educ. Psychol.* **2008**, *100*, 223–234. [CrossRef]
58. Ekstrom, R.; French, J.; Harman, H. *Manual for Kit of Factor-Referenced Cognitive Tests*; Educational Testing Service: Princeton, NJ, USA, 1976.
59. Kane, M.J.; Hambrick, D.Z.; Conway, A.R. Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle. *Psychol. Bull.* **2005**, *131*, 66–71. [CrossRef]
60. Marshalek, B.; Lohman, D.F.; Snow, R.E. The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence* **1983**, *7*, 107–127. [CrossRef]
61. Carruthers, M.J. *The Book of Memory: A Study of Memory in Medieval Culture*; Cambridge University Press: Cambridge, UK, 2008.
62. Hofstede, G. *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*; Sage: Thousand Oaks, CA, USA, 2001. [CrossRef]
63. Fiorella, L.; Mayer, R.E. Effects of Observing the Instructor Draw Diagrams on Learning from Multimedia Messages. *J. Educ. Psychol.* **2016**, *108*, 528–546. [CrossRef]
64. Sweller, J.; Ayres, P.; Kalyuga, S. *Cognitive Load Theory*; Springer: New York, NY, USA, 2011. [CrossRef]
65. Cerpa, N.; Chandler, P.; Sweller, J. Some Conditions under Which Integrated Computer-Based Training Software Can Facilitate Learning. *J. Educ. Comput. Res.* **1996**, *15*, 345–367. [CrossRef]
66. Cheng, Y.; Lee, P.L.; Yang, C.Y.; Lin, C.P.; Hung, D.; Decety, J. Gender Differences in the Mu Rhythm of the Human Mirror-Neuron System. *PLoS ONE* **2008**, *3*, e2113. [CrossRef] [PubMed]
67. Halpern, D.F.; Collaer, M.L. Sex Differences in Visuospatial Abilities. In *The Cambridge Handbook of Visuospatial Thinking*; Cambridge University Press: Cambridge, UK, 2005; pp. 170–212. [CrossRef]

68. Shepard, R.N.; Metzler, J. Mental rotation of three-dimensional objects. *Science* **1971**, *171*, 701–703. [CrossRef]
69. Gupta, R.; Marcus, N.; Ayres, P. Investigating the impact of gender-differences and spatial ability on learning from instructional animations. *Lannee Psychol.* **2022**, *122*, 537–561. [CrossRef]
70. Voyer, D.; Voyer, S.D.; Saint-Aubin, J. Sex differences in visual-spatial working memory: A meta-analysis. *Psychon. Bull. Rev.* **2017**, *24*, 307–334. [CrossRef]
71. Gillespie, M.; James, A.N.; Federmeier, K.D.; Watson, D.G. Verbal working memory predicts co-speech gesture: Evidence from individual differences. *Cognition* **2014**, *132*, 174–180. [CrossRef] [PubMed]
72. Pouw, W.T.; Mavilidi, M.F.; van Gog, T.; Paas, F. Gesturing during mental problem solving reduces eye movements, especially for individuals with lower visual working memory capacity. *Cogn. Process.* **2016**, *17*, 269–277. [CrossRef]
73. Cook, S.W.; Mitchell, Z.; Goldin-Meadow, S. Gesturing makes learning last. *Cognition* **2008**, *106*, 1047–1058. [CrossRef]
74. Cole, M. *Cultural Psychology: A Once and Future Discipline*; Harvard University Press: Cambridge, MA, USA, 1996.
75. Li, J. U.S. and Chinese cultural beliefs about learning. *J. Educ. Psychol.* **2003**, *95*, 258–267. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

A Cognitive Load Approach to Molecular Geometries: Augmented Reality Technology and Visuospatial Abilities in Chemistry

Conor Desmond Kenneally and Brendan Bentley *

School of Education, University of Adelaide, Adelaide 5005, Australia; conor.kenneally@adelaide.edu.au

* Correspondence: brendan.bentley@adelaide.edu.au

Abstract: Within chemistry education, methods for effectively teaching students the three-dimensional spatial arrangements of matter at the molecular level remains a topical issue. As a form of geometric problem solving, it requires learners to apply mental rotation abilities as an evolved visuospatial skill to obtain subject-specific content knowledge. Recent research into the use of Cognitive Load Theory (CLT) as a framework for instructional design in conjunction with augmented reality (AR) technology as a learning tool has begun to show promise in reducing unnecessary cognitive activity to improve learning. Yet, broader conclusions remain inconclusive, especially within the context of a learner's mental rotation abilities. This study investigated the relationship between these factors by collecting data using a 2×3 experimental design that divided a sample of Year 10 students ($n = 42$) into two groups. The intervention group ($n = 24$) used mobile devices utilising AR technology with instructional 3D molecular geometry content featuring design principles based on CLT to encourage hand movements to rotate three-dimensional molecular structures. The non-AR-based control group ($n = 18$) was taught using traditional methods. Analysis of the data revealed participants using AR technology that featured CLT design principles experienced less cognitive load and improved achievement in post-testing compared to those taught using traditional methods, suggesting under certain conditions, the use of hand movement applied to AR design material improves learning.

Citation: Kenneally, C.D.; Bentley, B.A Cognitive Load Approach to Molecular Geometries: Augmented Reality Technology and Visuospatial Abilities in Chemistry. *Educ. Sci.* **2024**, *14*, 1036.<https://doi.org/10.3390/educsci14091036>

Academic Editors: Fred Paas, Andre Tricot, Kim Ouwehand and Florence Lespiau

Received: 24 July 2024

Revised: 17 September 2024

Accepted: 17 September 2024

Published: 23 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: cognitive load theory; augmented reality; mental rotations; molecular geometry

1. Introduction

Human cognition, similar to other physiological traits, can be understood as a result of evolutionary processes, driven by natural selection to improve survival and reproductive fitness [1]. The evolution of innate cognitive knowledge that is essential to survival is known within the discipline of evolutionary educational psychology as biological primary knowledge [2]. Yet, the modern world requires the learning of much more complex and abstract culturally specific biological secondary knowledge. This is knowledge that lacks the same evolutionary impetus and requires conscious effort often facilitated by educational institutions [3]. Cognitive Load Theory (CLT) utilises this understanding of evolved human cognitive processes to develop a theory of instructional design underpinned by these evolutionary implications for human learning [4]. CLT informs the manner in which schemas are acquired and transferred from a learner's working memory (WM) to their long-term memory (LTM) during complex cognitive tasks [5]. The WM's capacity and duration plays a pivotal role in the efficacy of this process. When the WM's limited capacity is overloaded and the available cognitive resources exceed the WM's maximum cognitive capacity, the efficiency of schema transfer is inhibited and meaningful learning is prevented [6].

Within the context of educational institutions, the teaching of chemistry as a biological secondary knowledge is a subject that imposes high WM demands. Its challenge is somewhat unique amongst other subjects due to the multiple levels of representation

and interrelating levels of thought inherent to its content, requiring a learner to consider phenomena between macro, sub-micro, and representational levels of representation [7]. As such, cognitive limitations can quickly be exceeded through the restrictions of WM demands when a learner is expected to process such a volume of information [8]. Recent advances in the use of augmented reality (AR) technology to aid mental rotation and to reduce cognitive load and improve student-learning outcomes in chemistry have shown promise. AR technology involves imposing computer-generated virtual images onto a real environment, most often via an application utilising a phone's camera [9]. Within the context of the chemistry classroom and studying molecular geometries, a number of studies have suggested its benefit for optimising learner cognitive load by imposing cognitive load effects [10–12]. By substituting virtual objects onto real environments, AR technology can assist in directing and guiding the learner by using visual triggers and geographical information [13] to potentially reduce the number of elements required for processing by the WM.

The effect of AR technology to enhance the mental rotation ability of secondary chemistry students studying molecular geometries and the perceived task difficulty of utilising AR sympathetic technology to reduce cognitive load remains unresolved, with only a limited amount of research having been undertaken into this relationship [11]. By investigating how the implementation of this technology within an educational context can be optimised, a potential barrier to achievement in chemistry could be reduced [14]. This paper seeks to inform the current field of study by interrogating the following research question: What effect does AR technology with molecular geometric content designed using CLT have on secondary-science students' mental rotation abilities? Two further questions were also posed: What effect does mental rotation ability have on cognitive load? And can cognitive load and achievement be altered using AR technology when students engage in mental rotation?

1.1. Cognitive Load Theory and Visuospatial Abilities

Both the LTM and WM are considered essential characteristics of human cognitive architecture and fundamental to the acquisition of schemas. The LTM, as an information processing system, is elementary in determining the bulk of all human cognitive activity, highlighting its tremendous potential for total capacity [15]. In contrast, the WM is much more restricted in the amount of information that it can process at any given time, yet remains the main structure responsible for processing new information [16]. While the WM is limited when dealing with novel information, it is essentially limitless when dealing with biological secondary knowledge once it exists in the LTM [17]. Novel information that is to be acquired requires the WM to allocate significant resources to process and transfer information to the LTM. If the processing and transfer of the novel information exceeds the WM capacity, the WM will be overloaded and hinder effective learning [18]. As learners develop and increase their pre-existing cognitive schemas, the number of elements are able to be chunked and treated as one unit, as opposed to many singular or separate elements [19]. CLT provides the conceptual framework to design instructional procedures to reduce WM load and to enable domain-specific schemas to be more efficiently transferred to the LTM, allowing learners to function more effectively in a particular learning environment [20].

The adaptation of the human eye enables the WM to process and transfer visuospatial information to the LTM and takes advantage of both the WM and LTM functions. The transfer of visuospatial information from the WM that is to be retained by the LTM is the cognitive process of spatial learning, a crucial skill integral to the species survival and reproduction [21]. The evolution of the human brain has enabled the capacity for visuospatial processing, a cognitive process defined as the WM's ability to generate and transform visual and spatial information [22]. Geary and Berch [2] describe this visuospatial processing as a type of biological primary knowledge where different visuospatial abilities are responsible for controlling a diverse range of cognitive abilities [3]. Mental rotation

is considered as a specific type of visuospatial skill that involves the ability to transform an object by moving it through an imagined space, an ability that can be determined by how accurately and rapidly one is able to mentally move an object, e.g., by rotating it [23]. Whilst visuospatial abilities embody biological primary knowledge, much variation can still be observed in ability levels within a population [14]. Mental rotation, specifically, has been observed as particularly variable between individuals [24]. Yet, research suggests visuospatial ability is able to be influenced positively by educational interventions, implying the potential malleability that it has as a skill [25].

