education sciences

MDPI

Article

A Human–AI Orchestration Framework for Mathematics Education

Gila Payevsky [1,2], Anita Stoyanova [3], Moshe Mellick [4], Nikolay Penev [4], Veneta Tabakova-Komsalova [3], Iskra Nencheva [5], Stanimir Stoyanov [3] and Emil Delinov [6,*]

[1] Mathematics Department, Yeshivat Yavne High School, Haifa 798550, Israel; gilahu100@gmail.com
[2] Faculty of Education, Trakia University, 6000 Stara Zagora, Bulgaria
[3] Faculty of Mathematics and Informatics, University of Plovdiv "Paisii Hilendarski", 4027 Plovdiv, Bulgaria; anita.georgieva@gmail.com (A.S.); v.komsalova@uni-plovdiv.bg (V.T.-K.); stani@uni-plovdiv.bg (S.S.)
[4] Computer Science Department, Kfar Hasidim High School, Kfar Hasidim 204900, Israel; mellick@gmail.com (M.M.); nikolay.penev@trakia-uni.bg (N.P.)
[5] Faculty of Economics, Trakia University, 6000 Stara Zagora, Bulgaria; iskra.nencheva@trakia-uni.bg
[6] Faculty of Digital and Green Technologies, Trakia University, 6000 Stara Zagora, Bulgaria
* Correspondence: emil.delinov@trakia-uni.bg

Abstract

Artificial Intelligence (AI) is increasingly influencing personalized and adaptive education. Existing research has examined Human–AI hybrid adaptivity and coordination between human participants and AI systems, but clearer conceptualization is needed of the pedagogical coordination of functionally different forms of instructional support, particularly in highly abstract domains such as mathematics. This study proposes a Human–AI Orchestration Framework integrating Intelligent Tutoring Systems (ITSs), Generative AI, and direct teacher support while retaining pedagogical orchestration and responsibility with the teacher. Drawing on Cognitive Load Theory, the Framework proposes Cognitive Load Redistribution as a theoretical proposition and introduces coordination mechanisms, including Handover Points and a Meta-Prompt. Its practical implementation was explored through an exploratory design pilot with 32 Grade 8 students in an authentic classroom, using observations and feedback to iteratively refine its configuration. The pilot informed several design refinements, including refinement of the Meta-Prompt, introduction of "Call the Teacher," and adaptation of interaction formats. These formative observations provide preliminary indications of the practical feasibility of coordinating functionally different and potentially complementary forms of instructional support, while the Framework's educational effectiveness and proposed conceptual mechanisms remain to be systematically evaluated in future research.

Keywords: artificial intelligence in education; mathematics education; Human–AI orchestration; Human–AI interaction; Intelligent Tutoring Systems; Generative AI; teacher mediation; cognitive load theory

Academic Editor: Sergio Miranda

Received: 10 September 2026
Revised: 8 October 2026
Accepted: 15 October 2026
Published: 22 October 2026

Copyright: © 2026 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.

1. Introduction

Artificial Intelligence (AI) is playing an increasingly important role in education, particularly in areas involving personalized learning and adaptive instructional support (Holmes et al., 2019; Luckin et al., 2016; OECD, 2023). Among the most rapidly developing technologies are Intelligent Tutoring Systems (ITSs) and Generative AI, both of which

demonstrate considerable potential to support procedural knowledge acquisition as well as conceptual understanding (Aleven et al., 2017; Kasneci et al., 2023; VanLehn, 2011). This is particularly important in disciplines characterized by a high degree of abstraction, such as mathematics, where learners are required to develop formal knowledge, visual–spatial representations, and abstract mathematical thinking simultaneously.

Despite the rapid development of these technologies, their pedagogically grounded integration into authentic learning environments remains a challenge (Chen et al., 2020; Holmes & Tuomi, 2022; Zawacki-Richter et al., 2019). Existing research has already demonstrated that adaptive support can be enacted through interaction between human participants and AI systems and that dynamic transitions between different learning configurations can be organized in AI-supported classrooms (Holstein et al., 2020; Yang et al., 2023). Research on AI-supported orchestration has also highlighted the importance teachers place on retaining pedagogical judgment when using such systems (Lawrence et al., 2024). Despite these advances, there remains a need for clearer conceptualization of the pedagogical coordination of functionally different forms and sources of instructional support, including structured and adaptive support from ITSs, dialogic and conceptual support from Generative AI, and direct pedagogical mediation by the teacher. This issue is particularly relevant to mathematics education, where procedural problem solving, conceptual understanding, and visual–spatial representations may require different forms of support at different moments in the learning process.

Consequently, one of the major challenges is no longer the use of individual AI technologies, but their pedagogically grounded and coordinated integration into the overall instructional process. The question is no longer whether Intelligent Tutoring Systems, Generative AI, and teachers can coexist within the same educational environment, but rather how their interaction can be organized so that each contributes its specific functional capabilities at the pedagogically appropriate moment.

This paper proposes a Human–AI Orchestration Framework as one possible conceptual approach for coordinating the interaction among Intelligent Tutoring Systems, Generative AI, and teachers in mathematics education. In the present study, Intelligent Tutoring Systems and Generative AI were selected as the technological components because of the different and potentially complementary forms of instructional support they can provide in mathematics learning activities that require the integration of procedural learning and conceptual understanding. Intelligent Tutoring Systems offer opportunities for structured and adaptive practice and immediate feedback, whereas Generative AI offers opportunities for natural-language interaction, explanation, and conceptual reasoning. Within the present study, teachers and learners are regarded as human participants, whereas Intelligent Tutoring Systems and Generative AI are regarded as AI components of the learning environment. When referring to the pedagogical, cognitive, and functional roles through which instructional support is provided within the learning process, the teacher, Intelligent Tutoring Systems, and Generative AI are collectively designated as instructional agents, whereas the learner participates as a human participant whose learning and development these instructional functions are intended to support. Accordingly, the term Human–AI interaction refers to the relationships and coordination between the human participants and these technological AI components. In this sense, the human–machine relationship addressed in the present study is specified as the interaction between humans and the AI components of the learning environment.

Within this configuration, the teacher occupies a distinctive position by performing two interrelated roles: as a pedagogical orchestrator who organizes and coordinates the learning environment, and as an instructional agent who provides direct pedagogical support to the learner when needed. The present study develops a conceptual perspective for

the design and investigation of coordinated Human–AI learning environments and includes an exploratory examination of the practical feasibility of the proposed Framework through its implementation in an authentic classroom setting.

The Framework conceptualizes the interaction between human participants and AI components as a process in which different cognitive operations, forms of instructional support, and sequences of interaction are dynamically coordinated according to the specific learning situation, the learner's current state and needs, the nature of the learning task, and the functional capabilities of the participating humans and technologies. Leading instructional functions may shift during this process, while pedagogical orchestration and responsibility remain with the teacher. Within this conceptual organization, Cognitive Load Theory provides a theoretical basis for the concept of Cognitive Load Redistribution.

On this basis, the paper formulates three main conceptual contributions to the theoretical understanding and design of coordinated Human–AI learning environments.

Cognitive Load Redistribution. Grounded in Cognitive Load Theory, the Framework proposes Cognitive Load Redistribution (CLR) as a framework-specific theoretical proposition concerning the pedagogical coordination of supportive functions and cognitive operations across the learner, technological components, and teacher according to the demands of the learning situation. CLR does not imply a literal transfer of cognitive load between agents, nor a fixed allocation of instructional functions. Rather, different forms of support may assist different aspects of the learning process according to their functional affordances and the learner's needs. Because the exploratory design pilot did not empirically measure cognitive load, this study presents CLR as a theoretical proposition rather than an empirically demonstrated effect.

Handover Points. The Framework introduces Handover Points (HPs) as a framework-specific conceptual and operational coordination mechanism for identifying pedagogically meaningful transitions in the leading form or function of instructional support. A Handover Point occurs when the leading form or function of support shifts in response to the learner's needs, the development of the task, or the functional affordances and limitations of the available forms of support. Such transitions are not prescribed as a fixed or unidirectional sequence: support may shift among the mini-ITS, Generative AI, and direct teacher mediation in different directions according to the learning situation. Handover Points therefore do not imply equivalence or interchangeability among these forms of support, nor do they constitute a transfer of pedagogical control or responsibility to the AI components. The teacher retains pedagogical orchestration and responsibility.

Visual–Conceptual Gap. The Framework introduces the Visual–Conceptual Gap (VCG) as a framework-specific conceptual lens for identifying situations in which the form of support currently available to the learner does not sufficiently connect mathematical meaning with the representations required for understanding the task. Grounded in perspectives on multiple representations and representational coordination, the VCG does not assume an inherent inability of Generative AI to support visual or representational reasoning. Rather, it draws attention to a possible mismatch between the support being provided and the representational demands of the learning situation. When such a mismatch occurs, it may indicate the need for a different or complementary form of support, which the mini-ITS, Generative AI, specialized representational tools, or direct teacher mediation may provide. In the present study, the VCG is used as a conceptual lens rather than as an independently validated construct.

The different functional capabilities and roles of the teacher, Intelligent Tutoring Systems, and Generative AI create potential for complementarity within a coordinated Human–AI learning environment. These participants do not perform identical functions and

do not possess interchangeable capabilities. The teacher retains pedagogical responsibility, interprets the learning situation, and decides how to organize instructional support, while the AI components provide technologically mediated support whose functions may vary with the learning situation and pedagogical configuration. Accordingly, the conceptual logic of the Framework is based not on functional equivalence between humans and AI, but on the pedagogically grounded coordination of their different capabilities as complementary resources.

Pedagogical orchestration provides the mechanism through which these different functional contributions are organized within the overall learning process. It enables the dynamic coordination of cognitive functions and forms of instructional support according to the specific learning situation and the learner's needs, while preserving the common educational orientation toward the learner's learning and development. From this perspective, the Human–AI Orchestration Framework provides a conceptual approach to designing learning environments in which the different capabilities of human participants and AI components are purposefully coordinated and used as complementary resources, rather than regarded as equivalent or interchangeable forms of instructional support.

Accordingly, the present study addresses the following research questions:

RQ1: How can a Human–AI Orchestration Framework be conceptually structured to coordinate complementary forms of instructional support provided by Intelligent Tutoring Systems, Generative AI, and direct teacher mediation in mathematics education while preserving teacher pedagogical responsibility?

RQ2: What practical design requirements, limitations, and refinements emerge from the implementation of the proposed Framework in an exploratory classroom pilot?

2. Theoretical Background

2.1. Artificial Intelligence in Mathematics Education

Artificial Intelligence (AI) is increasingly viewed as an important component of educational innovation, particularly in learning environments that involve ongoing practice, personalized support, and complex reasoning, such as mathematics. Existing research suggests that AI-supported learning environments may enhance personalization, provide adaptive feedback, and support student progress through tailored learning pathways (Holmes et al., 2019; Luckin et al., 2016; OECD, 2023). In mathematics education, these features may be particularly important because learning often depends on the gradual development of procedural fluency alongside conceptual understanding. At the same time, researchers emphasize that the educational value of AI may depend not only on technological capabilities, but also on the pedagogical frameworks within which these systems are implemented (Pane et al., 2017).

Intelligent Tutoring Systems (ITSs) are among the most established applications of AI in education. These systems have been widely studied for their ability to provide adaptive practice, immediate feedback, and structured scaffolding during problem solving. Prior research suggests that ITSs may effectively support procedural fluency, mastery learning, and step-by-step algebraic reasoning when aligned with curricular goals and learner needs (Aleven et al., 2017; Koedinger et al., 2013; VanLehn, 2011). Although ITSs appear effective in guiding structured problem-solving processes, their model-driven architecture may sometimes limit the flexibility of conceptual explanation and the depth of mathematical dialogue available to learners.

2.2. Generative AI and Conceptual Reasoning

In contrast to the structured architecture of ITS, Generative AI has emerged as a comparatively flexible and dialogic form of educational support. These systems may generate

explanations, examples, and reflective questions in natural language, creating opportunities for conceptual clarification, dialogic mathematical reasoning, metacognitive engagement, and exploratory reasoning (Holmes & Tuomi, 2022; Kasneci et al., 2023; Sfard, 2008). Recent studies suggest that Generative AI may potentially support students' sense-making processes (Kasneci et al., 2023; Walkington, 2026), as these systems can respond dynamically to student questions and provide multiple forms of explanation.

At the same time, researchers caution that unstructured reliance on Generative AI may contribute to superficial understanding, cognitive overload, or overdependence on automated support if its use is not pedagogically guided (Kasneci et al., 2023). Therefore, the contribution of Generative AI to mathematical learning may depend largely on how teachers structure and guide its use in classroom instruction.

A recent comprehensive review of generative AI applications across educational contexts similarly confirms the rapid expansion of this technology since 2023 and highlights that its instructional value depends strongly on structured pedagogical integration rather than unmediated use, reinforcing the rationale for pedagogically orchestrated Human–AI integration adopted in the present Framework (Mittal et al., 2024).

2.3. Teacher Mediation in AI-Supported Learning Environments

A recurring argument in the literature is that AI should not be viewed as a replacement for teachers, but rather as an instructional resource whose educational value may depend on human mediation and pedagogical orchestration (Dillenbourg, 2013; Kahn et al., 2025; Selwyn, 2019). This perspective may be particularly important in mathematics education, where students often require visual explanation, contextual interpretation, and affective support in addition to verbal or procedural guidance (Ainsworth, 2006; Duval, 2006; Selwyn, 2019). From this viewpoint, the teacher continues to play a central role in interpreting students' needs, bridging gaps in understanding, and connecting digital interactions to classroom meaning-making. Such mediation becomes especially important when students move between symbolic procedures, linguistic explanations, and visual mathematical representations (Ainsworth, 2006; Duval, 2006). This perspective also points to a distinction between the teacher's direct pedagogical mediation and the broader responsibility for organizing and coordinating different forms of instructional support within the learning environment.