1.2. Mental Rotation and Molecular Geometries

Many aspects of content taught in modern educational institutions, such as geometric problem solving, require the processing of visuospatial information through mental rotation [26]. Geometric problems require learners to perform transformations of objects in order to analyse both the shape and orientation of 3D objects, in addition to translating this understanding between 2D and 3D representations [27]. The consequence of this is a high correlation between a learner's mental rotation ability level and achievement in mathematics being readily observed [28]. Processing complex mental rotation problems requires a learner to store, maintain, and interpret complex representations of visuospatial information using the limited resources of the visuospatial sketchpad, a feat less likely to overload those who have higher spatial ability capacity [29]. Chemistry, as the study of matter and how matter interacts, requires the navigation and manipulation of molecular geometries on a scale both unobservable and difficult to interact directly with. Mental rotation skills permeate numerous topics within a chemistry curriculum at all levels, requiring learners to use 2D diagrams to interpret 3D spatial arrangements [30]. As with geometry, a student's ability to understand and manipulate these 3D structures is influenced by their mental rotation visuospatial ability [11], ultimately inferring their ability to learn from visual representations [31].

1.3. Working Memory and Executive Function

The function of the WM to deal with mental rotation activity associated with 3D chemical structures can be optimised utilising CLT principles [16]. The WM can be considered to comprise of two levels of functioning composed of three components: a higher level consisting of a central executive structure responsible for controlling two lower-level storage components, the phonological loop, and the visuospatial sketchpad (refer to Figure 1) [3].

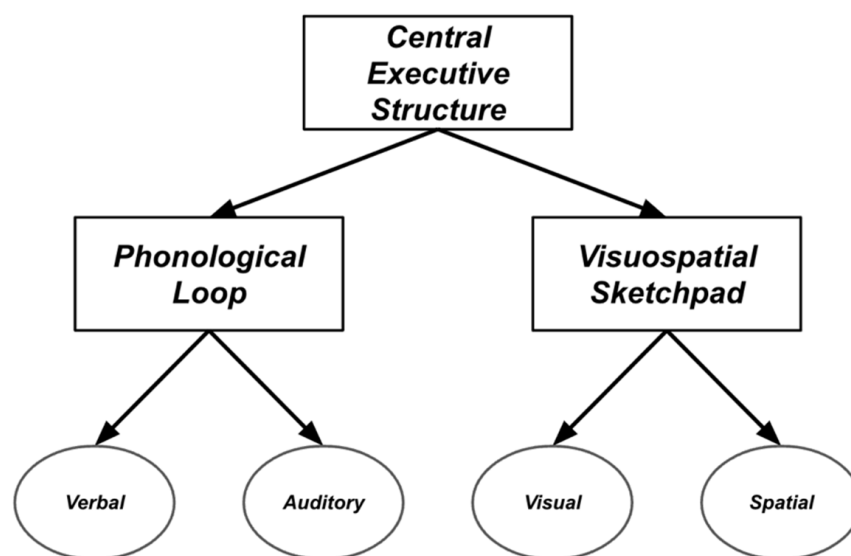


Figure 1. Components of WM.

Each component is responsible for processing differing types of information, with the phonological loop responsible for processing verbal and auditory information, whereas the visuospatial sketchpad is responsible for processing visual and spatial information [3]. The cognitive theory of multimedia learning (CTML) considers multiple representations of external information. The processing of dual and multiple information can be explained through dual coding theory, wherein the construction of meaning and knowledge through modality-specific mental representations arises from the two systems [32]. The effect of multiple sources of information requiring simultaneous processing by the limited resources available to the WM impacts the efficiency of schema transfer from the WM to the LTM and, thus, problem-solving capacity. To alleviate the limitations of the WM capacity and maximise visuospatial processing, Castro-Alonso, Ayres, and Sweller [16] propose that CLT and CTML principles can be utilised to optimise the visuospatial processing of the learner by guiding task design to reduce any unnecessary visuospatial processing a learner needs to perform. The adoption of CLT and CTML principles to improve visuospatial processing is, in part, supported by the malleability associated with spatial abilities and the potential that educational intervention can have in improving these types of skills [25], where one type of spatial skill can additionally transfer improvement to another type of spatial skill [33].

1.4. Augmented Reality Technology and Cognitive Load Effects

AR technology has been identified as an emerging technology exhibiting promising results in improving visuospatial processing by reducing unnecessary cognitive activity [11,34,35]. Within the context of studying molecular chemistry in the classroom, a number of studies have suggested AR may benefit learning by adopting CLT principles [10,12] through its imposition of computer-generated virtual images onto a real environment [9]. It is suggested that AR technology can assist in directing and guiding a learner's visual trigger and geographical information [13] and can signal important geospatial information to assist visuospatial processing to understand and visualise abstract concepts and complex spatial relationships [35]. The application of AR to recall or support the triggering of past schemas can enable the chunking together of separate elements so they can be treated as one singular unit, known as the signalling principle. This strategy encourages the reduction of cognitive load and can counter the effects of element interactivity by incorporating visual cues that signal only essential learning elements [36], reducing the amount of novel information in the visuospatial processor to optimise the visualisation process [22]. Comparable studies have rationalised AR's potential to promote a lower investment of cognitive effort required when manipulating molecular geometries by reducing the demands on visuospatial processing [11].

Instructional research involving human movements has also demonstrated the effectiveness of using hand gestures and object manipulation to facilitate learning across various subjects [37–39]. Researchers have found gesturing and object manipulation to benefit tasks such as solving math problems [40] and understanding complex science concepts [38], with De Koning and Tabbers [41] providing a comprehensive review further supporting the positive impact that these instructional approaches can have. While it is suggested actions such as gesturing and tracing have origins in biological primary knowledge, when they are applied to learning biological secondary knowledge, these types of actions may support cognition to positively influence a learner's experienced level of cognitive load [42,43]. AR-based learning readily embodies forms of gesturing and tracing by using hand gestures and body movements associated with the human movement effect. The effect suggests that certain movements used to support learning are automated and, therefore, have little impact on the use of the WM's resources. The technology adopts gesturing and tracing by using hand gestures and body movements to create and modify 3D digital structures, allowing the creation and manipulation of molecular geometries [10]. The success of gesturing and tracing on academic achievement within the context of visuospatial processing has been demonstrated to correspond with a learner's existing visuospatial ability levels,

with research suggesting its potential to be particularly effective in improving learning outcomes for lower visuospatial-ability learners [44].

Whilst some evidence exists suggesting the potential for AR technology to reduce the cognitive load demands of visuospatial processing during geometric manipulation by improving conceptual understanding through enhancing visualisation [45], broader conclusions remain inconsistent [36]. Some technology has been shown to have no significant effects in reducing cognitive load when implementing AR in the learning of molecular geometries, with certain cases suggesting it may actively be hindering the process [36,46,47]. Although some literature has explored the use of AR as a classroom tool to instruct molecular geometries, further research is required to strengthen and clarify the conclusions that can be drawn regarding its effectiveness as a learning tool. To further inform the field of study, additional research is required to better understand the effect optimised CLT science visualisations have on visuospatial processing skills, particularly in relation to the signalling principle [16]. This study aimed to achieve this by investigating the relationship between cognitive load and achievement in post-testing when utilising an AR-based learning intervention to support student mental rotation abilities in understanding three-dimensional molecular structure.

2. Method

2.1. Participants and Procedure

This study was conducted with an original group of 49 Year 10 (15- to 16-year-old) students from an independent Australian secondary school. The number of participants was reduced to 42 due to student non-attendance during the post-testing phase of the study. A 2×3 experimental method was used, with participants randomly placed into the non-AR control (C) group ($n = 18$) and AR-based intervention (I) group ($n = 24$). Both groups received 45 min instruction during a classroom lesson wherein the C group was provided with traditional teaching methods and the I group received instruction that utilised AR technology. Achievement and corresponding cognitive load data were collected at three times: pre-test (T1), immediate post-test (T3), and delayed post-test (T4). T3 data were collected one day after treatment, and T4 data were collected seven days after treatment (refer to Table 1). The 45 min classroom lesson (T2) was implemented after the pre-testing and was dedicated solely to the delivery of content; no data were collected during this session. All sessions with the I and C groups were conducted separately from each other, taking place at different times. Each session was undertaken between August and September 2023 and was implemented independently of participants' schooling. Ethical guidelines and regulations were carefully considered to ensure the safety and feasibility of participants, with ethical clearance being granted by the University of Adelaide (approval number: H-2023-086). Participation was voluntary, with informed consent collected from both the participants and their guardians.

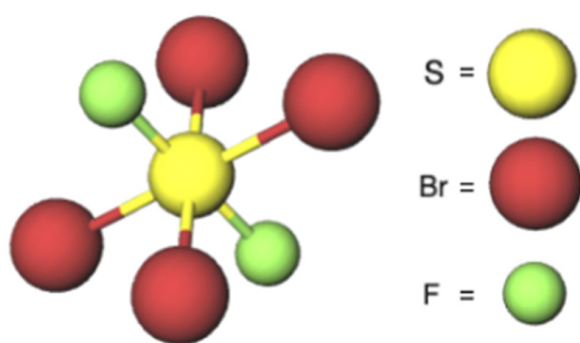
Table 1. Experimental methodology.

Group	T1 Pre-Test	T2 Classroom Lesson	T3 Immediate Post-Test	T4 Delayed Post-Test
Intervention (I)	6 content questions	Lesson taught using booklet with digital AR prompts accompanying molecular geometries	6 isomorphic content questions	6 isomorphic content questions
Control (C)	6 content questions	Lesson taught using booklet with diagrams accompanying molecular geometries	6 isomorphic content questions	6 isomorphic content questions

2.2. Materials

The pre-test assessed participants' prior content knowledge and associated levels of cognitive load. The six test questions were contained in a test booklet, with each participant given 10 min to complete the test for a total of 14 marks. The pre-test collected participants' baseline prior knowledge and cognitive load data. Cognitive load was measured using self-reporting methods by participants responding to the perceived difficulty based on a 9-point Likert scale from 'too easy' to 'too difficult' (refer to Figure 2). Utilising subjective measures of task difficulty via Likert scales has been extensively demonstrated to be a reliable and valid method [48], especially for implementation with more complex tasks [49]. This approach has also had its effectiveness demonstrated in similar studies regarding the relationship between CL and achievement when implementing AR technologies into the study of geometric principles in chemistry comparable to this study [11,50,51]. The two post-tests were designed with isomorphic questions being incorporated in the content section of the pre-test, again with each question accompanying the prompt to measure cognitive load.

Draw the following molecule using dash, wedge and solid-line notation in the space next to the diagram and name the molecular geometry.



S = 

Br = 

F = 

Geometry name: _____

How difficult did you find this problem?

Too easy	Very, very easy	Very easy	Easy	Neither difficult nor easy	Difficult	Very difficult	Very, very difficult	Too difficult

Figure 2. Example of content question and cognitive load measurement from the pre-test.

A treatment lesson at T2 was undertaken based on molecular geometric and Valence Shell Electron Pair Repulsion (VSEPR) theory content. Identical booklets were provided to the I and C groups differing only in the inclusion or exclusion of an AR digital prompt that accompanied the VSEPR model of the molecular geometry. The digital prompt was to be used with an app downloaded onto each participant's mobile device to generate an interactive 3D model of the molecular geometry. In lieu of the digital prompt, the C group was instead given diagrams typical of that which would accompany this content in a more traditional lesson or textbook (refer to Figure 3). Teacher instruction given during the teaching aspect of the class was controlled by utilising the same lesson slides during both sessions.

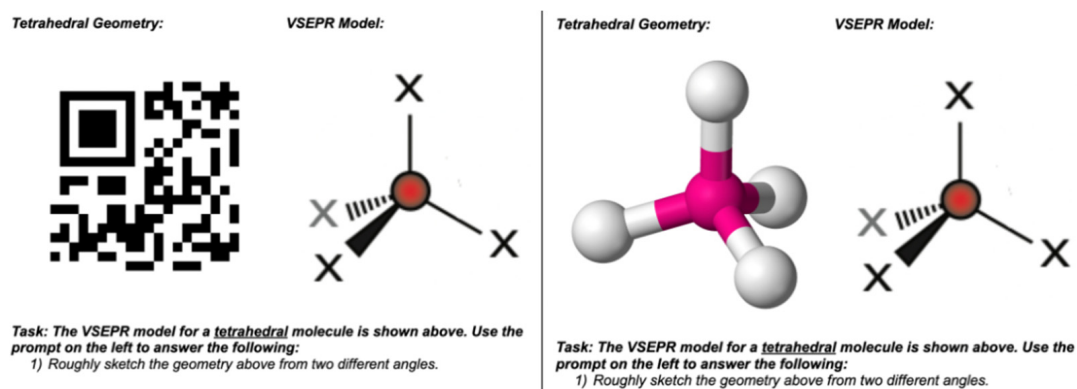


Figure 3. Example question from T2 lesson booklets illustrating the use of the AR digital prompt for the I group (on the **left**) compared with the diagram provided for the C group (on the **right**).

The MolecularAR App

A mobile app called ‘MolecularAR’ (version 1.1.3) was downloaded by the I group onto their mobile device for use with the digital AR prompt. The app selected provided the full functionality required to support the AR learning task, containing the necessary attributes of allowing participants to drag and pinch with their fingers when manipulating the perspectives of the geometric shapes that are aligned to the human movement effect. The body movement from the handheld device further enabled participants to reorient the position of the phone’s camera relative to the 3D model generated, adding to the human movement effect experienced (refer to Figure 4).

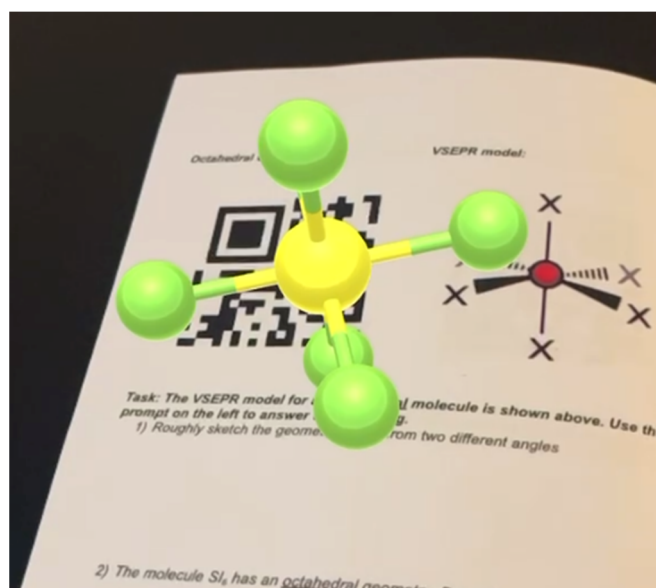


Figure 4. Example of AR app being used with learning booklet.

2.3. Validity and Reliability

The validity of the content delivery and testing phases of data collection were assured by aligning the content of the study with the relevant learning outcomes from the state curriculum implemented at the school. Careful selection of the year group was considered, as participants needed a certain level of prerequisite understanding without having been introduced to the content itself. A Year 10 cohort was ultimately selected, as the content from the curriculum is introduced in Year 11. The appropriateness of this choice was further informed by prior pilot testing undertaken with a younger cohort. Interpretation of the statistical analysis of this data suggested an inadequate level of pre-existing knowledge

present in the younger cohort for meaningful data to be obtained, ultimately guiding the final choice of participants. Reliability was considered by undertaking data collection within the familiar environment of the participants' school during schooling hours, in addition to the prior pilot testing, to circumvent potential issues such as fatigue, nervousness, and misinterpretation of questions.

3. Results

Data from the study were analysed using IBM SPSS (Statistical Package for Social Science) Statistics. Cognitive load was scored numerically utilising a 9-point scale corresponding to the Likert scale options (from 0 = Too easy to 9 = Too difficult), and achievement in each phase of testing was scored as a value corresponding to their achievement on the test (as a score out of 14).

Achievement. Initial analyses indicated no significant main or interaction effects ($F < 1$ in most instances) associated with gender on either the achievement or cognitive load scores. The two groups did not diverge in pre-test scores ($F(1, 41) = 1.7$, ns). The achievement test data were analysed using t -test procedures. In both groups, the achievement scores increased from the pre-test to the two post-tests, respectively (dependent t s of 7.2 and 6.4, $p < 0.01$). The two groups differed significantly in the T4 post-test ($p = 0.02$), although differences in the baseline tests and initial T3 tests failed to achieve significance (see means and t values in Table 2).

Table 2. Statistical comparisons of intervention and control groups on achievement total and sub scores.

Time	Control ($n = 18$)	Intervention ($n = 24$)	t	p
T1 Pre-test achievement	2.72 (1.22)	3.42 (1.25)	1.8	0.08
T3 Post-test achievement	4.78 (1.35)	5.96 (2.22)	1.9	0.06
T4 Delayed post-test achievement	4.50 (2.09)	6.21 (2.49)	2.4	0.02

Cognitive load. A 2×3 repeated measures ANOVA, with time as the repeated measure, revealed a significant effect for time on cognitive load scores: $F(1, 40) = 83$, $p < 0.01$. Significant interaction between time and treatment, $F(1, 40) = 6.2$, $p = 0.02$, was found, indicating the impact of the treatment varied with time. It was evident that the two groups diverged in both the immediate and delayed post-tests, respectively: $F(1, 40) = 6.8$, $p = 0.01$ and $F(1, 40) = 7.12$, $p = 0.01$. The two groups did not differ at the pre-test level (see Table 3 for mean scores). An incidental finding is that the achievement scores correlated with the cognitive load scores on the two post-tests (-0.6 , and -0.53); that is, as achievement increased, cognitive load decreased.

Table 3. Statistical comparisons of intervention and control groups on cognitive load total and sub scores.

Time	Control ($n = 18$)	Intervention ($n = 24$)	t	p
T1 Pre-test cognitive load	60.33 (5.80)	59.45 (6.86)	0.4	0.07
T3 Post-test cognitive load	52.66 (6.39)	43.95 (11.61)	3.1	0.004
T4 Delayed post-test cognitive load	51.44 (6.81)	42.38 (12.72)	3.0	0.005

4. Discussion

Similar to other studies using CLT and CTML principles to improve visuospatial processing by guiding task design, this study aimed to reduce any unnecessary visuospatial processing to highlight the benefits of using AR technology to positively impact schema

acquisition and reduce participant-perceived cognitive load. While it was expected participants from both the I and C groups would interpret the mental rotation images and engage in similar cognitive activity, the results suggest participants who were provided with AR technology visualisations (I group) perceived less cognitive load and had moderately improved achievement scores, an inference of the AR technology being the only change in variable between the two groups. Those participants who did not engage in AR technology visualisations (C group) perceived higher cognitive load and did not achieve as highly in test achievement when compared to the I group.

Understanding this perceived reduction of cognitive load and moderately improved achievement by the I group within the framework of CLT principles would suggest those who experienced AR technology:

1. Had greater WM-available resources to undertake the learning activity by drawing upon schemas from the LTM;
2. Were able to process novel information gained during the task due to the availability of WM resources; and
3. Were able to transfer new schemas to the LTM.

The content used in the study was designed to target both I and C participants' prior knowledge to support the construction of new schemas within their existing cognitive architecture. Yet, the I group participants were able to apply mental rotation visualisations more freely, with less obstruction. The interpretation of the self-reporting data suggesting that those I group participants who experienced a lower level of cognitive load ultimately achieved improved learning results in part due to the greater availability of cognitive resources that could be assigned to the learning of new content material. This understanding is supported by other studies [52] involving the integration of new information to reduce cognitive load to maximise working memory resources, offering benefit to schema acquisition [53] and enhancing the integration of lower-level schemas together to form higher-level schemas with an increasing level of complexity and autonomy [54].