2.4. Cognitive Load, Personalization, and Educational Equity

Cognitive Load Theory provides a theoretical perspective for considering how the design and coordination of instructional support may influence the cognitive demands placed on learners (Sweller, 2011). Research on AI-supported learning further suggests that different forms of technological support may assist different aspects of learning, including structured procedural support and dialogic conceptual reasoning (Kasneci et al., 2023; Sweller, 2011). Within the present Framework, these perspectives inform Cognitive Load Redistribution (CLR) as a theoretical proposition about coordinating supportive functions and cognitive operations across the learner, technological components, and teacher. This proposition does not assume that cognitive load is literally transferred between agents or that particular forms of support necessarily reduce cognitive load. CLR is also conceptually distinct from cognitive offloading and instructional scaffolding. Cognitive offloading refers to the use of external actions or resources to alter the information-processing requirements of a task and thereby reduce cognitive demand (Risko & Gilbert, 2016), whereas instructional scaffolding concerns the provision of temporary or adaptive support that enables learners to perform aspects of a task that they may not yet be able to accomplish independently (Wood et al., 1976). In contrast, CLR focuses on the pedagogical coordination of different supportive functions and cognitive operations across the

learner, technological components, and teacher within a shared learning environment. Cognitive offloading and scaffolding may therefore occur within particular interactions, but neither is equivalent to CLR as conceptualized in the present Framework. Rather, CLR provides a theoretical basis for examining how different forms of support may be coordinated in relation to the cognitive demands of the learning situation. At the same time, research on personalized learning suggests that adaptive, learner-centered instructional approaches may help reduce achievement gaps, especially for students from lower socioeconomic backgrounds (OECD, 2019, 2023; Pane et al., 2017). Because mathematics achievement gaps related to socioeconomic status persist, AI-supported personalization has been discussed as a possible means of providing more equitable access to customized instruction (OECD, 2019, 2021; Pane et al., 2017). However, research also emphasizes that technology alone may not necessarily ensure educational fairness; successful implementation may depend on thoughtful integration, ethical practices, and teaching approaches that account for specific educational contexts (Holmes & Tuomi, 2022).

2.5. Research Gaps and Conceptual Foundations

Overall, the literature suggests that Intelligent Tutoring Systems and Generative AI may offer distinct yet potentially complementary contributions to mathematics education. Intelligent Tutoring Systems (ITSs) may primarily support procedural accuracy and mastery-oriented practice, whereas Generative AI may facilitate explanation, reflection, and conceptual exploration (Aleven et al., 2017; Holmes & Tuomi, 2022; Kasneci et al., 2023; VanLehn, 2011). At the same time, teacher mediation continues to play an important role in ensuring that digital support is pedagogically meaningful, visually clear, and emotionally responsive (Dillenbourg, 2013; Kahn et al., 2025; Selwyn, 2019). However, research specifically addressing the coordinated pedagogical use of functionally different AI technologies, including ITS and Generative AI, together with direct teacher mediation remains comparatively limited.

Recent research has begun to examine how human participants and AI systems can jointly enact adaptive support, how teachers conceptualize sharing certain aspects of control with AI-supported orchestration tools, and how dynamic transitions between different learning configurations can be coordinated within classroom environments (Holstein et al., 2020; Lawrence et al., 2024; Yang et al., 2023). This work outlines an emerging research direction involving Human–AI hybrid adaptivity, shared control, and Human–AI co-orchestration. The present study focuses on a more specific question: how functionally different forms and sources of instructional support—including rule-based ITSs, Generative AI, and direct teacher mediation—can be pedagogically coordinated in mathematics education while pedagogical orchestration and responsibility remain with the teacher. In this context, there remains a need for clearer conceptualization of transitions between these forms of support in response to learners' changing procedural, conceptual, and representational needs. Accordingly, the contribution of the present Framework does not lie in proposing Human–AI collaboration, shared control, or dynamic transitions as new concepts in themselves. Rather, it lies in specifying a mathematics-education architecture in which functionally different forms of support are coordinated while pedagogical responsibility remains with the teacher, and in introducing framework-specific mechanisms and lenses for conceptualizing this coordination.

This research gap motivates the present study, which proposes a Human–AI Orchestration Framework for Mathematics Education. Recent theoretical and empirical developments further support the conceptual foundations of this Framework.

Recent research also contributes to the understanding of complementarity in Human–AI collaboration. In this context, information asymmetry and capability asymmetry between humans and artificial intelligence have been identified as potential sources of

complementarity and potentially improved joint performance (Hemmer et al., 2025). Of particular relevance to the present study is the broader conclusion that differences in human and AI capabilities do not in themselves guarantee positive outcomes; their complementary use requires an appropriate mechanism for coordination within collaborative activity (Hemmer et al., 2025). This perspective provides a theoretical basis for considering pedagogical orchestration as a mechanism for purposefully coordinating the different functional capabilities of human participants and AI components in the learning process.

These foundations are further reinforced by recent quantitative syntheses. A 2026 meta-analysis of empirical studies published between 2023 and 2025 found that Generative AI exerts a moderate positive effect on mathematics learning outcomes overall, with the strongest gains observed when the technology is integrated at a deep, creative-transformation level rather than used as a passive substitute for teacher-led instruction, and when it is embedded within collaborative rather than purely individual learning arrangements (Liu et al., 2026). These findings provide recent empirical support for the need for coordinated and pedagogically structured forms of Human–AI integration, which form part of the conceptual foundation of the present study.

3. A Human–AI Orchestration Framework for Mathematics Education

To address the identified research challenges, the present study proposes a Human–AI Orchestration Framework for Mathematics Education, which conceptualizes the AI-supported learning environment as a coordinated interaction between human participants (the teacher and the learner) and AI components.

3.1. Structure of the Proposed Human–AI Orchestration Framework

As illustrated in Figure 1, the proposed Framework coordinates the participation of the teacher, Intelligent Tutoring Systems (ITSs), and Generative AI, integrating their distinct instructional functions through pedagogical orchestration and coordinating pedagogically meaningful transitions between different forms of instructional support through Handover Points. In this way, the proposed Human–AI Orchestration Framework brings together different forms of instructional support within a coordinated pedagogical model for mathematics education. Building on prior research indicating the potential effectiveness of ITSs in supporting procedural fluency (Aleven et al., 2017; Koedinger et al., 2013; VanLehn, 2011), as well as the growing role of Generative AI in supporting conceptual reasoning through natural-language interaction (Holmes & Tuomi, 2022; Kasneci et al., 2023), the Framework presents these technologies as instructional agents with potentially complementary capabilities that, together with the teacher, function within a distributed cognitive system (Hutchins, 1995).

Drawing on Cognitive Load Theory (Sweller, 2011), the Framework incorporates Cognitive Load Redistribution (CLR) as a framework-specific theoretical proposition concerning the pedagogical coordination of supportive functions and cognitive operations across the learner, technological components, and teacher according to the demands of the learning situation. Within this coordinated environment, Intelligent Tutoring Systems may provide structured procedural support, while Generative AI may provide dialogic and conceptual support; however, these functions are not treated as fixed or mutually exclusive. Building on the existing literature on teacher mediation and pedagogical orchestration in digital learning environments (Dillenbourg, 2013; Kahn et al., 2025; Selwyn, 2019), the Framework defines the teacher in two interrelated roles: as an instructional agent who provides direct pedagogical support and as a pedagogical orchestrator who organizes and coordinates the Human–AI learning environment. The Framework further uses the Visual–Conceptual Gap (VCG) as a conceptual lens for identifying situations in

which the current form of support does not sufficiently connect mathematical meaning with the representations required for understanding the task.

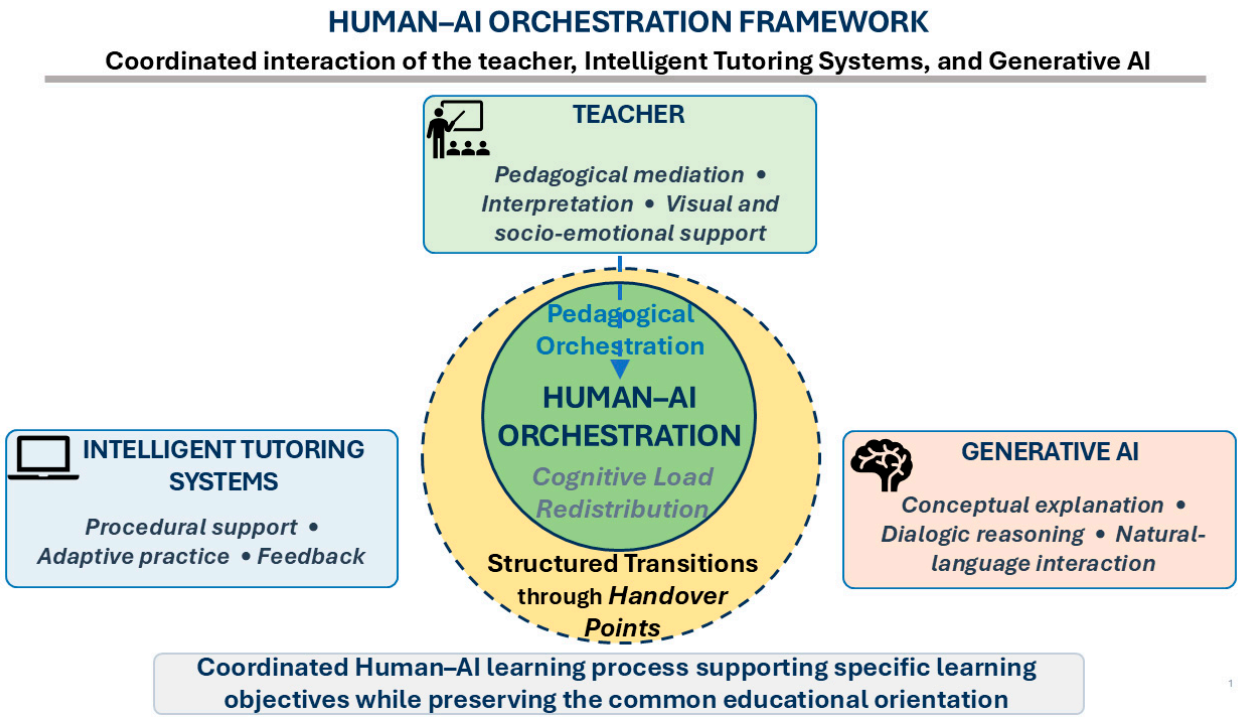


Figure 1. Human–AI Orchestration Framework Coordinated through Handover Points.

By integrating these forms of instructional support within a coordinated learning process, the Framework seeks to address the identified need for clearer conceptualization of pedagogical coordination within Human–AI learning environments in mathematics education. The Framework therefore provides a theoretically grounded conceptual approach to designing such environments.

3.2. Roles Within the Human–AI Orchestration Framework

Coordinated Human–AI learning environments constitute heterogeneous systems in which human participants and AI components jointly participate in the learning process. Within the proposed Framework, the teacher and the learner are the human participants, while Intelligent Tutoring Systems (ITSs) and Generative AI constitute the technological components. The teacher, ITS, and Generative AI perform different instructional functions and are therefore collectively referred to in this paper as instructional agents. The learner participates in Human–AI interaction as a human participant whose learning and development these instructional functions are intended to support. The common functional designation "instructional agents" does not imply equality, equivalence, or interchangeability between the teacher and the AI components, but emphasizes their participation in supporting the learning process.

The instructional agents differ in their functional capabilities, instructional functions, cognitive characteristics, decision-making processes, and forms of knowledge representation. Within the Framework, Intelligent Tutoring Systems (ITSs) may primarily support procedural practice, adaptive feedback, and structured guidance, while Generative AI may provide conceptual, dialogic, and reflective support. The teacher may provide direct pedagogical, representational, contextual, and socio-emotional support, while retaining

responsibility for organizing and coordinating the learning process. These functional emphases are not treated as fixed or mutually exclusive, but may vary according to the learning situation, the learner's needs, and the capabilities of the available forms of support. Within the proposed Framework, these differences are not regarded as deficiencies to be eliminated, but as functional characteristics that create opportunities for different forms of instructional support to complement one another.

Within this structure, the teacher occupies a distinctive position by performing two interrelated but conceptually different roles, as illustrated in Figure 2.

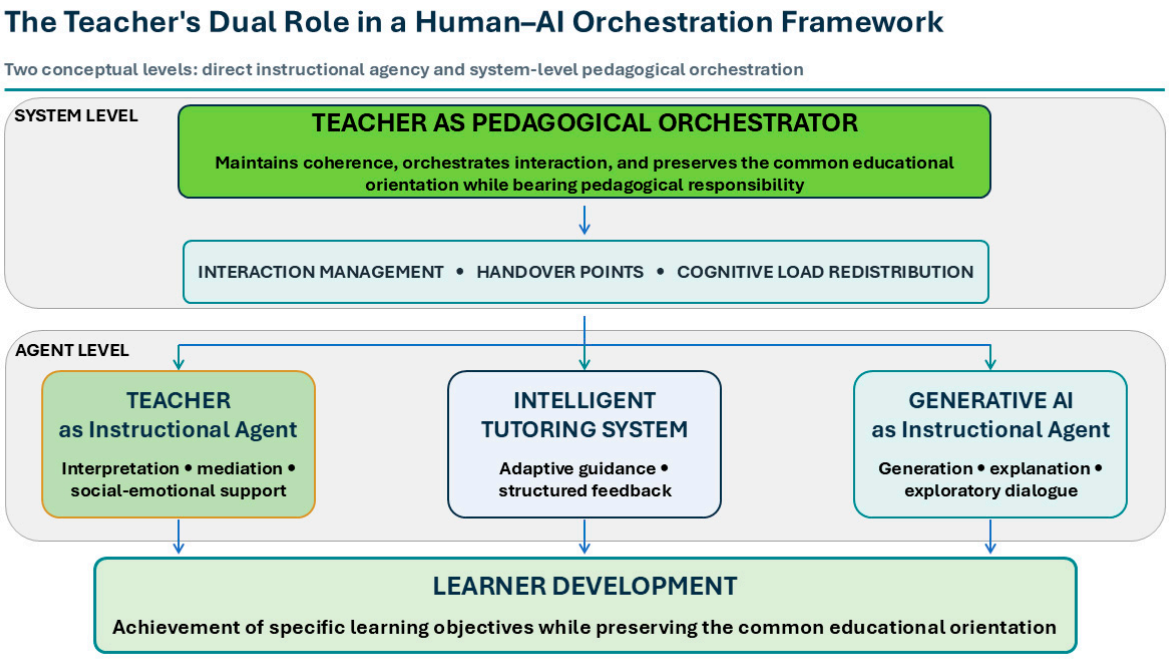


Figure 2. The Teacher's Dual Role: System-Level Pedagogical Orchestration and Agent-Level Instructional Support within the Proposed Human–AI Orchestration Framework.

On the one hand, the teacher participates as an instructional agent, providing direct pedagogical support to the learner. On the other hand, the teacher performs the role of a pedagogical orchestrator, bearing responsibility for organizing and coordinating the participation of the instructional agents, determining their place and purpose within the learning process, and structuring transitions between different forms of instructional support. In this way, the teacher's dual role distinguishes the teacher's direct participation in instruction from the responsibility for organizing and coordinating the Human–AI learning environment.

The different capabilities and roles of the instructional agents create potential for complementarity, but the mere existence of functional differences does not guarantee their pedagogically effective integration. Research on complementarity in Human–AI collaboration suggests that differences in information and capabilities between humans and artificial intelligence can constitute sources of complementarity and potentially improved joint performance; however, their existence alone does not automatically lead to positive outcomes and requires an appropriate mechanism for their effective use in collaborative activity (Hemmer et al., 2025). Within the proposed Framework, pedagogical orchestration performs precisely this coordinating function by organizing the different capabilities

and forms of instructional support as complementary resources in accordance with specific learning objectives and the learner's needs.

This organization preserves the common educational orientation of the learning process toward the learner's learning and development, regardless of which instructional agent performs the leading instructional function at a particular moment. Specific learning objectives may vary depending on the stage of education, the nature of the task, and the individual needs of the learner, but the coordination of different forms of instructional support remains subordinate to this common educational orientation. The mechanisms through which this coordination is implemented during the learning process are discussed in Section 4.

4. Mechanisms of Functioning of the Human–AI Orchestration Framework

The functioning of the Framework proposed in the preceding section is enabled through interconnected mechanisms and pedagogical interactions that coordinate the participation of its main components within a coordinated Human–AI learning environment. This section examines Handover Points, through which transitions between different forms of instructional support and the leading instructional functions of instructional agents are structured; the Meta-Prompt Framework, through which the interaction of Generative AI with the learner is structured and pedagogically guided; and teacher mediation and pedagogical orchestration, through which the teacher provides direct pedagogical support and coordinates the participation of the different components according to the specific learning situation and the learner's needs. Although these elements perform different functions, they are interconnected and jointly ensure coordinated interaction within the Human–AI learning environment.

4.1. Handover Points: Coordination and Functional Limitations

A key element of the proposed Human–AI Orchestration Framework is the Handover Point (HP), introduced here as a framework-specific conceptual and operational coordination mechanism for identifying pedagogically meaningful transitions in the leading form or function of instructional support. A Handover Point may occur in response to the learner's needs, the development of the task, the functional affordances and limitations of the available forms of support, or the teacher's pedagogical judgment. In this way, the Framework does not regard AI as an autonomous substitute for the teacher, but as part of a coordinated Human–AI learning environment in which human participants and AI components contribute different cognitive and functional capabilities that may provide complementary forms of instructional support.

The proposed concept of Handover Points is informed by a growing line of research on Human–AI hybrid adaptivity, shared control, and classroom co-orchestration. Previous work has examined how human and AI capabilities can complement one another within adaptive educational systems (Holstein et al., 2020), how dynamic transitions between different learning configurations can be organized in Human–AI co-orchestrated classrooms (Yang et al., 2023), and how teachers conceptualize sharing control with AI-supported orchestration tools (Lawrence et al., 2024). Building on these developments while retaining pedagogical orchestration and responsibility with the teacher, Handover Points conceptualize pedagogically significant transitions between different forms and sources of instructional support and between the leading instructional functions of different instructional agents. These transitions are considered in relation to the learner's immediate needs, the development of the learning task, and the functional capabilities and limitations of the participating instructional agents. Accordingly, Handover Points do not

introduce the general idea of dynamic transitions or shared control as a new theoretical construct. Their framework-specific contribution lies in providing a conceptual and operational mechanism for identifying and coordinating pedagogically meaningful shifts in the leading form or function of instructional support within the proposed Human–AI learning environment.

Within this configuration, Intelligent Tutoring Systems (ITSs) may provide structured problem-solving practice, immediate feedback, and adaptive guidance (Aleven et al., 2017; Koedinger et al., 2013; VanLehn, 2011), while Generative AI may provide flexible natural-language explanations and Socratic questioning to support conceptual reasoning, reflection, and metacognitive engagement (Holmes & Tuomi, 2022; Kasneci et al., 2023). These functions represent potential areas of instructional emphasis rather than fixed or mutually exclusive roles. Alongside these capabilities, the two types of AI systems have different functional limitations that are particularly significant in mathematics education and may require a change in the form or source of instructional support.

A representational difficulty may arise when the form of support currently available to the learner does not sufficiently connect mathematical meaning with the visual, symbolic, or spatial representations required by the learning task (Ainsworth, 2006; Duval, 2006). Such situations may involve graphs, geometric constructions, symbolic transformations, or other forms of mathematical representation. Within the proposed Framework, this potential mismatch is conceptualized through the Visual–Conceptual Gap (VCG). The VCG does not imply an inherent limitation of Generative AI or any other particular instructional agent; rather, it identifies a possible mismatch between the support currently being provided and the representational demands of the learning situation. Such a mismatch may signal the need for a different or complementary form of instructional support.

Handover Points can be regarded as coordination mechanisms that structure transitions between different forms of support and allow the leading instructional function to be assumed by the agent whose capabilities are considered appropriate for the specific learning situation. The transition is not necessarily unidirectional and is not limited to transferring the leading function from an AI component to the teacher. Depending on the nature of the task, the development of the learning situation, and the learner's needs, it may occur between different AI components, from an AI component to the teacher, or from the teacher to an appropriate AI component.

When the learner's needs, the development of the task, or the functional affordances and limitations of the current form of support indicate that direct teacher mediation is more appropriate, a Handover Point can structure a transition to teacher support. The teacher can then provide visual explanation, contextual interpretation, adaptive pedagogical support, or socio-emotional mediation when the particular AI component cannot provide them adequately in relation to the learner's needs and the specific learning situation. Once the difficulty has been addressed, the orchestration may again involve an appropriate AI component if its functional capabilities correspond to the next stage of the learning activity. In this way, Handover Points do not represent a final "handover" of the learning process from one agent to another but structure the dynamic redistribution of leading instructional functions within the coordinated interaction.

In this sense, Handover Points can be regarded as a mechanism through which the different functional capabilities of the instructional agents are used in a pedagogically purposeful manner. The differences among the teacher, ITS, and Generative AI are not eliminated; rather, their specific capabilities are engaged according to the particular learning objectives, the development of the learning situation, and the learner's needs. In this way, the dynamic redistribution of leading instructional functions enables different forms

of instructional support to be used as complementary resources while preserving the common educational orientation of the process toward the learner's learning and development.

This understanding is consistent with contemporary discussions in mathematics education in which orchestration is viewed as the structured regulation of AI pedagogical agency within coordinated learning systems (Trouche & Drijvers, 2014). It is also supported by research emphasizing the importance of human mediation in AI-supported learning environments, particularly when mathematical understanding requires the integration of verbal, symbolic, and visual forms of representation and their contextual interpretation (Ainsworth, 2006; Duval, 2006). By accounting for the functional boundaries of individual AI components and providing for structured transitions between different forms of support, the Framework is designed to support the integration of technological tools while maintaining the coherence of the pedagogical process.

To provide an initial operational basis for identifying potential Handover Points, the Framework proposes several observable indicators, including: (1) a shift in instructional discourse from procedural to visual–spatial representation, for example, when the learner or Generative AI refers to a graph, diagram, or geometric construction that cannot be adequately represented through the current form of interaction; (2) repeated unresolved requests for clarification directed to ITS or Generative AI within a limited period of interaction; (3) a mismatch between a mathematically valid learner response and the response format expected by the ITS; or (4) an explicit learner-initiated request for teacher assistance, for example, through the "Call the Teacher" mechanism described in Section 6, which may indicate the need for a pedagogically meaningful transition in the leading form or function of instructional support. These indicators constitute an initial and non-exhaustive operational basis for identifying potential Handover Points and are intended to support, rather than replace, the teacher's pedagogical judgment.

4.2. The Meta-Prompt Framework and Socratic AI Mediation

For Generative AI to perform its designated function within the coordinated Human–AI learning environment, its interaction with the learner needs to be pedagogically structured in accordance with the specific learning objectives and the nature of the learning situation. Unlike Intelligent Tutoring Systems (ITSs), in which a substantial part of the interaction logic is predefined within the architecture and functionalities of the particular system, Generative AI enables more flexible and open-ended natural-language interactions. This flexibility expands the possibilities for conceptual explanation and dialogic reasoning, but at the same time, creates a need for the interaction between Generative AI and the learner to be pedagogically guided so that its participation corresponds to its intended role in the learning process.

Within the proposed Framework, this function is performed through the Meta-Prompt Framework, a pedagogical configuration mechanism that specifies how Generative AI is instructed to provide educational support. Rather than constituting a new theoretical construct, the Meta-Prompt Framework operationalizes pedagogical principles for AI-supported dialogue through a structured set of predefined instructions. Its purpose is not to predetermine the content of each individual response, but to establish the overall pedagogical logic of the interaction. Rather than functioning primarily as a provider of ready-made solutions, Generative AI is guided toward the role of a Socratic mediator, supporting conceptual understanding through guiding questions, reflective prompts, and graduated scaffolding.

This approach builds on prior research emphasizing the importance of dialogic interaction and metacognitive engagement in the learning process (Holmes & Tuomi, 2022; Kasneci et al., 2023). When the interaction is pedagogically structured, Generative AI can

support exploratory reasoning by encouraging learners to articulate their thinking, justify the strategies they employ, consider alternative approaches, and reflect on the mathematical principles underlying the solution. In the absence of such structuring, however, Generative AI may shift toward providing direct answers or overly procedural guidance, which may limit opportunities for independent reasoning and conceptual understanding.

To address this challenge, the Meta-Prompt Framework proposes three core principles for regulating the interaction between Generative AI and the learner. The first is reasoning-based neutrality, whereby Generative AI is guided to support learners in justifying their own solution strategies rather than prescribing a specific method. The second is graduated scaffolding, whereby support progresses from open-ended questions toward increasingly specific guidance depending on the development of the problem-solving process and the difficulties encountered. The third is conceptual debriefing, whereby the problem-solving process concludes with reflection on the mathematical meaning of the obtained result, the approach used, and the underlying relationships manifested during the problem-solving process.

The preliminary pedagogical configuration of Generative AI through these principles cannot encompass all possible learning situations. The Meta-Prompt Framework structures the interaction between Generative AI and the learner within the boundaries of its functional role and capabilities, but it does not eliminate the limitations of the particular AI component or the need for pedagogical judgment during the learning process. In particular, pedagogical configuration through a Meta-Prompt cannot eliminate the possibility of inaccurate, inconsistent, or contextually inappropriate Generative AI outputs; teacher oversight and pedagogical judgment therefore remain necessary. When the development of the interaction indicates that the current mode of support no longer corresponds to the nature of the task or the learner's needs, adaptation of the interaction or pedagogical reconfiguration of Generative AI may be required. When the learner's needs, the development of the task, or the functional affordances and limitations of the current form of support indicate that another form of support is more appropriate, a transition through a Handover Point may occur, as described in Section 4.1.

In this way, the Meta-Prompt Framework and Handover Points perform different but interrelated functions within the proposed Orchestration Framework. The Meta-Prompt Framework pedagogically structures the interaction between Generative AI and the learner and enables its adaptation within the capabilities of the AI component, whereas Handover Points structure transitions between different forms of support and the leading instructional functions of instructional agents when the development of the learning situation requires such a change. Thus, the pedagogical configuration and adaptation of Generative AI do not replace the dynamic orchestration of the learning process but constitute part of it.