The participants' ability to use visuospatial representations and cueing may have also supported the drawing of the pre-requisite schemas required to interpret and experience these representations. Known as the signalling principle, this ability enables students to chunk pre-existing schemas, sometimes incorporating multiple representations into fewer elements or even a single element, freeing up cognitive resources. The lower level of cognitive load experienced encourages the I group to construct schemas, optimising the relationship observed between mental rotation and achievement in both phases of post-testing. This is congruent with other findings that demonstrate the effectiveness that AR technology can have in reducing the level of cognitive load influence by mental rotation when applied to studying three-dimensional structures at the molecular level [11]. It also reiterates claims regarding the efficacy of the technology to increase learning gains in molecular chemistry [55].

A significant feature of the AR technology required students to use a mobile device utilising AR technology and required hand movements to touch the screen to rotate the three-dimensional molecular structures. Finger movement and manipulation, a commonly evolved automated skill, requires very few WM resources and enables students to move the 3D molecular geometry content with little instruction due to the intuitive design features based on CLT principles. The use of innate and evolved physical resources, such as the use of fingers to manipulate objects, can be applied to manipulating geometric shapes on a mobile device and is aligned with the human movement effect. The lower performance from the C group participants who did not engage in AR technology would suggest that they were not able to engage in mental rotation visualisation as successfully when compared to the I group. The instructional approach used by the C group implies that many of the C participants' WM resources were depleted during the instructional phase of the task and while attempting to mentally rotate the molecular geometry rather than being applied to solving the problem and creating and transferring new schemas to their LTM.

Limitations and Future Research

Due to logistical restraints, the sample size of the study was limited. Extending the scope of the study to incorporate more participants in future studies could allow for additional meaningful insights and strengthen the confidence in the conclusions drawn. Mental rotation abilities have been observed to vary due to a number of factors, including socio-economic status, education level, and language [14]. Future directions could consider the relationship between AR-based learning interventions designed using CLT principles and these additional factors by collecting data from a more diverse range of contexts. The method of collecting cognitive load measurements using a Likert scale also comes with inherent limitations [56]. Future directions could incorporate alternate or additional measurements of cognitive load, e.g., through measurements of individual types of cognitive load or via direct physiological measurements. A mixed-method approach to CLT research could also yield interesting insight [57,58], where the incorporation of items such as open-ended questions in the data collection may allow for more complex inferences to take shape.

5. Conclusions

This study investigated the effect of AR technology using mobile devices containing instructional 3D molecular geometry content, designed based on CLT principles, to encourage hand movements to rotate three-dimensional molecular structures. Statistical analysis using SPSS revealed that AR technology improved the ability of participants to mentally rotate molecular geometries and perceive less cognitive load. The study found mental rotation ability has an effect on cognitive load, but cognitive load can be altered using AR technology.

The optimised implementation of AR technology within the context of the study requires careful consideration. CTML suggests that active engagement is required to construct internal mental representations when processing multiple representations of external information [59]. Processing and drawing referential connections between multiple representations, such as those composed with AR learning, imposes high cognitive expenses [60]. Further investigation is required to determine whether this technology becomes more beneficial to participants who are less adept in their visuospatial abilities or to those who have mental rotation mastery. The possibility of creating specific content tailored to both lower and higher levels of expertise may prove to add more benefit to the application of AR in the classroom. Additionally, further investigation is required to establish whether other subject areas that require mental rotation, such as mathematical spatial reasoning, may also benefit from the use of AR technology.

The study provides a strong rationale for the application of AR technology when teaching chemistry and three-dimensional spatial arrangements of matter at the molecular level. The benefits of using AR technology can be explained from both the human movement effect and the signalling principle, highlighting its advantageous and beneficial effects on student learning.

Author Contributions: Conceptualisation, C.D.K. and B.B.; methodology, C.D.K. and B.B.; software, C.D.K. and B.B.; validation, C.D.K. and B.B.; formal analysis, C.D.K. and B.B.; investigation, C.D.K. and B.B.; resources, C.D.K. and B.B.; data curation, C.D.K. and B.B.; writing—original draft preparation, C.D.K.; writing—review and editing, C.D.K. and B.B.; visualisation, C.D.K. and B.B.; supervision, B.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of the University of Adelaide (H-2023-086, 8 May 2023).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Requests to access the data should be directed to the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Geary, D.C. Evolutionary Educational Psychology. In *APA Educational Psychology Handbook*; American Psychological Association: Washington, DC, USA, 2012; pp. 597–621.
- Geary, D.C.; Berch, D.B. Evolution and Children’s Cognitive and Academic Development. In *Evolutionary Perspectives on Child Development and Education*; Evolutionary Psychology; Springer International Publishing: Cham, Switzerland, 2016; pp. 217–249.
- Castro-Alonso, J.C.; Atit, K. Different Abilities Controlled by Visuospatial Processing. In *Visuospatial Processing for Education in Health and Natural Sciences*; Springer International Publishing: Cham, Switzerland, 2019; pp. 23–51.
- Sweller, J.; Ayres, P.; Kalyuga, S. *Cognitive Load Theory*, 1st ed.; Springer: New York, NY, USA, 2011.
- Sweller, J. Cognitive load theory and educational technology. *Educ. Technol. Res. Dev.* **2020**, *68*, 1–16. [CrossRef]
- Paas, F.; van Gog, T.; Sweller, J. Cognitive Load Theory: New Conceptualizations, Specifications, and Integrated Research Perspectives. *Educ. Psychol. Rev.* **2010**, *22*, 115–121. [CrossRef]
- Johnstone, A.H. Teaching of chemistry—logical or psychological? *Chem. Educ. Res. Pract.* **2000**, *1*, 9–15. [CrossRef]
- Taber, K.S. Revisiting the chemistry triplet: Drawing upon the nature of chemical knowledge and the psychology of learning to inform chemistry education. *Chem. Educ. Res. Pract.* **2013**, *14*, 156–168. [CrossRef]
- Mazzuco, A.; Krassmann, A.L.; Reategui, E.; Gomes, R.S. A systematic review of augmented reality in chemistry education. *Rev. Educ.* **2022**, *10*, e3325. [CrossRef]
- Habig, S. Who can benefit from augmented reality in chemistry? Sex differences in solving stereochemistry problems using augmented reality. *Br. J. Educ. Technol.* **2020**, *51*, 629–644. [CrossRef]
- Keller, S.; Rumann, S.; Habig, S. Cognitive Load Implications for Augmented Reality Supported Chemistry Learning. *Information* **2021**, *12*, 96. [CrossRef]
- Yang, F.-Y.; Wang, H.-Y. Tracking visual attention during learning of complex science concepts with augmented 3D visualizations. *Comput. Educ.* **2023**, *193*, 104659. [CrossRef]
- Sommerauer, P.; Müller, O. Augmented Reality in Informal Learning Environments: Investigating Short-term and Long-term effects. In Proceedings of the 51st Hawaii International Conference on System Sciences, Waikoloa Village, HI, USA, 3–6 January 2018.
- Farrell, S.; Duffy, G.; Bowe, B. A cross-cultural exploration of spatial visualisation abilities of first year STEM students: Students from Gulf States and Ireland. In Proceedings of the 2015 International Conference on Interactive Collaborative Learning (ICL), Florence, Italy, 20–24 September 2015; pp. 922–926.
- Sweller, J.; Sweller, S. Natural Information Processing Systems. *Evol. Psychol.* **2006**, *4*, 434–458. [CrossRef]
- Castro-Alonso, J.C.; Ayres, P.; Sweller, J. Instructional Visualizations, Cognitive Load Theory, and Visuospatial Processing. In *Visuospatial Processing for Education in Health and Natural Sciences*; Springer International Publishing: Cham, Switzerland, 2019; pp. 111–143.
- Sweller, J.; Van Merriënboer, J.J.G.; Paas, F. Cognitive Architecture and Instructional Design: 20 Years Later. *Educ. Psychol. Rev.* **2019**, *31*, 261–292. [CrossRef]
- Sweller, J. The Role of Evolutionary Psychology in Our Understanding of Human Cognition: Consequences for Cognitive Load Theory and Instructional Procedures. *Educ. Psychol. Rev.* **2021**, *34*, 2229–2241. [CrossRef]
- Sweller, J. Evolution of human cognitive architecture. *Psychol. Learn. Motiv.* **2003**, *43*, 215–266. [CrossRef]
- Sweller, J. Human Cognitive Architecture: Why Some Instructional Procedures Work and Others Do Not. In *APA Educational Psychology Handbook*, 1st ed.; American Psychological Association: Washington, DC, USA, 2012.
- Cassilhas, R.C.; Tufik, S.; de Mello, M.T. Physical exercise, neuroplasticity, spatial learning and memory. *Cell. Mol. Life Sci.* **2016**, *73*, 975–983. [CrossRef] [PubMed]
- Castro-Alonso, J.C. Overview of Visuospatial Processing for Education in Health and Natural Sciences. In *Visuospatial Processing for Education in Health and Natural Science*; Springer International Publishing: Cham, Switzerland, 2019; pp. 1–21.
- Lowrie, T.; Logan, T.; Hegarty, M. The Influence of Spatial Visualization Training on Students’ Spatial Reasoning and Mathematics Performance. *J. Cogn. Dev.* **2019**, *20*, 729–751. [CrossRef]
- Lippa, R.A.; Collaer, M.L.; Peters, M. Sex Differences in Mental Rotation and Line Angle Judgments Are Positively Associated with Gender Equality and Economic Development Across 53 Nations. *Arch. Sex. Behav.* **2009**, *39*, 990–997. [CrossRef]
- Li, X.; Wang, W. Exploring Spatial Cognitive Process Among STEM Students and Its Role in STEM Education: A Cognitive Neuroscience Perspective. *Sci. Educ.* **2020**, *30*, 121–145. [CrossRef]
- Harris, D.; Lowrie, T.; Logan, T.; Hegarty, M. Spatial reasoning, mathematics, and gender: Do spatial constructs differ in their contribution to performance? *Br. J. Educ. Psychol.* **2021**, *91*, 409–441. [CrossRef]
- Kaur, N.; Pathan, R.; Khwaja, U.; Murthy, S. GeoSolvAR: Augmented Reality Based Solution for Visualizing 3D Solids. In Proceedings of the IEEE Ninth International Conference on Technology for Education, Mumbai, India, 9–13 July 2018; pp. 372–376.

28. Weckbacher, L.M.; Okamoto, Y. Mental rotation ability in relation to self-perceptions of high school geometry. *Learn. Individ. Differ.* **2014**, *30*, 58–63. [CrossRef]
29. Jaeger, A.J.; Shipley, T.F.; Reynolds, S.J. The Roles of Working Memory and Cognitive Load in Geoscience Learning. *J. Geosci. Educ.* **2017**, *65*, 506–518. [CrossRef]
30. Stieff, M. Mental rotation and diagrammatic reasoning in science. *Learn. Instr.* **2007**, *17*, 219–234. [CrossRef]
31. Lee, D.Y.; Shin, D.-H. Effects of spatial ability and richness of motion cue on learning in mechanically complex domain. *Comput. Hum. Behav.* **2011**, *27*, 1665–1674. [CrossRef]
32. Sadoski, M.; Paivio, A. *Imagery and Text: A Dual Coding Theory of Reading and Writing*, 2nd ed.; Routledge: New York, NY, USA, 2013.
33. Uttal, D.H.; Meadow, N.G.; Tipton, E.; Hand, L.L.; Alden, A.R.; Warren, C.; Newcombe, N.S. The Malleability of Spatial Skills: A Meta-Analysis of Training Studies. *Psychol. Bull.* **2013**, *139*, 352–402. [CrossRef] [PubMed]
34. Fatemah, A.; Rasool, S.; Habib, U. Interactive 3D Visualization of Chemical Structure Diagrams Embedded in Text to Aid Spatial Learning Process of Students. *J. Chem. Educ.* **2020**, *97*, 992–1000. [CrossRef]
35. Weng, C.; Otanga, S.; Christianto, S.M.; Chu, R.J.-C. Enhancing Students' Biology Learning by Using Augmented Reality as a Learning Supplement. *J. Educ. Comput. Res.* **2020**, *58*, 747–770. [CrossRef]
36. Peeters, H.; Habig, S.; Fechner, S. Does Augmented Reality Help to Understand Chemical Phenomena during Hands-On Experiments?—Implications for Cognitive Load and Learning. *Multimodal Technol. Interact.* **2023**, *7*, 9. [CrossRef]
37. de Koning, B.B.; Marcus, N.; Brucker, B.; Ayres, P. Does observing hand actions in animations and static graphics differentially affect learning of hand-manipulative tasks? *Comput. Educ.* **2019**, *141*, 103636. [CrossRef]
38. Bentley, B.; Walters, K.; Yates, G.C.R. Using iconic hand gestures in teaching a year 8 science lesson. *Appl. Cogn. Psychol.* **2023**, *37*, 496–506. [CrossRef]
39. Castro-Alonso, J.C.; Ayres, P.; Paas, F. Animations showing Lego manipulative tasks: Three potential moderators of effectiveness. *Comput. Educ.* **2015**, *85*, 1–13. [CrossRef]
40. Wang, B.; Ginns, P.; Mockler, N. Sequencing tracing with imagination. *Educ. Psychol. Rev.* **2022**, *34*, 421–449. [CrossRef]
41. De Koning, B.B.; Tabbers, H.K. Facilitating understanding of movements in dynamic visualizations: An embodied perspective. *Educ. Psychol. Rev.* **2011**, *23*, 501–521. [CrossRef]
42. Ginns, P.; Kydd, A. Learning Human Physiology by Pointing and Tracing: A Cognitive Load Approach. In *Advances in Cognitive Load Theory*; Routledge: New York, NY, USA, 2019; pp. 119–129.
43. Sepp, S.; Tindall-Ford, S.; Agostinho, S.; Pass, F. Capturing Movement: A Tablet App, Geometry Touch, for Recording Onscreen Finger-based Gesture Data. *IEEE Trans. Learn. Technol.* **2023**, *17*, 73–83. [CrossRef]
44. Brucker, B.; Ehrlis, A.-C.; Häußinger, F.B.; Fallgatter, A.J.; Gerjets, P. Watching corresponding gestures facilitates learning with animations by activating human mirror-neurons: An fNIRS study. *Learn. Instr.* **2015**, *36*, 27–37. [CrossRef]
45. Pang, C.G.; Cai, Y. Transforming Learning Experiences Through Affordances of Virtual and Augmented Reality. In *Mixed Reality for Education*; Cai, Y., Mangina, E., Goei, S.L., Eds.; Springer Nature Singapore: Singapore, 2023; pp. 109–165.
46. Elford, D.; Lancaster, S.J.; Jones, G.A. Exploring the Effect of Augmented Reality on Cognitive Load, Attitude, Spatial Ability, and Stereochemical Perception. *J. Sci. Educ. Technol.* **2022**, *31*, 322–339. [CrossRef] [PubMed]
47. Elford, D.; Lancaster, S.J.; Jones, G.A. Fostering Motivation toward Chemistry through Augmented Reality Educational Escape Activities. A Self-Determination Theory Approach. *J. Chem. Educ.* **2022**, *99*, 3406–3417. [CrossRef]
48. Leppink, J.; Gog van, T.; Paas, F.; Sweller, J. Cognitive Load Theory: Researching and Planning Teaching to Maximise Learning. In *Researching Medical Education*; Wiley Blackwell: Chichester, UK, 2015; pp. 207–218.
49. Schmeck, A.; Opfermann, M.; Van Gog, T.; Paas, F.; Leutner, D. Measuring cognitive load with subjective rating scales during problem solving: Differences between immediate and delayed ratings. *Instr. Sci.* **2015**, *43*, 93–114. [CrossRef]
50. Chen, S.-Y.; Liu, S.-Y. Using augmented reality to experiment with elements in a chemistry course. *Comput. Hum. Behav.* **2020**, *111*, 106418. [CrossRef]
51. Abdinejad, M.; Talaie, B.; Qorbani, H.S.; Dalili, S. Student perceptions using augmented reality and 3d visualization technologies in chemistry education. *J. Sci. Educ. Technol.* **2021**, *30*, 87–96. [CrossRef]
52. Klepsch, M.; Schmitz, F.; Seufert, T. Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. *Front. Psychol.* **2017**, *8*, 1997. [CrossRef]
53. Bentley, B.; Yates, G.C.R. Facilitating proportional reasoning through worked examples: Two classroom-based experiments. *Cogent Educ.* **2017**, *4*, 1297213. [CrossRef]
54. Leppink, J.; Hanham, J. Human Cognitive Architecture Through the Lens of Cognitive Load Theory. In *Instructional Design Principles for High-Stakes Problem-Solving Environments*; Springer: Singapore, 2018; pp. 9–23.
55. Bullock, M.; Graulich, N.; Huwer, J. Using an Augmented Reality Learning Environment to Teach the Mechanism of an Electrophilic Aromatic Substitution. *J. Chem. Educ.* **2024**, *101*, 1534–1543. [CrossRef]
56. Greenberg, K.; Zheng, R. Cognitive load theory and its measurement: A study of secondary tasks in relation to working memory. *J. Cogn. Psychol.* **2022**, *34*, 497–515. [CrossRef]
57. Dohaney, J.; Brogt, E.; Kennedy, B. Strategies and Perceptions of Students' Field Note-Taking Skills: Insights From a Geothermal Field Lesson. *J. Geosci. Educ.* **2015**, *63*, 233–249. [CrossRef]

58. Tremblay, M.-L.; Leppink, J.; Leclerc, G.; Rethans, J.-J.; Dolmans, D.H.J.M. Simulation-based education for novices: Complex learning tasks promote reflective practice. *Med. Educ.* **2019**, *53*, 380–389. [CrossRef]
59. Kapp, S.; Thees, M.; Beil, F.; Weatherby, T.; Burde, J.-P.; Wilhelm, T.; Kuhn, J. The Effects of Augmented Reality: A Comparative Study in an Undergraduate Physics Laboratory Course. In Proceedings of the 12th International Conference on Computer Supported Education (CSEDU 2020), Prague, Czech Republic, 2–4 May 2020; pp. 197–206.
60. Scheid, J.; Müller, A.; Hettmannsperger, R.; Schnotz, W. Improving learners' representational coherence ability with experiment-related representational activity tasks. *Phys. Rev. Phys. Educ. Res.* **2019**, *15*, 010142. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Review

How Scientific Is Cognitive Load Theory Research Compared to the Rest of Educational Psychology?

Amedee Marchand Martella ^{1,2,*}, Alyssa P. Lawson ³ and Daniel H. Robinson ⁴¹ Department of Psychological & Brain Sciences, University of California, Santa Barbara, CA 93106, USA² Department of Educational Psychology, University of Georgia, Athens, GA 30602, USA³ Institute for Research and Training, Landmark College, Putney, VT 05346, USA; alyssalawson@landmark.edu⁴ Department of Higher Education, Adult Learning and Organizational Studies, The University of Texas at Arlington, Arlington, TX 76019, USA; daniel.robinson@uta.edu

* Correspondence: amedeemartella@uga.edu

Abstract: Cognitive load theory (CLT) has driven numerous empirical studies for over 30 years and is a major theme in many of the most cited articles published between 1988 and 2023. However, CLT articles have not been compared to other educational psychology research in terms of the research designs used and the extent to which recommendations for practice are justified. As Brady and colleagues found, a large percentage of the educational psychology articles reviewed were not experimental and yet frequently made specific recommendations from observational/correlational data. Therefore, in this review, CLT articles were examined with regard to the types of research methodology employed and whether recommendations for practice were justified. Across several educational psychology journals in 2020 and 2023, 16 articles were determined to directly test CLT. In contrast to other articles, which employed mostly observational methods, all but two of the CLT articles employed experimental or intervention designs. For the two CLT articles that were observational, recommendations for practice were not made. Reasons for the importance of experimental work are discussed.