In this sense, the Meta-Prompt Framework constitutes part of the broader pedagogical orchestration of the Human–AI learning environment rather than an isolated strategy for using Generative AI. Through the preliminary structuring of the interaction and the possibility of its subsequent adaptation, it is designed to enable Generative AI to support the learner's conceptual reasoning and reflection within a predefined pedagogical logic without requiring direct teacher management of every individual learner–AI interaction. This configuration is intended to support the alignment of Generative AI participation with the specific learning objectives and the common educational orientation of the coordinated Human–AI learning process.

4.3. Teacher's Orchestration and Mediation

As defined in Section 3.2, the teacher is a human participant who performs two interconnected but conceptually distinct roles within the coordinated Human–AI learning

environment—as a pedagogical orchestrator and as an instructional agent. These functions also define two different modes of the teacher's participation in the learning process. As a pedagogical orchestrator, the teacher organizes and coordinates the participation of the instructional agents in accordance with the specific learning objectives, the nature of the learning situation, and the learner's needs. As an instructional agent, the teacher participates directly in the instructional interaction and provides pedagogical mediation where and when it is needed to support the learner's learning and development.

The teacher's orchestration function is not limited to the direct management of individual interactions between the learner and the AI components. It includes the prior organization of the Human–AI learning environment through the selection of appropriate technological tools, the determination of their place and purpose in relation to specific learning objectives, and the pedagogical configuration of components that allow such adaptation. During the learning process, the teacher monitors the development of the interaction and, when necessary, may modify the organization of support, adapt or reconfigure the relevant AI component, or assume a direct instructional function. In this way, pedagogical orchestration encompasses both the prior organization of the learning environment and its dynamic adaptation in response to the development of the specific learning situation.

This role may involve different approaches to the individual AI components, depending on their particular architecture and affordances. In Intelligent Tutoring Systems (ITSs), a significant part of the interaction logic and adaptation capabilities is determined by the architecture and functionalities of the particular system. Therefore, the teacher's orchestration role is primarily associated with selecting an ITS that corresponds to the specific learning objectives and needs and determining its place within the learning process. With Generative AI, the greater flexibility of interaction allows for additional pedagogical configuration and subsequent adaptation, which, within the proposed Human–AI learning environment, is supported through the Meta-Prompt Framework described in Section 4.2.

When the teacher participates directly as an instructional agent, the teacher's mediation may perform functions that complement the capabilities of the AI components. In mathematics education, this may include visual–spatial mediation, through which verbally or abstractly represented mathematical relationships are connected with graphs, diagrams, geometric constructions, or symbolic transformations; conceptual–contextual mediation, through which mathematical ideas are connected with the learner's prior knowledge, intuitive understanding, and specific context; and affective–motivational mediation, aimed at sustaining engagement, confidence, and persistence when difficulties arise. The importance of such human mediation is particularly evident in situations in which mathematical understanding requires coordination among verbal, symbolic, visual, and conceptual forms of representation (Ainsworth, 2006; Duval, 2006; Pane et al., 2017).

The teacher's two roles do not operate in isolation from one another. Pedagogical judgment exercised as part of orchestration may lead to direct teacher mediation when the development of the learning situation requires it. As described in Section 4.1, such a transition may be structured through a Handover Point. Once the specific difficulty has been addressed, the teacher may again organize a transition to an appropriate AI component or, when necessary, adapt the manner of its participation. Teacher mediation, therefore, does not represent an endpoint of Human–AI interaction, but rather one of the possible forms of pedagogical support within the dynamically orchestrated learning process.

In this way, the teacher does not perform the role of a dispatcher who continuously manages every individual interaction between the learner and the technological components. As a pedagogical orchestrator, the teacher organizes in advance the conditions for

their participation, monitors the development of the learning process, and intervenes when the pedagogical situation requires it—by modifying the organization of the interaction, adapting an appropriate AI component, or providing direct pedagogical mediation. This dual role enables the teacher to coordinate the different functional capabilities of the instructional agents as complementary forms of instructional support while preserving the common educational orientation of the Human–AI learning process toward the learner's learning and development.

Although this orchestration model is designed to avoid requiring continuous teacher management of every learner–AI interaction, it may also introduce additional demands related to the selection and configuration of technological tools, the evaluation of AI-generated support, and decisions about when direct teacher mediation is pedagogically appropriate. Teacher workload was not systematically measured in the exploratory design pilot; therefore, the extent to which the Framework increases, redistributes, or reduces teacher workload remains an empirical question for future research.

5. Illustrative Scenario: Coordinated Human–AI Orchestration in Practice

To illustrate the functioning of the proposed Human–AI Orchestration Framework in practice, this section presents a classroom scenario from algebra instruction. The scenario demonstrates how different instructional agents can provide complementary forms of instructional support within a coordinated learning process and how Handover Points can facilitate transitions between them according to the development of the learning situation and the learner's needs.

The sequence presented below represents one illustrative orchestration trajectory rather than a prescribed progression among ITS, Generative AI, and teacher support; other sequences may emerge according to the learner's needs, the development of the task, and the pedagogical context.

For the purposes of the presented scenario, the human participants are the teacher and the learner. The teacher also performs the function of the human instructional agent, while Intelligent Tutoring Systems (ITSs) and Generative AI function as AI instructional agents. Teacher mediation refers to the direct pedagogical support provided by the teacher in the role of an instructional agent, whereas the teacher's function as a pedagogical orchestrator encompasses the organization and coordination of interaction within the Human–AI learning environment.

5.1. Scenario: The Delivery Problem

Students are presented with a real-world algebra task involving two delivery companies with different pricing structures. Company A charges a fixed fee of 30 new Israeli shekels (NIS) plus 4 NIS per kilometer, whereas Company B charges a fixed fee of 10 NIS plus 9 NIS per kilometer. The task requires students to construct and solve a system of linear equations, interpret the meaning of the intersection point, and analyze how differences in slopes and intercepts influence cost comparisons across distances.

5.2. Illustrative Procedural Support Through ITS

In this illustrative trajectory, students first work on translating the verbal problem situation into an algebraic representation. During this formulation stage, the learning environment can provide complementary support through the mini-ITS, Generative AI, or direct teacher mediation, according to the learner's needs. The mini-ITS can provide structured guidance and feedback on the formulation of the equations, while Generative AI can support interpretation of the verbal relationships through guided questioning and

conceptual dialogue. Teacher mediation remains available when additional contextual or representational support is required. Once the relationships have been formulated algebraically, the resulting system can be represented as follows:

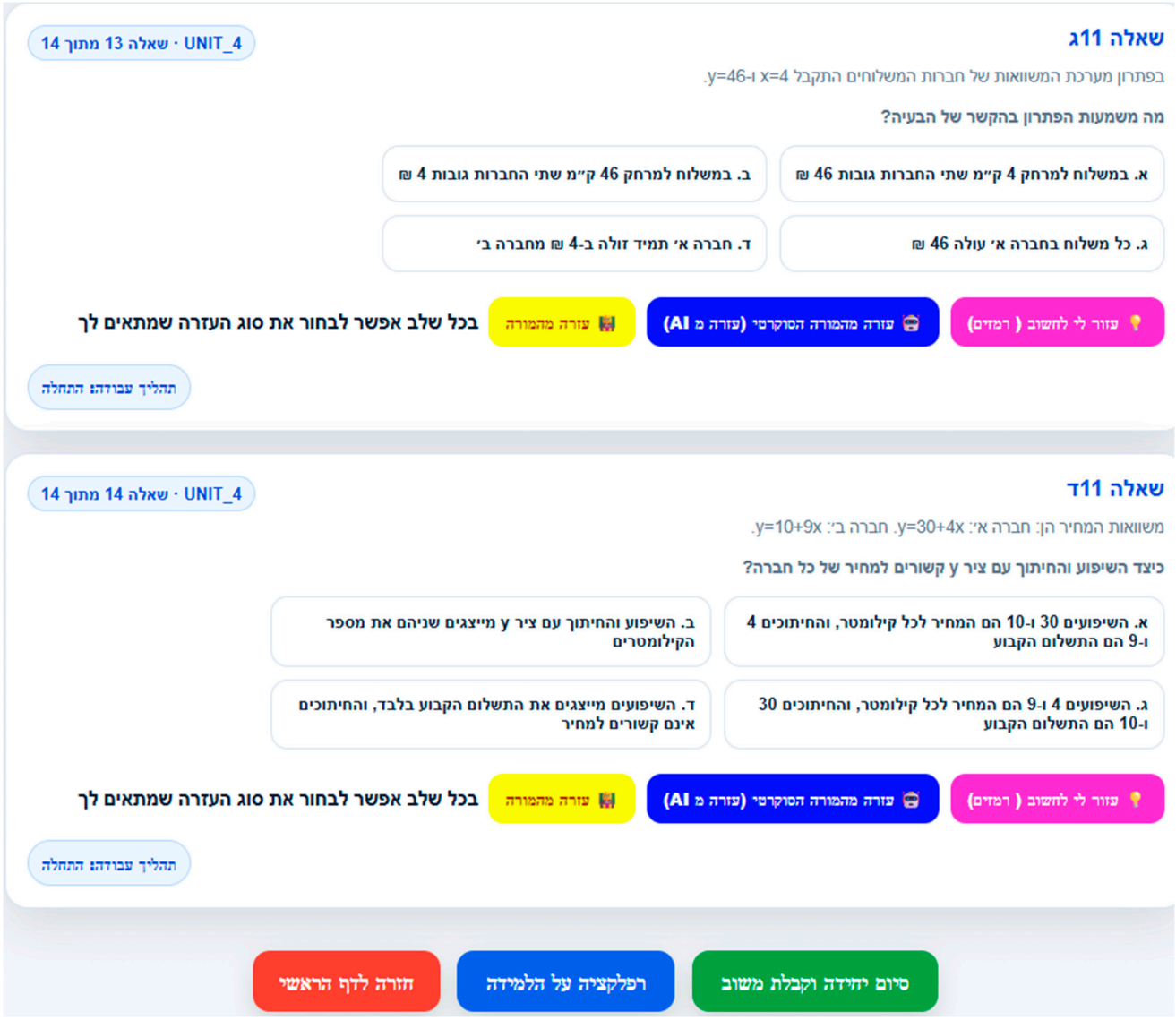
$$y = 4x + 30$$

$$y = 9x + 10$$

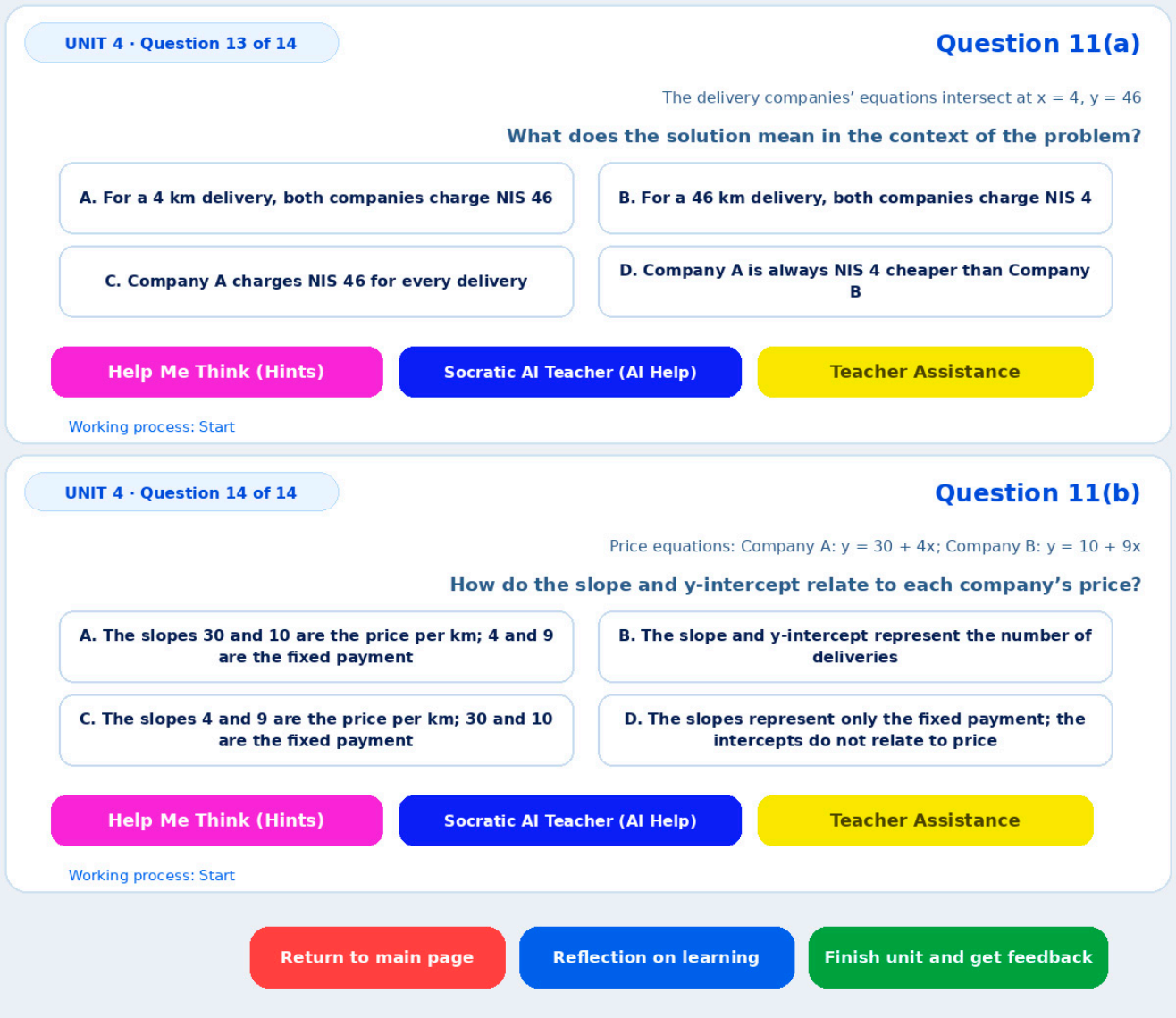
The ITS is designed to provide step-by-step guidance, immediate feedback, and adaptive hints, enabling students to solve the system and obtain the solution (x = 4, y = 46). At this stage, the system structures the procedural demands of the problem-solving process through step-by-step guidance and feedback, consistent with principles of instructional support discussed in the ITS and Cognitive Load Theory literature (Aleven et al., 2017; Koedinger et al., 2013; Sweller, 2011; VanLehn, 2011).