Keywords: causation; correlational; cognitive load theory; experimental; recommendations for practice

Citation: Martella, A.M.; Lawson, A.P.; Robinson, D.H. How Scientific Is Cognitive Load Theory Research Compared to the Rest of Educational Psychology? *Educ. Sci.* **2024**, *14*, 920. <https://doi.org/10.3390/educsci14080920>

Academic Editors: Fred Paas, Andre Tricot, Kim Ouwehand and Florence Lespiau

Received: 10 July 2024

Revised: 14 August 2024

Accepted: 19 August 2024

Published: 22 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. How Scientific Is Cognitive Load Theory Research Compared to the Rest of Educational Psychology?

1.1. Overview

Cognitive load theory (CLT) emerged in the late 1980s and early 1990s as a theory of learning intended to guide instructional design. The term “cognitive load” first appeared in an article title in 1988 by John Sweller [1]. In a recent study by Hassan et al. [2] that examined the most cited authors and articles appearing in 12 educational psychology journals from 1988 to 2023, five of the top 30 most cited articles addressed CLT, and the most cited author was John Sweller. Along with his colleagues, including Paul Chandler and Paul Ayres, Sweller advanced the notion of cognitive load into a more formalized theory in 1991, e.g., [3]. Strengths of CLT include being based on knowledge of human cognitive architecture [4] and having important implications for instructional design [5].

CLT rose in popularity within educational psychology and instructional design journals after its updated description in 1998 [4]. CLT has maintained its prominence within the educational research literature since its introduction [4,6]. In fact, in a *Scopus* database search using terms “cognitive load” and “cognitive load theory” in articles published between 1988 and 2023 in the twelve educational psychology journals reviewed by Hassan et al. [2], there have been 373 articles published with a combined citation number of 50,262. When looking year by year, there has been a general increase in the number of articles that include “cognitive load” and/or “cognitive load theory” and in the citation rate (see Figure 1).

There is no doubt that CLT has left an indelible mark on the field of educational psychology over the past 35 years.

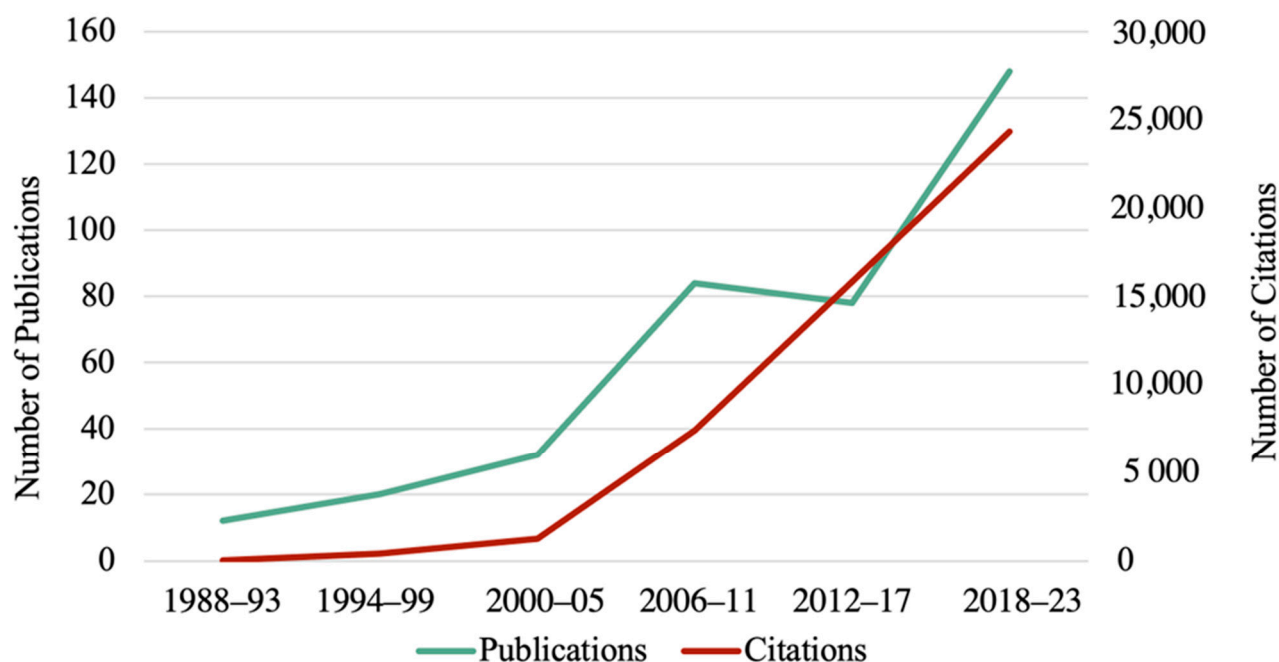


Figure 1. Articles from 1988 to 2023 that contain the terms “Cognitive Load” and/or “Cognitive Load Theory”.

However, despite its highly cited research base, there has been no specific comparative examination of CLT research to determine if there are any differences or similarities in methodological trends and recommendations for practice as compared to other educational psychology research. Such a comparison is important given that in the field of educational psychology as a whole, there has been a decline in intervention research and an increase in recommendations for practice being made from non-experimental studies [7]. With a decline in studies that provide strong causal evidence, current recommendations continue to be made based on shaky ground [8]. As such, the question remains: can the same be said about CLT research?

1.2. Causal Conclusions and Recommendations for Practice

Several reviews have been conducted over the years to investigate the rigor of methodology used in educational psychology research. For example, Hsieh and colleagues [9] published an article that examined studies published in 1983 and between 1995 and 2004 in three prominent educational psychology journals and one broader educational journal. They assessed the types of methodologies used in these studies and found that there was a decrease in the frequency of intervention studies and that the ‘quality indicators’ of these studies (e.g., length of intervention, treatment integrity) were rather poor.

In a similar line of investigation, Robinson and colleagues [10] investigated the types of methodologies used in the studies published in five teaching-and-learning research journals across two years, 1994 and 2004. They classified each article as an intervention study or as a nonintervention study (i.e., correlational, qualitative, or descriptive). They also examined whether causal conclusions were made in each study. Results indicated there was a decline in the percentage of intervention studies published in these journals from 1994 to 2004, with correlational and qualitative studies becoming increasingly popular. Further, there was an increase in the number of causal conclusions made, particularly in the nonintervention studies.

Extending the work of Hsieh et al. [9] and Robinson et al. [10], Reinhart and colleagues [11] also wanted to understand the trends occurring in educational psychology concerning methodology and recommendations for practice. In their work, the authors investigated the use of statistical modeling analyses used between 2000 and 2010. Similar to the previous studies mentioned, they found that less than a quarter of studies from these journals were categorized as intervention studies, and almost half of the nonintervention studies included recommendations for practice. Additionally, when modeling analyses were used within the nonintervention studies, they more commonly resulted in recommendations for practice than those without the modeling analyses.

In the most recent review, Brady et al. [7] found a decrease in the percentage of empirical articles that employed intervention and experimental methods, an increase in the percentage of articles that employed observational methods, and an increase in the percentages of the latter that included recommendations for practice (close to two-thirds as of 2020). These findings were in line with the findings from the previous reviews, making it evident that the field as a whole is moving away from experimental research and yet continues to make recommendations for practice.

As noted by Brady et al. [7], there are a myriad of empirical methods that can be used to study questions within the field of educational psychology, each serving an important purpose in the “exploration-to-intervention study sequence” (p. 2). However, problems arise when the systematic study of a hypothesis via experimental research is skipped (often the final step in the sequence) and cause-and-effect claims are made from correlational research. Most introductory research methods and statistics textbooks caution that correlation does not imply causation. Although most agree that causal claims and recommendations for practice should be based on experimental research, e.g., [12–15], others have argued that recommendations can be based on non-experimental studies, e.g., [16–18].

This debate over when causality has been established and when it is appropriate to make specific recommendations was recently featured in *Educational Psychology Review*, where several authors [8,14,16,17,19] responded to the Brady et al. [7] article. Some argued that experimental designs are not the only way to establish causal relationships and suggested that the right analysis (e.g., structural equation modeling—SEM) could enable causation from correlational data, e.g., [16,17]. For example, Dumas and Edelsbrunner [16] presented a five-phase process in which SEM could be used to develop implications for educational practice. They emphasized that it is not a “yes/no” question as to whether specific recommendations can be made but rather is a continuum based on the strength of evidence.

With this continuum in mind, there are certain methods that are stronger with regard to establishing causality and making subsequent recommendations. Mayer [14] rated modeling studies as having moderate strength in testing causal claims and experimental studies as having the highest strength. Randomized control trials have long been considered the gold standard and are the ideal design for establishing causality and making recommendations for practice [14,17,20,21]. The benefit of experimental designs is the ability to control for possible confounds and to specifically isolate the independent variable as the causal variable. When non-experimental research is conducted, one runs the risk of extraneous or confounding variables compromising the internal validity of the study—that is, the ability to make valid inferences or draw causal conclusions [20,22,23]. As noted by statistician Donald Rubin, “when it comes to causal inference, design always trumps analysis” (personal communication).

To inform educational research and practice and improve the educational experiences of students, effective interventions need to be identified by isolating and controlling variables. This hallmark of scientific reasoning [24] helps to ensure that “researchers can conclude that changes in the dependent variable (i.e., outcome) are caused by manipulations to the independent variable” [25] (p. 111). With educational psychology moving away from one of the most rigorous methods to investigate research questions, there is a risk of making decisions based on shaky ground—or at the very least, based on results that

are not of the highest strength. Shaky ground decisions could affect the validity of instructional practices and influence whether they are embraced and adopted by the educational practice community.

1.3. The Present Review

Given the shifts in the types of research designs being prioritized within educational psychology, the purpose of the present review was to compare research designs and recommendations for CLT research to other educational psychology research, with the primary research question being “What types of research designs and recommendations for practice are found within CLT research as compared to non-CLT research?”

2. Method

2.1. Journal Selection and Search Process

Brady et al. [7] reviewed 255 empirical articles that were published in one of five journals during the year 2020. The journals included *Journal of Educational Psychology*, *Contemporary Educational Psychology*, *Cognition and Instruction*, *Journal of Experimental Education*, and *American Educational Research Journal*. These journals are similar to those included in three previous studies by Hsieh et al. [9], Reinhart et al. [11], and Robinson et al. [10]. With permission, their 2020 database was accessed.

To identify CLT articles in the 2020 database, the question “What makes an article a CLT article?” needed to be asked. To identify articles related to cognitive load theory, the authors of the present review examined each of the 255 articles to determine whether CLT was an integral part of the article (herein termed a CLT article). “Integral” was operationalized as including measures of cognitive load within the study, using CLT as the overarching cognitive framework, using CLT in research hypotheses/predictions, or testing CLT as part of the study. In total, six articles were identified as CLT articles [26–31].

Due to only identifying a handful of CLT articles from the original Brady et al. [7] database, the present review’s search was expanded to include (a) the most recent publication year of 2023 for the same five journals represented in the 2020 search and (b) an additional journal, *Educational Psychology Review*, given that it has published empirical CLT research in the past. Through this updated search, an additional 10 CLT articles were identified [32–41]. Thus, the total sample for 2020 and 2023 included 16 CLT articles.

2.2. Coding and Analysis

Brady et al. [7] categorized articles based on their research designs. Articles were coded as observational/correlational, intervention, experimental, qualitative, and mixed method/multi-method. If the independent variable was manipulated, the article was either intervention or experimental. If an intervention study involved random assignment of individuals, it was further coded as experimental. If the independent variable was not manipulated, the article was categorized as observational/correlational, qualitative, or mixed method. Observational/correlational studies involved quantitative data, qualitative studies involved qualitative data, and mixed method/multi-method studies involved both quantitative and qualitative data. This same categorization system was used for 2023 CLT articles.

In addition to categorizing the research design used in each article, Brady et al. [7] coded for whether recommendations for practice (RFP) were included in studies that were not intervention. RFPs were defined as “if practice X is adopted/avoided or increased/decreased, then teacher or student outcome Y will improve” [11] (p. 244). This same coding system was used for the CLT articles pulled from the six journals published in 2023.

3. Results

3.1. Research Designs

Because the number of CLT articles was so few, these articles were simply compared to all of the educational psychology articles reviewed by Brady et al. [7] without having to adjust the denominator. See Table 1 for a breakdown of each CLT article and coding included in this analysis.

Table 1. Breakdown of each CLT article included.

Authors	Journal	Method	Recommendations for Practice?
2020			
Bichler et al. [26]	<i>Journal of Educational Psychology</i>	Experimental	--
de Koning et al. [27]	<i>Contemporary Educational Psychology</i>	Experimental	--
Merkt et al. [28]	<i>Journal of Educational Psychology</i>	Experimental	--
Miller-Cotto & Byrnes [29]	<i>Journal of Educational Psychology</i>	Observational/ Correlational	No
Schneider et al. [30]	<i>Journal of Educational Psychology</i>	Experimental	--
Zu et al. [31]	<i>Journal of Educational Psychology</i>	Experimental	--
2023			
Buchin & Mulligan [32]	<i>Journal of Educational Psychology</i>	Experimental	--
Ehrhart & Lindner [33]	<i>Contemporary Educational Psychology</i>	Experimental	--
Hoch et al. [34]	<i>Educational Psychology Review</i>	Experimental	--
Martin et al. [35]	<i>Contemporary Educational Psychology</i>	Observational/ Correlational	No
Park et al. [36]	<i>Educational Psychology Review</i>	Experimental	--
Pengelley et al. [37]	<i>Educational Psychology Review</i>	Experimental	--
Rau & Beier [38]	<i>Journal of Educational Psychology</i>	Intervention	--
Sondermann & Merket [39]	<i>Contemporary Educational Psychology</i>	Experimental	--
Wang et al. [40]	<i>Journal of Educational Psychology</i>	Experimental	--
Yang et al. [41]	<i>Journal of Educational Psychology</i>	Experimental	--

Across the CLT articles published in 2020 and 2023, all but three used random assignment experiments. One of the articles that was not experimental was intervention (without randomization of individuals), while the other two were observational/correlational. These results are not surprising and confirm the claim by Sweller et al. [4] that CLT research is characterized by experimental studies. This contrasts sharply with much of the field of educational psychology. Brady et al. [7] found that 25% of the empirical articles published in 2020 were intervention and only 20% were experimental.

3.2. Recommendations for Practice

Recommendations for practice made based on the research findings are appropriate when the methodology—experimental and intervention—supports causal claims [7]. Fourteen of the CLT articles were experimental or intervention, and as such, any recommendations made in those articles could be justified. The two articles that were observational/correlational did not include such recommendations. This contrasts sharply with much of the field of educational psychology—as demonstrated by Brady et al.’s [7] analysis, 66% of non-intervention articles published in 2020 included recommendations for practice.

4. Discussion

As confirmed by the several reviews conducted by Robinson and colleagues [9–11], a large proportion of educational psychology research does not involve experimental methods. The present review found that thirteen of the sixteen CLT articles published in the five reviewed journals in 2020 and six reviewed journals in 2023 used experimental methods, with an additional article categorized as “intervention.” Only two articles were

observational/correlational. These results differ greatly from the rest of the educational psychology literature examined by Brady et al. [7] and are not surprising given the emphasis on experiments and experimental data discussed throughout Sweller et al.'s [4] review summarizing research on CLT over the past 20 years. In fact, failure of experiments to replicate helped cognitive load theory to expand with the identification of boundary conditions [42]. As noted by Sweller [42], "Cognitive load theory's continual adaptation to new data is one of its primary virtues. The fact that those data are collected primarily from full experiments using randomised, controlled trials with real students studying materials from their own curricula is also a virtue" (pp. 95–96).

Although each type of empirical approach can serve a distinct and important purpose in research, a decline in experimental research affects the degree of confidence researchers and practitioners may have that the independent variable caused changes in the dependent variable. Without randomization and the control of variables, recommendations for practice are based on shaky ground. To be able to provide teachers with specific guidance on when and how to implement specific learning strategies and teaching approaches in their classroom, strong experimental research needs to be conducted to establish causality. CLT research, however, has consistently used experimental methods, and, as such, the recommendations for practice offered are based on solid ground. For example, CLT research on explicit instruction and problem-solving is rooted in randomized, controlled trials, strengthening recommendations for teachers looking to effectively teach their students.

Despite these results and justified recommendations from CLT research and other experimental educational psychology research, there remains a gap between research and practice [43]. In such areas as science education, much of the research-validated data on effective instructional practices have been overlooked in science educational policy. Sweller [44] noted, "it is regrettably rare for instructional design to be based on human cognitive architecture. Frequently, instructional design principles are promulgated as though human cognition either does not exist or if it does exist, it has no implications for instruction" (p. 37). Issues concerning the adoption of evidence-based practices are nothing new and were noted by Dempster [45] prior to the start of CLT. With the wealth of experimental data on CLT, particularly compared to other educational psychology research in general, its implications for education should not be overlooked.

4.1. Limitations and Future Directions

The present review has two primary limitations. First, the sample size for CLT articles was small and from only five/six journals, so it may not be representative of all CLT research published in other journals. However, based on Sweller and colleagues' [4] observations, it would be expected that CLT research appearing in other journals would be mostly experimental. Second, the present review's operational definition of a CLT article may not match other researchers' definition of CLT, potentially changing the sample of included and excluded articles.

As for future directions, in addition to examining the research designs and recommendations for practice as performed by Brady et al. [7] and in the present review, it is recommended that researchers conduct reviews on the controlled nature of educational research. More specifically, controlling variables is essential in rigorous research, with confounds threatening the internal validity of a study. Different threats to internal validity could be assessed, as done by Martella et al. [20] and Lawson et al. [22], to determine if the experimental/intervention research that is being conducted is rigorous. Although recommendations for practice are inappropriately made in non-experimental/nonintervention research, they may also be inappropriately made in experimental/intervention research if variables have not been well controlled. As stated by John Sweller, "the 'C' in 'RCT' is at least as important as the 'R'... Randomisation without control of variables gives random results" (personal communication).

4.2. Conclusions

The present review found that the CLT research published in five educational psychology journals in 2020 and six educational psychology journals in 2023 were primarily experimental. When not experimental or intervention, recommendations for practice were not made. These results differ from non-CLT educational psychology research as a whole. The importance of experimental research cannot be overstated, as it is an essential component in the “exploration-to-intervention study sequence” [7] (p. 2). Experimental research provides the strongest evidence and allows for causal conclusions to be made on solid rather than shaky ground. As such, CLT research and the accompanying recommendations appear to be based on solid ground.

Author Contributions: Conceptualization: A.M.M. and D.H.R.; Methodology: A.M.M., A.P.L. and D.H.R.; Validation: D.H.R.; Formal Analysis: A.M.M., A.P.L. and D.H.R.; Writing—original draft preparation: A.M.M., A.P.L. and D.H.R.; writing—review and editing: A.M.M., A.P.L. and D.H.R.; visualization: A.P.L. All authors have read and agreed to the published version of the manuscript.