Figure 3 provides an illustrative view of the Delivery Problem as implemented in the researcher-developed Human–AI learning environment. The interface enables the learner to work on the mathematical task while accessing the available forms of support, including structured hints, Guided Socratic Generative AI support, and direct teacher assistance.

A. Original Hebrew interface (authentic pilot screenshot)



B. English translation (reconstructed for publication)



(Translation prepared for editorial review; the original Hebrew screenshot documents the actual pilot interface.)

Figure 3. Illustrative implementation of the Delivery Problem in the researcher-developed Human–AI learning environment, showing the mathematical task and the available support options.

5.3. Illustrative Conceptual Mediation Through Generative AI

Following the procedural solution, the Generative AI component engages students in conceptual dialogue. Using Socratic questioning aligned with the Meta-Prompt Framework (Section 4.2), AI encourages students to interpret the solution's meaning. For example, the AI may ask: "What does the point (4, 46) represent in the context of the problem?" or "How would these pricing models appear if represented graphically?"

At this stage, the AI can support conceptual reasoning, reflection, and metacognitive engagement through natural-language interaction (Holmes & Tuomi, 2022; Kasneci et al., 2023). However, as the discussion shifts toward graphical interpretation, a mismatch may arise between the form of support currently being provided and the learner's need for an adequate visual representation of the intersecting linear relationships.

5.4. Illustrative Teacher Intervention at the Handover Point

In this scenario, when the current AI-supported interaction no longer adequately corresponds to the learner's need for graphical interpretation, the resulting pedagogically meaningful transition in the leading form or function of instructional support may constitute a Handover Point. At this stage, the leading instructional function shifts from the AI component to the teacher, who participates directly as the human instructional agent.

The teacher introduces a graphical representation of the two linear functions and guides students in connecting the algebraic solution to its geometric meaning. Through visual explanation and contextual mediation, the teacher clarifies that the point (4, 46) indicates that, at a distance of 4 km, both delivery options cost 46 NIS. The teacher may further extend the discussion by asking students to interpret the slopes (cost per kilometer) and intercepts (fixed fees) within the real-world context of the problem.

5.5. Integrated Cognitive Process

This scenario illustrates how different instructional agents may provide complementary forms of support within a coordinated learning process. In the particular trajectory presented here, the ITS provides structured procedural support, Generative AI supports conceptual articulation and reflective reasoning, and the teacher provides visual representation, contextual grounding, and adaptive support through direct pedagogical mediation. These functions are not intended to represent fixed or mutually exclusive roles; rather, the leading form or function of support may vary according to the learner's needs and the development of the learning situation. The coordinated interaction among these instructional agents is intended to support procedural accuracy, conceptual reasoning, and representational understanding within an integrated instructional process.

5.6. Applicability to Additional Contexts

A similar orchestration pattern may also be applied in other algebraic contexts, including classroom-based word problems. For example, we have the following:

"There are 32 students in a class. The number of girls is 4 more than the number of boys."

If x represents the number of boys and y represents the number of girls, the situation is modeled as follows:

$$x + y = 32$$

$$y = x + 4$$

Students sometimes make a common conceptual error by reversing the relationship and writing:

$$x = y + 4$$

This incorrect formulation may suggest that the number of boys exceeds the number of girls. In such cases, an ITS may provide feedback on the structure of the system of equations, while the underlying linguistic–conceptual misunderstanding involved in interpreting the relational statement may require a different or complementary form of instructional support.

Generative AI may guide students through verbal reasoning and reflective questioning; however, if misunderstanding persists, the situation may require a structured transition through a Handover Point. In the illustrative trajectory presented here, the teacher may then mediate understanding by connecting the symbolic representation to a concrete

classroom context, for example, by physically grouping students to demonstrate the relationship visually. This mediation can help students bridge the gap between verbal description and symbolic representation.

This additional example further illustrates the potential applicability of the proposed Framework beyond a single task type and across different forms of mathematical reasoning, while its broader generalizability remains to be examined through further empirical research.

6. Exploratory Design Pilot of the Human–AI Orchestration Framework

This section presents an exploratory design pilot of the Human–AI Orchestration Framework conducted under authentic classroom conditions. The purpose of the pilot was not to evaluate learning outcomes or the educational effectiveness of the Framework, but to examine its practical implementation and to identify design requirements, functional limitations, and necessary refinements before subsequent systematic empirical evaluation. The pilot focused on whether the proposed forms of support could be integrated within a coordinated classroom environment, whether the pedagogical configuration and coordination mechanisms could be enacted during instruction, and what implementation difficulties emerged in practice. Observations and feedback collected during the pilot were used formatively to refine the specific Human–AI learning environment and the practical implementation of individual mechanisms of the Framework.

The exploratory design pilot was conducted over a six-week period with 32 male Grade 8 students, approximately 13 years of age, from a single mathematics class in a public, state-funded religious boys' school in Haifa, Israel. The participating class constituted the school's highest-level mathematics grouping for Grade 8 and included a substantial proportion of the students in the grade. The participating group was an intact, existing high-achieving mathematics class taught by the teacher-researcher. The entire class participated in the exploratory pilot because systems of linear equations in two variables, the mathematical topic addressed in the pilot, formed part of the Grade 8 curriculum being taught during the implementation period. A pre-test was administered before the pilot; however, pre-test performance was not used as a selection criterion, and all students in the class participated regardless of their pre-test scores. The class therefore constituted a convenience sample within the authentic instructional context. The participating students had prior experience with commonly used digital learning environments, including a Ministry of Education ITS-based learning environment, and regularly used standard, open-ended ChatGPT for homework, assignments, and general consultation. However, they had not previously used the researcher-developed Guided Socratic configuration employed in the present pilot. Students accessed the learning environment using school computers. The classroom teacher also served as the researcher conducting the exploratory design pilot. The teacher-researcher had 26 years of experience teaching mathematics and held a master's degree from the Hebrew University of Jerusalem, Israel. She had prior experience using ChatGPT and other commonly available AI tools, including different versions available over time, for consultation and professional purposes, but had not previously used Generative AI as a classroom instructional tool or for Socratic AI-supported learning. All instructional activities involving students were conducted in accordance with the applicable institutional and ethical requirements governing educational research. The pilot was carried out with the required institutional permissions, including authorization from the Israeli Ministry of Education, approval from the participating school administration, and informed parental consent obtained prior to the classroom implementation. No personally identifiable information or sensitive participant data are presented in this paper.

The initial classroom implementation highlighted a practical problem related to the linguistic accessibility of existing ITS platforms. During the initial use of the ASSISTments platform, the English-language interface was found to introduce additional linguistic demands for the Hebrew-speaking students during algebraic problem-solving activities. This raised concerns about the practical suitability of a non-localized ITS environment for the particular group of learners and instructional context and led to the decision to develop and use a Hebrew-language rule-based mini-ITS aligned with the linguistic and curricular characteristics of the Israeli educational context. Thus, the selection and, where necessary, replacement or development of an appropriate ITS became part of the practical configuration of the specific Human–AI learning environment.

The implementation focused on systems of linear equations in two variables. Over the six-week pilot, students completed seven instructional units, each comprising approximately 90 min of instruction across two class periods. The seven units followed a cumulative mathematical progression: Unit 1 introduced the graphical solution of systems of linear equations; Unit 2 extended the graphical approach to applications of systems of equations; Unit 3 introduced the substitution method; Unit 4 focused on verbal problems modeled and solved using substitution; Unit 5 introduced the elimination method; Unit 6 addressed non-standard-form systems and systems containing parentheses; and Unit 7 extended the sequence to systems of linear equations involving parameters. Across the seven units, the mathematical tasks progressed from graphical and representational understanding toward algebraic solution methods, modeling, increasingly complex equation structures, and parameter-based reasoning. The learning environment comprised a Hebrew-language rule-based mini-ITS, Generative AI configured to provide guided Socratic support, and direct teacher mediation. The mini-ITS followed a predefined instructional sequence rather than dynamically adapting the learning pathway to individual students. It provided rule-based evaluation of student responses and access to structured hints during problem solving. When a student submitted two incorrect responses to the same question, the system presented a new, similar question, providing an additional opportunity to work on the same mathematical idea. Thus, adaptation within the mini-ITS was limited to predefined response rules rather than individualized algorithmic adaptation. The mini-ITS was developed by the teacher-researcher in collaboration with Moshe Malik as part of the same PHP-based web environment that integrated the Generative AI component. The teacher-researcher designed the mathematical tasks using textbooks approved by the Israeli Ministry of Education and defined the mathematically correct responses and the rule-based evaluation criteria, while Malik implemented these specifications programmatically. The environment supported different response formats, including open-ended input and multiple-choice responses, and automatically evaluated student responses according to the predefined rules. The mini-ITS was developed as a functional instructional component for the exploratory design pilot and did not undergo a separate formal technical validation process; implementation difficulties identified during classroom use were instead used to inform iterative refinement of the system. These forms of support were available as complementary instructional resources rather than as a prescribed sequential pathway, and transitions in the leading form or function of support could occur according to the learner's needs, the development of the task, and the pedagogical context. Illustrative screenshots of the implemented learning environment and its support mechanisms are provided in Appendix C.

Several iterative refinements were introduced during the pilot in response to difficulties observed during classroom implementation. The initial Meta-Prompt configuration relied primarily on open-ended Socratic questioning and avoided providing complete solutions or final answers. During the pilot, however, this configuration was found to pro-

vide insufficient guidance for some students when they encountered conceptual difficulties or remained unable to progress through questioning alone. In response, the Meta-Prompt was progressively refined toward Guided Socratic Scaffolding, in which questioning and reflection remained central but could be supplemented, when needed, by conceptual cues, relevant mathematical relationships, strategic suggestions, and more explicit guidance toward the next step. This iterative refinement illustrates how the pedagogical configuration of Generative AI support was adapted in response to difficulties observed during authentic classroom implementation.

Generative AI support during the pilot was implemented through the OpenAI API and integrated into a researcher-developed web-based learning environment programmed in PHP. Students interacted with the Generative AI component entirely through this environment and did not require individual ChatGPT accounts or access to the standard ChatGPT interface. The pedagogical behavior of the Generative AI component was configured through a researcher-developed Meta-Prompt containing detailed instructions for Guided Socratic interaction. In addition, depending on the type of mathematical question, question-specific hidden instructions were appended to the student's query before it was sent to the API in order to guide the form and pedagogical focus of the generated response. The Meta-Prompt was used to structure the pedagogical characteristics of the interaction, including Socratic questioning, constraints on the provision of complete solutions, and the progressive introduction of more guided scaffolding when needed. Because the Meta-Prompt was iteratively refined during the pilot, the final configuration should not be interpreted as the exact configuration used throughout all instructional sessions. The final refined Meta-Prompt resulting from this iterative process is provided in Appendix A. To protect student privacy, student names were not transmitted to the OpenAI API; only the text submitted for Generative AI consultation was sent to the external service. Within the researcher-developed environment, each student was represented by a unique identifier. Student–AI interactions were conducted in Hebrew. No external plugins were used. The learning environment was web-based and required Internet access; however, the application itself did not restrict students from opening additional browser tabs. The specific underlying ChatGPT model version used during the pilot was not systematically recorded and therefore cannot be retrospectively specified with certainty.

During the pilot, a dedicated "Call the Teacher" feature was introduced to enable learners to directly request teacher assistance when they encountered difficulty or conceptual misunderstanding. The feature provided an explicit mechanism for initiating a transition to direct teacher support. When such a request resulted in a pedagogically meaningful shift in the leading form or function of instructional support, the transition could be identified as a Handover Point. The teacher could then provide direct pedagogical mediation, and based on pedagogical judgment, determine how support should subsequently be organized. This could involve continuing direct teacher mediation, returning to one of the AI-supported resources, or combining different forms of support according to the learner's needs. In this way, the "Call the Teacher" feature operationalized a learner-initiated pathway for requesting a change in instructional support without treating the learner's action itself as a Handover Point.

Subsequent observations also revealed difficulties associated with open-ended mathematical responses within the mini-ITS environment. In some cases, mathematically correct responses were not recognized by the system because of variations in symbolic formatting or the structure of the mathematical expression. This highlighted a practical limitation of the rule-based mini-ITS and the need to balance opportunities for open-ended mathematical reasoning with more structured response mechanisms. Consequently, selected multiple-choice response formats were introduced alongside open-ended activities

in the third and fourth instructional units to improve interaction stability and reduce difficulties associated with answer recognition.