Funding: The first author acknowledges support from the National Science Foundation Postdoctoral Research Fellowship Program under award number 2222208. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Acknowledgments: The authors would also like to acknowledge and thank Waseem Hassan for his assistance in examining the frequency with which the terms “cognitive load” and “cognitive load theory” have appeared in 12 educational psychology articles since 1988.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Sweller, J. Cognitive load during problem solving: Effects on learning. *Cogn. Sci.* **1988**, *12*, 257–285. [CrossRef]
- Hassan, W.; Martella, A.M.; Robinson, D.H. Identifying the most cited articles and authors in educational psychology journals from 1988–2023. *Educ. Psychol. Rev.* **2024**. *in press*.
- Sweller, J.; Chandler, P. Evidence for cognitive load theory. *Cogn. Instr.* **1991**, *8*, 351–362. [CrossRef]
- Sweller, J.; van Merriënboer, J.J.G.; Paas, F. Cognitive architecture and instructional design: 20 years later. *Educ. Psychol. Rev.* **2019**, *31*, 261–292. [CrossRef]
- Schnotz, W.; Kürschner, C. A reconsideration of cognitive load theory. *Educ. Psychol. Rev.* **2007**, *19*, 469–508. [CrossRef]
- de Jong, T. Cognitive load theory, educational research, and instructional design: Some food for thought. *Instr. Sci.* **2010**, *38*, 105–134. [CrossRef]
- Brady, A.; Griffin, M.M.; Lewis, A.R.; Fong, C.J.; Robinson, D.H. How scientific is educational psychology research? The increasing trend of squeezing causality and recommendations from non-intervention studies. *Educ. Psychol. Rev.* **2023**, *35*, 37. [CrossRef]
- Robinson, D.H.; Wainer, H. It’s just an observation. *Educ. Psychol. Rev.* **2023**, *35*, 83. [CrossRef]
- Hsieh, P.; Acee, T.; Chung, W.-H.; Hsieh, Y.-P.; Kim, H.; Thomas, G.D.; You, J.-I.; Levin, J.R.; Robinson, D.H. Is educational intervention research on the decline? *J. Educ. Psychol.* **2005**, *97*, 523–529. [CrossRef]
- Robinson, D.H.; Levin, J.R.; Thomas, G.D.; Pituch, K.A.; Vaughn, S. The incidence of “causal” statements in teaching-and-learning research journals. *Am. Educ. Res. J.* **2007**, *44*, 400–413. [CrossRef]
- Reinhart, A.L.; Haring, S.H.; Levin, J.R.; Patall, E.A.; Robinson, D.H. Models of not-so good behavior: Yet another way to squeeze causality and recommendations for practice out of correlational data. *J. Educ. Psychol.* **2013**, *105*, 241–247. [CrossRef]
- Alexander, P.A. In praise of (reasoned and reasonable) speculation: A response to Robinson et al.’s moratorium on recommendations for practice. *Educ. Psychol. Rev.* **2013**, *25*, 303–308. [CrossRef]
- Harris, K.R. Disallowing recommendations for practice and policy: A proposal that is both too much and too little. *Educ. Psychol. Rev.* **2013**, *25*, 309–316. [CrossRef]
- Mayer, R.E. How to assess whether an instructional intervention has an effect on learning. *Educ. Psychol. Rev.* **2023**, *35*, 64. [CrossRef]
- Renkl, A. Why practice recommendations are important in use-inspired basic research and why too much caution is dysfunctional. *Educ. Psychol. Rev.* **2013**, *25*, 317–324. [CrossRef]
- Dumas, D.; Edelsbrunner, P. How to make recommendations for educational practice from correlational data using structural equation models. *Educ. Psychol. Rev.* **2023**, *35*, 48. [CrossRef]
- Grosz, M.P. Should researchers make causal inferences and recommendations for practice on the basis of nonexperimental studies? *Educ. Psychol. Rev.* **2023**, *35*, 57. [CrossRef]

18. Grosz, M.P.; Rohrer, J.M.; Thoemmes, F. The taboo against explicit causal inference in nonexperimental psychology. *Perspect. Psychol. Sci.* **2020**, *15*, 1243–1255. [CrossRef]
19. Zitzmann, S.; Machts, N.; Hübner, N.; Schaubert, S.; Möller, J.; Lindner, C. The yet underestimated importance of communicating findings from educational trials to teachers, schools, school authorities, or policy makers (Comment on Brady et al. 2023)). *Educ. Psychol. Rev.* **2023**, *35*, 65. [CrossRef]
20. Martella, A.M.; Martella, R.C.; Yacilla, J.K.; Newson, A.; Shannon, E.N.; Voorhis, C. How rigorous is active learning research in STEM education? An examination of key internal validity controls in intervention studies. *Educ. Psychol. Rev.* **2023**, *35*, 107. [CrossRef]
21. Shavelson, R.J.; Towne, L. *Scientific Research in Education*; National Academy Press: Washington, DC, USA, 2002.
22. Lawson, A.P.; Martella, A.M.; LaBonte, K.; Delgado, C.Y.; Zhao, F.; Gluck, J.A.; Munns, M.E.; Wells LeRoy, A.; Mayer, R.E. Confounded or controlled? A systematic review of media comparison studies involving immersive virtual reality for STEM education. *Educ. Psychol. Rev.* **2024**, *36*, 69. [CrossRef]
23. Martella, R.C.; Nelson, J.R.; Morgan, R.L.; Marchand-Martella, N.E. *Understanding and Interpreting Educational Research*; The Guilford Press: New York, NY, USA, 2013.
24. Chen, Z.; Klahr, D. All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Dev.* **1999**, *70*, 1098–1120. [CrossRef]
25. Kaya, C. Internal validity: A must in research designs. *Educ. Res. Rev.* **2015**, *10*, 111–118. [CrossRef]
26. Bichler, S.; Schwaighofer, M.; Stadler, M.; Bühner, M.; Greiff, S.; Fischer, F. How working memory capacity and shifting matter for learning with worked examples—A replication study. *J. Educ. Psychol.* **2020**, *112*, 1320–1337. [CrossRef]
27. de Koning, B.B.; Rop, G.; Paas, F. Learning from split-attention materials: Effects of teaching physical and mental learning strategies. *Contemp. Educ. Psychol.* **2020**, *61*, 101873. [CrossRef]
28. Merkt, M.; Lux, S.; Hoogerheide, V.; van Gog, T.; Schwan, S. A change of scenery: Does the setting of an instructional video affect learning? *J. Educ. Psychol.* **2020**, *112*, 1273–1283. [CrossRef]
29. Miller-Cotto, D.; Byrnes, J.P. What's the best way to characterize the relationship between working memory and achievement? An initial examination of competing theories. *J. Educ. Psychol.* **2020**, *112*, 1074–1084. [CrossRef]
30. Schneider, S.; Nebel, S.; Beege, M.; Rey, G.D. The retrieval-enhancing effects of decorative pictures as memory cues in multimedia learning videos and subsequent performance tests. *J. Educ. Psychol.* **2020**, *112*, 1111–1127. [CrossRef]
31. Zu, T.; Hutson, J.; Loschky, L.C.; Rebello, N.S. Using eye movements to measure intrinsic, extraneous, and germane load in a multimedia learning environment. *J. Educ. Psychol.* **2020**, *112*, 1338–1352. [CrossRef]
32. Buchin, Z.L.; Mulligan, N.W. Retrieval-based learning and prior knowledge. *J. Educ. Psychol.* **2023**, *115*, 22–35. [CrossRef]
33. Ehrhart, T.; Lindner, M.A. Computer-based multimedia testing: Effects of static and animated representational pictures and text modality. *Contemp. Educ. Psychol.* **2023**, *73*, 102151. [CrossRef]
34. Hoch, E.; Sidi, Y.; Ackerman, R.; Hoogerheide, V.; Schiter, K. Comparing mental effort, difficulty, and confidence appraisals in problem-solving: A metacognitive perspective. *Educ. Psychol. Rev.* **2023**, *35*, 61. [CrossRef]
35. Martin, A.J.; Ginns, P.; Nagy, R.P.; Collie, R.J.; Bostwick, K.C.P. Load reduction instruction in mathematics and English classrooms: A multilevel study of student and teacher reports. *Contemp. Educ. Psychol.* **2023**, *72*, 102147. [CrossRef]
36. Park, B.; Korbach, A.; Ginns, P.; Brücken, R. How learners use their hands for learning: An eye-tracking study. *Educ. Psychol. Rev.* **2023**, *35*, 116. [CrossRef]
37. Pengelley, J.; Whipp, P.R.; Rovis-Hermann, N. A testing load: Investigating test mode effects on test score, cognitive load and scratch paper use with secondary school students. *Educ. Psychol. Rev.* **2023**, *35*, 67. [CrossRef]
38. Rau, M.A.; Beier, J.P. Exploring the effects of gesture-based collaboration on students' benefit from a perceptual training. *J. Educ. Psychol.* **2023**, *115*, 267–289. [CrossRef]
39. Sondermann, C.; Merkt, M. What is the effect of talking heads in educational videos with different types of narrated slides? *Contemp. Educ. Psychol.* **2023**, *74*, 102207. [CrossRef]
40. Wang, F.; Cheng, M.; Mayer, R.E. Improving learning-by-teaching without audience interaction as a generative learning activity by minimizing the social presence of the audience. *J. Educ. Psychol.* **2023**, *115*, 783–797. [CrossRef]
41. Yang, X.; Wang, F.; Mayer, R.E.; Hu, X.; Gu, C. Ocular foundations of the spatial contiguity principle: Designing multimedia materials for parafoveal vision. *J. Educ. Psychol.* **2023**, *115*, 1125–1140. [CrossRef]
42. Sweller, J. The development of cognitive load theory: Replication crises and incorporation of other theories can lead to theory expansion. *Educ. Psychol. Rev.* **2023**, *35*, 95. [CrossRef]
43. Zhang, L.; Kirschner, P.A.; Cobern, W.W.; Sweller, J. There is an evidence crisis in science educational policy. *Educ. Psychol. Rev.* **2022**, *34*, 1157–1176. [CrossRef]
44. Sweller, J. Chapter two: Cognitive load theory. *Psychol. Learn. Motiv.* **2011**, *55*, 37–76. [CrossRef]
45. Dempster, F.N. The spacing effect: A case study in the failure to apply the results of psychological research. *Am. Psychol.* **1988**, *43*, 627–634. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Education Sciences Editorial Office
E-mail: education@mdpi.com
www.mdpi.com/journal/education



Disclaimer/Publisher's Note: The title and front matter of this reprint are at the discretion of the Guest Editors. The publisher is not responsible for their content or any associated concerns. The statements, opinions and data contained in all individual articles are solely those of the individual Editors and contributors and not of MDPI. MDPI disclaims responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-4039-7