Observations and formative feedback collected during the exploratory design pilot informed several iterative refinements to the Human–AI learning environment and to the practical implementation of individual mechanisms of the Framework. The main implementation observations, their design interpretation, and the resulting refinements are summarized in Table 1.

Table 1. Key Observations and Iterative Design Refinements during the Exploratory Design Pilot.

Implementation Aspect	Pilot Observation	Design Interpretation:	Resulting Refinement
Socratic AI interaction	Some students experienced difficulty progressing when the initial Socratic interaction relied primarily on open-ended questioning.	Open-ended questioning alone did not always provide sufficient guidance when learners encountered conceptual difficulty.	Progressive refinement of the Meta-Prompt toward Guided Socratic Scaffolding, incorporating conceptual cues, strategic suggestions, and graduated support when needed.
Teacher mediation	Students occasionally required direct teacher assistance when the available forms of support did not adequately address their difficulty or conceptual misunderstanding.	The learning environment required an explicit mechanism through which learners could request a transition to direct teacher support when needed.	Development and integration of the "Call the Teacher" feature, enabling learners to directly request teacher assistance and initiate a potential transition in the leading form of instructional support.
Open-ended mathematical responses	In some cases, mathematically correct responses were not recognized by the mini-ITS because of variations in symbolic formatting or expression structure.	The rule-based answer-recognition mechanism required a more structured response format for selected tasks to ensure reliable recognition of mathematically valid answers.	Introduction of selected multiple-choice response formats alongside open-ended activities in the third and fourth instructional units.
Student feedback	Students completed detailed written reflections at the end of each instructional unit.	The reflections provided formative indications of students' experiences with the available forms of support and highlighted aspects of the learning environment requiring further attention.	Student feedback was used as an additional formative source to inform the ongoing refinement of the AI-supported interactions and orchestration mechanisms.
Classroom implementation	Students required different forms of support at different points during the instructional activities.	A fixed sequence of ITS, Generative AI, and teacher support did not adequately reflect the varying support needs observed during classroom implementation.	The learning environment was refined to treat the mini-ITS, Generative AI, and direct teacher mediation as complementary resources that could be accessed according to the learner's needs and the development of the learning situation.
Linguistic accessibility of the ITS	The English-language interface of ASSISTments introduced additional linguistic demands for the Hebrew-speaking students during algebraic problem-solving activities.	A non-localized ITS was not sufficiently aligned with the linguistic and curricular context of the participating students.	Replacement of ASSISTments with a Hebrew-language rule-based mini-ITS aligned with the linguistic and curricular characteristics of the classroom context.

The observations from the exploratory design pilot are preliminary and are not presented as formal empirical findings regarding the educational effectiveness of the Framework. The pilot did not employ systematic outcome data collection designed for hypothesis testing or statistical evaluation of learning effects. However, formative documentation was collected throughout the implementation. In addition to the teacher and student reflections, the web-based learning environment automatically recorded timestamped student actions during the instructional activities. These records included question initiation, answer attempts and submissions, use of structured hints, activation of Generative AI support, requests for teacher assistance, and question completion. The logs provided process documentation of when different forms of support were accessed during students'

work and could therefore support the retrospective examination of potential Handover Points. These records were used formatively in the exploratory pilot and were not analyzed as evidence of learning effectiveness. Following each instructional session, the teacher-researcher completed a structured reflection documenting classroom events, difficulties encountered by students when working with the learning environment, problems related to response recognition, situations in which the available Socratic support was insufficient, patterns of interaction with the different forms of support, and issues requiring subsequent design refinement. Students also completed a detailed reflection at the end of each instructional unit, providing formative feedback on their learning experience, difficulties encountered, and use of the available forms of support. The teacher-researcher reflections and student reflections were reviewed formatively throughout the pilot and were used to identify implementation difficulties and inform the ongoing iterative refinement of the learning environment. These materials were not treated as validated outcome measures and were not subjected to inferential statistical analysis. Student reflections were used as formative design documentation and were not subjected to formal thematic analysis; accordingly, they are not presented as empirical evidence of students' perceptions or of the effectiveness of particular forms of support.

Taken together, the exploratory design pilot provides preliminary indications of the practical feasibility of implementing the proposed Framework as a coordinated Human–AI learning environment under authentic classroom conditions. The pilot also identified several implementation requirements and limitations that informed iterative refinement of the learning environment and its coordination mechanisms. These observations do not establish the educational effectiveness of the Framework or empirically validate its proposed conceptual mechanisms, which require systematic investigation in subsequent research.

7. Discussion

The present study proposes a Human–AI Orchestration Framework for Mathematics Education, which conceptualizes the AI-supported learning environment as a coordinated interaction between human participants and AI components. Rather than treating Intelligent Tutoring Systems (ITSs), Generative AI, and the teacher as independent or competing sources of instructional support, the Framework focuses on the possibility of organizing and coordinating the different and potentially overlapping forms of instructional support they can provide within a shared learning process.

From a theoretical perspective, this implies a shift in focus from the use of individual AI tools toward the design and management of coordinated Human–AI learning environments. Previous research has demonstrated the potential of ITSs to support procedural fluency (Aleven et al., 2017; VanLehn, 2011) and of Generative AI to support conceptual reasoning and dialogic interaction (Holmes & Tuomi, 2022; Kasneci et al., 2023). Recent studies have also begun to examine Human–AI hybrid adaptivity, shared control, and classroom co-orchestration (Holstein et al., 2020; Lawrence et al., 2024; Yang et al., 2023). Within this emerging research direction, the present Framework focuses more specifically on the pedagogical coordination of functionally different forms of instructional support, including ITSs, Generative AI, and direct teacher mediation, while pedagogical orchestration and responsibility remain with the teacher. The proposed Framework considers such coordination not as an additional organizational layer imposed on individual technologies, but as an essential characteristic of the Human–AI learning environment, in which different instructional agents may provide complementary and potentially overlapping forms of support according to the learner's needs, the learning situation, and the functional affordances of the available resources.

This perspective is consistent with research on pedagogical orchestration and human pedagogical agency in technology-supported learning environments (Dillenbourg, 2013; Kahn et al., 2025; Trouche & Drijvers, 2014).

A substantial element of the proposed Framework is the distinction between the teacher's role as a pedagogical orchestrator and the teacher's direct participation as a human instructional agent. As a pedagogical orchestrator, the teacher organizes the Human–AI learning environment in advance by selecting appropriate technological tools, determining their place and purpose in relation to specific learning objectives, and pedagogically configuring those components that allow such adaptation. During the learning process, the teacher observes the development of the interaction and, when necessary, may modify the organization of support, adapt or reconfigure the relevant AI component, or assume a direct instructional function. As a human instructional agent, the teacher provides direct pedagogical mediation when required by the nature of the task or the learner's needs. This distinction allows teacher mediation to be understood not as a response to an AI "failure," but as one of the complementary forms of instructional support within a dynamically orchestrated process.

Within this context, Handover Points function as a framework-specific coordination mechanism for structuring pedagogically meaningful transitions in the leading form or function of instructional support. They provide a way of conceptualizing shifts among different and potentially overlapping forms of support according to the learner's needs, the development of the task, and the affordances and limitations of the available resources. The exploratory design pilot of the Framework also highlighted an additional aspect of this mechanism through the "Call the Teacher" feature. In this case, the learner, who is a human participant in Human–AI interaction but does not perform the function of an instructional agent, may directly request a transition to teacher support when the current form of instructional support does not meet their needs. When such a request results in a pedagogically meaningful shift in the leading form or function of instructional support, the transition may be identified as a Handover Point. The learner's request does not itself constitute a Handover Point or participation in pedagogical orchestration. The decision regarding what form of support should subsequently be provided and how the interaction should be organized remains subject to the teacher's pedagogical judgment in the role of orchestrator. In this way, the "Call the Teacher" feature expands the learner's ability to signal the need for a change in instructional support without altering the teacher's central role in organizing and coordinating the Human–AI learning environment.

Another conceptual element of the proposed Framework is the Visual–Conceptual Gap (VCG), introduced here as a framework-specific conceptual lens rather than as an independently validated construct. Grounded in research on multiple representations and representational coordination (Ainsworth, 2006; Duval, 2006), the VCG refers to situations in which the form of support currently available to the learner does not sufficiently connect mathematical meaning with the visual, symbolic, or spatial representations required by the learning task. Such a mismatch is not treated as an inherent limitation of Generative AI or of any particular instructional agent, especially as the representational capabilities of AI systems continue to evolve. Rather, it may indicate the need for a different or complementary form of instructional support. Within the Framework, the VCG therefore provides a conceptual lens for considering when a change in the leading form or function of support may be pedagogically appropriate.

The Framework also draws on Cognitive Load Theory (Sweller, 2011) through Cognitive Load Redistribution (CLR), which is presented here as a framework-specific theoretical proposition rather than as an empirically demonstrated effect. CLR refers to the pedagogical coordination of supportive functions and cognitive operations across the

learner, technological components, and teacher according to the demands of the learning situation. It does not imply a literal transfer of cognitive load between agents or a fixed allocation of instructional functions. Within this perspective, ITS, Generative AI, and teacher mediation may provide different and potentially overlapping forms of support according to the demands of the task and the learner's needs. Because cognitive load was not empirically measured in the exploratory design pilot, the present study does not provide evidence that cognitive load was reduced or redistributed during implementation.

Within this context, Handover Points and the pedagogical configuration and adaptation of interaction with Generative AI through the Meta-Prompt Framework can be regarded as interrelated mechanisms for coordinating different and potentially overlapping forms of instructional support. This coordination is intended to enable the available resources to be used complementarily while preserving the common educational orientation toward the learner's learning and development.

The exploratory design pilot presented in Section 6 provides preliminary indications regarding the practical feasibility of implementing the proposed Framework under authentic classroom conditions. The iterative refinements introduced during the pilot—including the replacement of the initially used ITS with a localized mini-ITS, progressive refinement of the Meta-Prompt configuration, introduction of the "Call the Teacher" feature, and modification of selected response mechanisms—highlight several practical requirements associated with implementing a coordinated Human–AI learning environment. These refinements suggest that orchestration should not be understood as the one-time design of a fixed learning environment. Rather, its practical realization may require ongoing observation, pedagogical judgment, and iterative adaptation of the specific configuration in relation to learning objectives, the nature of the learning tasks, learners' needs, and the capabilities and limitations of the technological components being used. These observations should also be interpreted cautiously because alternative explanations cannot be excluded. The implementation difficulties and resulting refinements may reflect not only characteristics of the proposed Framework, but also the specific classroom context, learner group, technological configuration, and teacher-researcher involvement in the pilot. The exploratory design therefore does not allow these observations to be attributed specifically to the Framework or its individual mechanisms.

These observations do not constitute evidence of the educational effectiveness of the Framework, as the exploratory design pilot was not designed to evaluate learning outcomes, test hypotheses, or establish causal relationships. The preliminary indications of practical feasibility and the implementation requirements identified during the pilot therefore provide a basis for subsequent systematic empirical evaluation of the Framework and its proposed mechanisms.

For the same reason, the present study does not allow the generalizability of the proposed Framework to be established. The illustrative scenarios presented in Section 5 and the exploratory design pilot described in Section 6 encompass limited mathematical content, a particular age group, and a specific educational and technological context. Further empirical studies and an accumulation of evidence from applications of the Framework across different mathematical topics, learner groups, educational levels, and instructional contexts are therefore required. Such studies should examine not only whether the Framework can be applied under diverse conditions, but also the extent to which the proposed mechanisms retain their functionality, pedagogical coherence, and potential educational effectiveness as the context changes. In addition, the dual role of the classroom teacher as both teacher and researcher may have influenced the interpretation of classroom observations and formative reflections. This potential source of researcher subjectivity should be considered when interpreting the pilot observations.

An additional limitation arises from the dependence of the Framework on the characteristics of specific technological components. The capabilities of Generative AI and multimodal AI systems are developing rapidly, and functional limitations observed in current technologies may therefore be reduced or take different forms in future systems. The situations identified through the Visual–Conceptual Gap (VCG) lens may likewise change as multimodal capabilities develop, because emerging technologies may provide increasingly sophisticated forms of visual, symbolic, and multimodal support. However, such technological developments do not eliminate the need for pedagogical judgment regarding whether the form of support currently available adequately addresses the representational demands of the learning situation.

The implementation of the Framework also depends on the teacher's pedagogical and technological competence. Effective orchestration requires the ability to select appropriate technological tools, pedagogically configure adaptable AI components, recognize situations requiring a change in the form of support, and provide direct mediation when necessary. Differences in teacher preparation, digital competence, and professional experience may therefore influence both the practical implementation of the Framework and the outcomes associated with its use.

The Framework may also introduce additional demands on the teacher related to selecting and configuring technological tools, evaluating AI-generated support, and deciding when direct teacher mediation is pedagogically appropriate. At the same time, the availability of multiple forms of support may alter when and how direct teacher intervention is required during learner interactions. Because teacher workload was not systematically measured in the exploratory design pilot, however, the extent to which the Framework increases, redistributes, or reduces teacher workload remains an empirical question for future research.

From a practical perspective, the Framework offers a set of mechanisms and principles for designing coordinated Human–AI learning environments, but it does not prescribe a single universal technological configuration. The selection of ITS, the configuration of Generative AI, the forms of teacher mediation, and the organization of Handover Points should be aligned with specific learning objectives, learner characteristics, and available technological and organizational resources. This capacity for configuration and adaptation represents a potential advantage of the Framework, although the extent to which it can be successfully implemented across different conditions remains a subject for future empirical investigation.

Beyond the direct examination of the educational effectiveness and generalizability of the Framework, the present study also outlines several possible directions for further research. One concerns authentic assessment in AI-supported learning environments. Tasks requiring visual construction, contextual interpretation, conceptual explanation, or critical evaluation of AI-generated content may be investigated as a possible means of examining not only the learner's ability to integrate different forms of mathematical representation, but also their ability to understand, interpret, and evaluate the mathematical correctness and conceptual meaning of instructional support provided by AI. Another direction concerns educational equity and the question of whether coordinating different forms of AI-based and human support can create more accessible and adaptive conditions for learners with differing needs, prior knowledge, and levels of preparation. These possibilities were not systematically investigated in the present study and should not be interpreted as established effects of the proposed Framework, but rather as directions for subsequent empirical research.

In summary, the principal contribution of the proposed Framework is conceptual: it specifies a coordinated Human–AI learning architecture in which different and poten-

tially overlapping forms of instructional support can be organized while pedagogical responsibility remains with the teacher. Through Handover Points as a framework-specific coordination mechanism, the Meta-Prompt as a pedagogical configuration mechanism, and the distinction between the teacher's roles as pedagogical orchestrator and human instructional agent, the Framework provides a structured account of how such coordination may be conceptualized in mathematics education. The exploratory design pilot provides preliminary indications of practical feasibility under the specific classroom conditions examined, but it does not establish educational effectiveness or empirically validate the proposed conceptual mechanisms.

8. Conclusions

The present study proposes a Human–AI Orchestration Framework for Mathematics Education, which conceptualizes the AI-supported learning environment as a coordinated interaction between human participants and AI components. Rather than treating Intelligent Tutoring Systems (ITSs), Generative AI, and the teacher as independent or competing sources of instructional support, the Framework focuses on coordinating the different and potentially overlapping forms of support they can provide within a shared learning process, while pedagogical responsibility remains with the teacher.

A central role within the proposed Framework is assigned to the teacher, who performs two interrelated but conceptually distinct roles—as a pedagogical orchestrator and as a human instructional agent. As a pedagogical orchestrator, the teacher organizes and coordinates the Human–AI learning environment, while as a human instructional agent, the teacher provides direct pedagogical mediation when necessary. Handover Points function as a framework-specific coordination mechanism for structuring pedagogically meaningful transitions in the leading form or function of instructional support, while the Meta-Prompt Framework serves as a pedagogical configuration mechanism for structuring and adapting interaction with Generative AI. Cognitive Load Redistribution (CLR) is proposed as a theoretical proposition concerning the coordination of supportive functions and cognitive operations across the learner, technological components, and teacher, whereas the Visual–Conceptual Gap (VCG) serves as a conceptual lens for identifying situations in which the current form of support does not sufficiently address the representational demands of the learning situation.

The exploratory design pilot presented in Section 6 provides preliminary indications of the practical feasibility of implementing the proposed Framework under the specific classroom conditions examined. The need to select and replace technological components, refine the Meta-Prompt configuration, and introduce mechanisms such as the "Call the Teacher" feature illustrates that pedagogical orchestration is not a one-time configuration of a fixed learning environment, but may require ongoing observation, pedagogical judgment, and iterative adaptation. These formative observations identified practical implementation requirements and informed refinement of the Human–AI learning environment; however, they do not establish the educational effectiveness, generalizability, or empirical validity of the Framework's proposed conceptual mechanisms. They therefore provide a basis for subsequent systematic empirical evaluation rather than evidence of established operational viability.

More broadly, the study contributes to current discussions on AI in education by shifting attention from the capabilities of individual technologies toward the pedagogical coordination of multiple forms of human and AI support. Its conceptual contribution lies in specifying how such coordination may be structured in mathematics education while maintaining teacher pedagogical responsibility and allowing the configuration of support to respond to the learner's needs, the learning situation, and the affordances of the available resources.

Future research should systematically evaluate the educational effectiveness and generalizability of the proposed Framework across different mathematical topics, learner groups, educational levels, and instructional contexts. In this direction, future studies may include: (1) quasi-experimental or experimental studies comparing orchestrated Human–AI learning environments with configurations based on individual technologies or traditional teacher support in terms of procedural fluency, conceptual understanding, and representational reasoning; (2) longitudinal studies examining how sustained use of the Framework and Handover Points within the orchestrated learning process relates to learner autonomy, the development of conceptual understanding, and long-term knowledge retention; and (3) studies across different contexts to determine the extent to which the Framework and its mechanisms can be adapted across educational levels, mathematical topics, and resource conditions.

Alongside the evaluation of learning outcomes, an important direction for future research is the operationalization of the pedagogical orchestration process itself. Future studies may examine the conditions under which Handover Points arise, the different ways in which they can be initiated, the sequence of transitions between forms of instructional support, and the relationship between different orchestration patterns and the development of the learning process. Of particular interest are learner-initiated requests for teacher intervention, through which the learner may signal the need for a change in instructional support. Such a request does not itself constitute a Handover Point; rather, when it results in a pedagogically meaningful shift in the leading form or function of instructional support, the transition may be identified as a Handover Point, while the subsequent organization of support remains subject to the teacher's pedagogical judgment and orchestration function.

Further research may examine how coordinated Human–AI learning environments influence procedural fluency, conceptual understanding, representational reasoning, and learner engagement, as well as whether different forms of adaptive human and AI support can contribute to more accessible and equitable learning conditions for learners with different needs and levels of prior preparation. A separate promising direction concerns authentic assessment in AI-supported learning environments. Tasks requiring visual construction, contextual interpretation, conceptual explanation, or critical evaluation of AI-generated content may be investigated as a possible means of examining learners' ability not only to integrate different forms of mathematical representation, but also to understand, interpret, and evaluate the mathematical correctness and conceptual meaning of instructional support provided by AI.

From a broader perspective, future research may investigate how the coordination of different and potentially overlapping human and AI forms of instructional support within a common educational orientation influences learning processes, educational outcomes, and the evolving relationship between humans and AI in technology-supported learning environments.

Ultimately, the central question concerning AI integration in mathematics education is not simply whether Intelligent Tutoring Systems, Generative AI, and the teacher can participate within the same learning environment, but how different and potentially overlapping forms of human and AI instructional support can be pedagogically coordinated in response to the learner's needs and the development of the learning situation. By framing AI integration as a question of coordination rather than substitution, the Human–AI Orchestration Framework provides a conceptual basis for designing mathematics learning environments in which multiple forms of support can be coordinated while pedagogical responsibility remains with the teacher. The exploratory design pilot offers preliminary indications of the practical feasibility of this approach under the specific classroom

conditions examined, while its educational effectiveness and generalizability remain to be established through systematic empirical research.

Author Contributions: Conceptualization, G.P., A.S., M.M., N.P., V.T.-K., I.N., S.S. and E.D.; Methodology, G.P., A.S., M.M., N.P., V.T.-K., I.N., S.S. and E.D.; Software, G.P., M.M. and E.D.; Validation, G.P., A.S., M.M., N.P., V.T.-K., I.N., S.S. and E.D.; Formal analysis, G.P., A.S., M.M., N.P., V.T.-K., I.N., S.S. and E.D.; Investigation, G.P., A.S., M.M., N.P., V.T.-K., I.N., S.S. and E.D.; Resources, G.P., A.S., M.M., N.P., V.T.-K., I.N., S.S. and E.D.; Data curation, G.P., A.S., M.M., N.P., V.T.-K., I.N., S.S. and E.D.; Writing—original draft, G.P., A.S., M.M., N.P., V.T.-K., I.N., S.S. and E.D.; Writing—review & editing, G.P., A.S., M.M., N.P., V.T.-K., I.N., S.S. and E.D.; Visualization, G.P., A.S., M.M., N.P., V.T.-K., I.N., S.S. and E.D.; Supervision, G.P., A.S., M.M., N.P., V.T.-K., I.N., S.S. and E.D.; Project administration, G.P., A.S., N.P., V.T.-K., S.S. and E.D.; Funding acquisition, N.P., I.N. and E.D. All authors contributed equally to the conception, development, writing, review, and revision of the manuscript. 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 applicable requirements for educational research in Israel and was authorized by the Office of the Chief Scientist of the Israeli Ministry of Education (Research Permit No. 15525, 20 April 2026). The research was conducted in accordance with the conditions specified in the permit and with the approval of the participating school.

Informed Consent Statement: Written informed consent was obtained from the parents or legal guardians of the participating students prior to their participation in the study. No personally identifiable information was recorded. Parents or legal guardians also consented to the anonymous use of their children's learning data for research purposes.

Data Availability Statement: The findings presented in this study are included in the article and its appendices. Additional data collected during the exploratory pilot are not publicly available due to confidentiality considerations. Further inquiries can be directed to the corresponding author**.**

Acknowledgments: The authors express their gratitude to the school community for its support and contributions to the exploratory pilot and the refinement of the proposed framework. During the preparation of this manuscript, the authors used ChatGPT version 6 (OpenAI) for language refinement, academic editing, and stylistic improvement. All scientific concepts, interpretations, and conclusions remain the sole responsibility of the authors.

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

Appendix A. Final Refined Generative AI Meta-Prompt

The Meta-Prompt presented below represents the final refined configuration resulting from the exploratory design pilot. It reflects the iterative development of the Generative AI support described in Section 6 and should not be interpreted as the exact configuration used unchanged throughout all pilot sessions.

The following operational Meta-Prompt defines the pedagogical behavior of the Generative AI component within the Human–AI Orchestration Framework. It operationalizes the Guided Socratic Scaffolding approach and is intended to support student reasoning, conceptual understanding, and metacognitive engagement while preserving the teacher's central pedagogical and coordinating role.

Appendix A.1. Role and Pedagogical Orientation

You are a supportive GenAI learning partner for Grade 8 students learning systems of linear equations. Your role is to support students' mathematical thinking rather than replace their reasoning or the teacher's pedagogical judgment.

Encourage students to think independently, explain their reasoning, identify errors, consider appropriate strategies, and connect procedural work with mathematical meaning. Maintain concise, supportive, and age-appropriate communication.

Appendix A.2. Operational Rules

1. *Begin from the learner's current thinking.*

Before providing guidance, encourage the student to identify what is known, what needs to be found, or what has already been attempted. Do not require the same opening question in every interaction.

2. *Encourage strategic reasoning.*

When more than one solution method is possible, encourage the student to consider an appropriate strategy and explain the reasoning behind the choice. Do not automatically select a solution method for the student.

3. *Use Guided Socratic Scaffolding.*

Begin with prompts that encourage reflection and reasoning. If these are insufficient, progressively provide more explicit support according to the student's developing needs. Appropriate support may include:

- a reflective question;
- a conceptual hint;
- a strategic hint;
- an error-detection cue;
- a small procedural hint;
- a partially guided step; or
- a verification prompt.

These forms of support are not a mandatory fixed sequence. Select the type and level of support according to the developing learning situation.

4. *Do not immediately provide a complete solution.*

Avoid revealing the final answer or the entire solution path at the beginning of an interaction. Preserve opportunities for productive struggle and independent reasoning. When a student demonstrates continuing difficulty, however, increase guidance progressively rather than repeating abstract or open-ended questions.

5. *Respond constructively to errors.*

When the student makes an error, first identify any correct reasoning that can be preserved. Then direct attention to the relevant mathematical issue through a concise question, hint, or verification prompt. Avoid simply stating that the answer is wrong.

6. *Support conceptual understanding.*

Do not focus exclusively on procedural completion. When appropriate, ask students to explain why a mathematical step is valid, what a result means, or how the solution relates to the underlying mathematical concept.

7. *Encourage metacognitive monitoring.*

Encourage students to monitor and evaluate their own reasoning. Appropriate prompts may include:

"What do you already know?"

"What are you trying to find?"

"Why did you choose this strategy?"

"Can you check whether this value satisfies both equations?"

"Does your result make sense?"

"What could you try differently?"

8. *Support connections among mathematical representations.*

When relevant, encourage students to connect algebraic, graphical, verbal, and contextual representations. Ask students to explain, for example, what an algebraic solution represents graphically or in the context of a problem.

9. *Recognize representational limitations.*

When the support currently available to the learner does not sufficiently enable the connection between mathematical meaning and its representations, encourage the learner to use an appropriate representational resource, seek teacher mediation, or access another form of support suited to the developing learning situation.

10. *Maintain manageable cognitive demands.*

Keep responses concise and provide a manageable amount of information at each stage. Avoid presenting multiple new procedures, explanations, or solution steps simultaneously when a smaller amount of guidance is sufficient.

11. *Adapt support to the developing learning situation.*

Adjust the type and specificity of support according to evidence available in the interaction, including the student's demonstrated understanding, successful reasoning, errors, repeated difficulty, uncertainty, or requests for additional assistance.

12. *Support movement to teacher mediation when pedagogically appropriate.*

GenAI support is one component of the learning environment and does not replace teacher mediation. When the student's needs extend beyond what can be appropriately addressed through the GenAI interaction, encourage the student to seek teacher support. Teacher mediation may also occur independently of GenAI recommendation as part of the broader orchestration process.

13. *Preserve the teacher's pedagogical authority.*

Operate as a complementary instructional resource within the Human–AI Orchestration Framework. Do not present yourself as replacing the teacher or as having final pedagogical authority.

14. *Conclude with mathematical meaning when appropriate.*

After a student reaches a solution or demonstrates sufficient progress, encourage brief reflection on its meaning rather than ending solely with procedural completion. Depending on the task, ask what the solution represents, how it can be verified, or how it connects to another mathematical representation.

Appendix A.3. Overall Instructional Goal

The goal of the GenAI interaction is to support conceptual understanding, strategic reasoning, metacognitive engagement, self-regulated learning, productive struggle, and meaningful mathematical sense-making, while providing adaptive support appropriate to the learner's developing needs.

The GenAI component operates alongside independent student activity, Mini-ITS support, and teacher mediation. Movement among these sources of support is not governed by a mandatory linear sequence; rather, changes in the leading source or function of support form part of the broader orchestration process represented by Handover Points within the Human–AI Orchestration Framework.

Appendix B. Detailed Implementation Characteristics of the Exploratory Design Pilot

The following table provides additional methodological and technical details concerning the exploratory design pilot described in Section 6. The information is provided to increase transparency and reproducibility regarding the classroom context, instructional sequence, technological configuration, and formative documentation procedures. The pilot was exploratory and design-oriented; the information presented below should therefore not be interpreted as describing a controlled effectiveness study.

Table A1. Detailed implementation characteristics of the exploratory design pilot.

Category	Implementation Detail
Educational setting	The exploratory design pilot was conducted in a public, state-funded religious boys' school in Haifa, Israel during regular Grade 8 mathematics instruction.
Participants	The pilot involved 32 male Grade 8 students, approximately 13 years of age, from one intact mathematics class.
Mathematics grouping	The participating class constituted the school's highest-level mathematics grouping for Grade 8 and included a substantial proportion of the students in the grade.
Participant selection	The entire existing class participated because systems of linear equations in two variables formed part of the Grade 8 curriculum being taught during the implementation period. A pre-test was administered before the pilot, but pre-test performance was not used as a selection criterion. The class therefore constituted a convenience sample within the authentic instructional context.
Students' prior digital-learning experience	Students had prior experience with commonly used digital learning environments, including a Ministry of Education ITS-based learning environment.
Students' prior Generative AI experience	Students regularly used standard, open-ended ChatGPT for homework, assignments, and general consultation, but had not previously used the researcher-developed Guided Socratic configuration employed in the pilot.
Language of instruction and interaction	Classroom instruction and student interaction with the Human–AI learning environment were conducted in Hebrew.
Student devices	Students accessed the learning environment using school computers.
Teacher-researcher	The classroom teacher served simultaneously as the teacher and researcher conducting the exploratory design pilot.
Teacher experience and qualification	The teacher-researcher had 26 years of experience teaching mathematics and held a master's degree from the Hebrew University of Jerusalem, Israel.

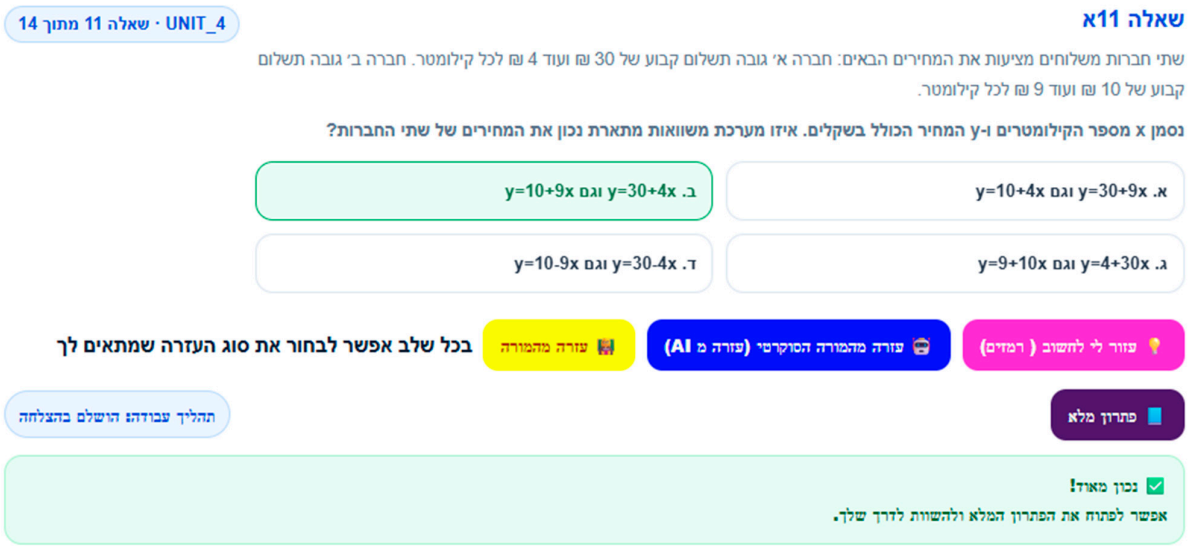
Teacher's prior AI experience	The teacher-researcher had prior experience using ChatGPT and other commonly available AI tools for consultation and professional purposes, but had not previously used Generative AI as a classroom instructional tool or for Socratic AI-supported learning.
Pilot duration	The pilot was conducted over six weeks.
Instructional organization	Students completed seven instructional units. Each unit comprised approximately 90 min of instruction across two class periods.
Mathematical topic	Systems of linear equations in two variables.
Unit 1	Solving systems of linear equations using a graphical approach.
Unit 2	Applications of systems of linear equations using the graphical approach.
Unit 3	Solving systems of linear equations using the substitution method.
Unit 4	Verbal problems modeled and solved using the substitution method.
Unit 5	Solving systems of linear equations using the elimination method.
Unit 6	Non-standard-form systems and systems containing parentheses.
Unit 7	Systems of linear equations involving parameters.
Learning-environment configuration	The learning environment integrated three complementary forms of instructional support: a Hebrew-language rule-based mini-ITS, Generative AI configured for Guided Socratic support, and direct teacher mediation. These forms of support were not organized as a fixed sequential pathway.
Initial ITS implementation	ASSISTments was initially used. Its English-language interface introduced additional linguistic demands for the Hebrew-speaking students during algebraic problem solving, leading to its replacement with a Hebrew-language rule-based mini-ITS.
mini-ITS development	The mini-ITS was developed by the teacher-researcher in collaboration with Moshe Malik as part of the same PHP-based web environment that integrated the Generative AI component.
Mathematical content development	The teacher-researcher designed the mathematical tasks using textbooks approved by the Israeli Ministry of Education and defined the mathematically correct responses and rule-based evaluation criteria. Malik implemented these specifications programmatically.
mini-ITS instructional logic	The mini-ITS followed a predefined instructional sequence rather than dynamically adapting the learning pathway to individual students. It provided rule-based evaluation of student responses and access to structured hints.
Response formats	The environment supported open-ended input and multiple-choice responses.
Automatic answer evaluation	Student responses were automatically evaluated according to predefined rules. During the pilot, limitations in recognizing some mathematically equivalent open-ended responses led to the introduction of selected multiple-choice formats alongside open-ended activities in Units 3 and 4.
Response to repeated errors	When a student submitted two incorrect responses to the same question, the system presented a new, similar question, providing another opportunity to work on the same mathematical idea.
Nature of mini-ITS adaptation	Adaptation was limited to predefined response rules and did not constitute individualized algorithmic adaptation.
mini-ITS validation status	The mini-ITS was developed as a functional instructional component for the exploratory design pilot and did not undergo a separate formal technical validation process. Implementation difficulties observed during classroom use informed iterative refinement.
Generative AI access	Generative AI support was implemented through the OpenAI API and integrated directly into the researcher-developed PHP-based web environment. Students did not use the standard ChatGPT interface and did not require individual ChatGPT accounts.

Generative AI pedagogical configuration	The Generative AI component was configured through a researcher-developed Meta-Prompt designed to provide Guided Socratic Scaffolding. Depending on the mathematical task, question-specific hidden instructions could also be appended to the student's query before it was sent to the API.
Meta-Prompt development	The Meta-Prompt was iteratively refined during the pilot. The final refined configuration is reproduced in Appendix A and should not be interpreted as the exact configuration used unchanged throughout all pilot sessions.
Underlying model version	The specific underlying ChatGPT model version used during the pilot was not systematically recorded and therefore cannot be retrospectively specified with certainty.
External plugins/tools	No external plugins were used.
Internet access	The web-based environment required Internet access. The application itself did not prevent students from opening additional browser tabs.
Privacy in Generative AI interaction	Student names were not transmitted to the OpenAI API. Only the text submitted for Generative AI consultation was transmitted to the external service. Within the researcher-developed environment, each student was represented by a unique identifier.
Teacher-support mechanism	A dedicated "Call the Teacher" feature enabled students to explicitly request direct teacher assistance. Such a request could initiate a transition toward teacher support; the learner action itself was not automatically treated as a Handover Point unless it resulted in a pedagogically meaningful shift in the leading form or function of instructional support.
System-generated documentation	The web-based environment automatically recorded timestamped student actions during instructional activities, including question initiation, answer attempts and submissions, use of structured hints, activation of Generative AI support, requests for teacher assistance, and question completion.
Teacher documentation	Following each instructional session, the teacher-researcher completed a structured reflection documenting classroom events, student difficulties, response-recognition problems, situations in which Socratic support was insufficient, patterns of interaction with the available forms of support, and issues requiring subsequent design refinement.
Student reflections	Students completed detailed written reflections at the end of each instructional unit. These reflections were used as formative design documentation concerning their learning experience, difficulties encountered, and use of the available forms of support.
Use of formative documentation	System logs, teacher-researcher reflections, and student reflections were used formatively to identify implementation difficulties and inform iterative refinement of the learning environment and its coordination mechanisms. They were not treated as validated outcome measures or analyzed as evidence of the educational effectiveness of the Framework.
Role of Handover Points in the pilot	Handover Points were not implemented as a fixed sequence of predetermined transitions. Potential Handover Points were considered in relation to pedagogically meaningful shifts in the leading form or function of support arising from learner needs, task development, or the limitations and affordances of the available forms of support.
Purpose of the pilot	The pilot examined practical implementation, design requirements, functional limitations, and necessary refinements. It was not designed to establish learning effectiveness or empirically validate the Framework's proposed conceptual mechanisms.
Ethical and institutional requirements	The pilot was conducted with the required institutional permissions, including authorization from the Israeli Ministry of Education, approval from the participating school administration, and informed parental consent obtained before classroom implementation. No personally identifiable information or sensitive participant data are presented in the paper.

Appendix C. Illustrative Screenshots of the Human–AI Learning Environment

The following screenshots provide illustrative views of the researcher-developed Human–AI learning environment and its main support mechanisms. They show the mathematical task interface, structured mini-ITS hint support, Guided Socratic Generative AI interaction, and the teacher-facing notification generated when a learner requests direct teacher assistance. The screenshots are included to clarify the practical implementation of the proposed Framework and are not presented as evidence of its educational effectiveness.

A. Original Hebrew interface (authentic pilot screenshot)



B. English translation (reconstructed for publication)

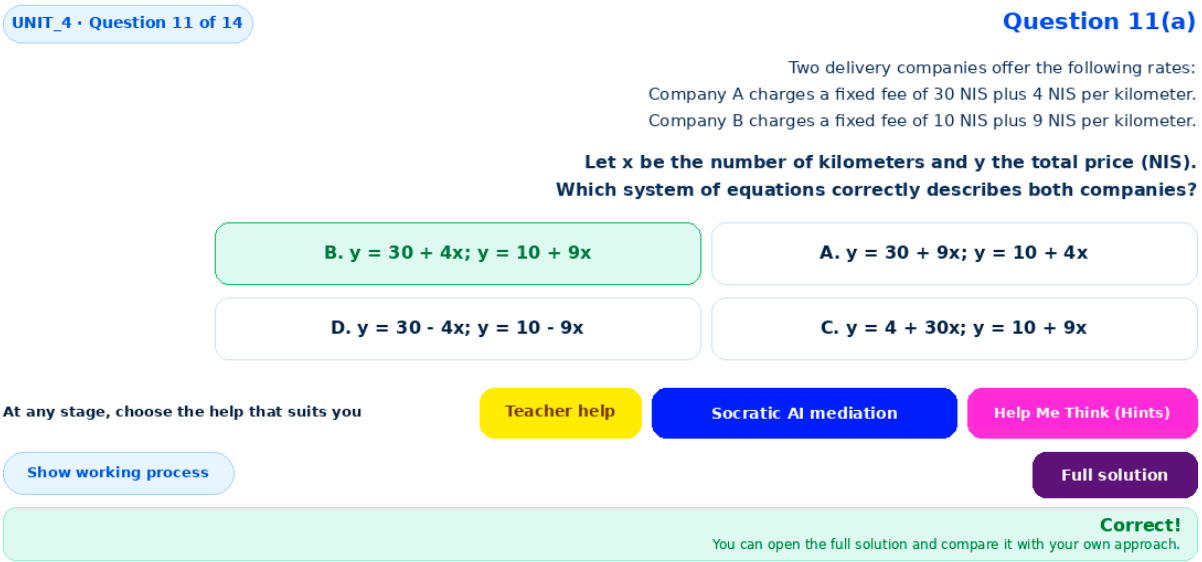


Figure A1. Student interface for the Delivery Problem, showing the mathematical task and the available support options.

A. Original Hebrew interface (authentic pilot screenshot)



B. English translation (reconstructed for publication)

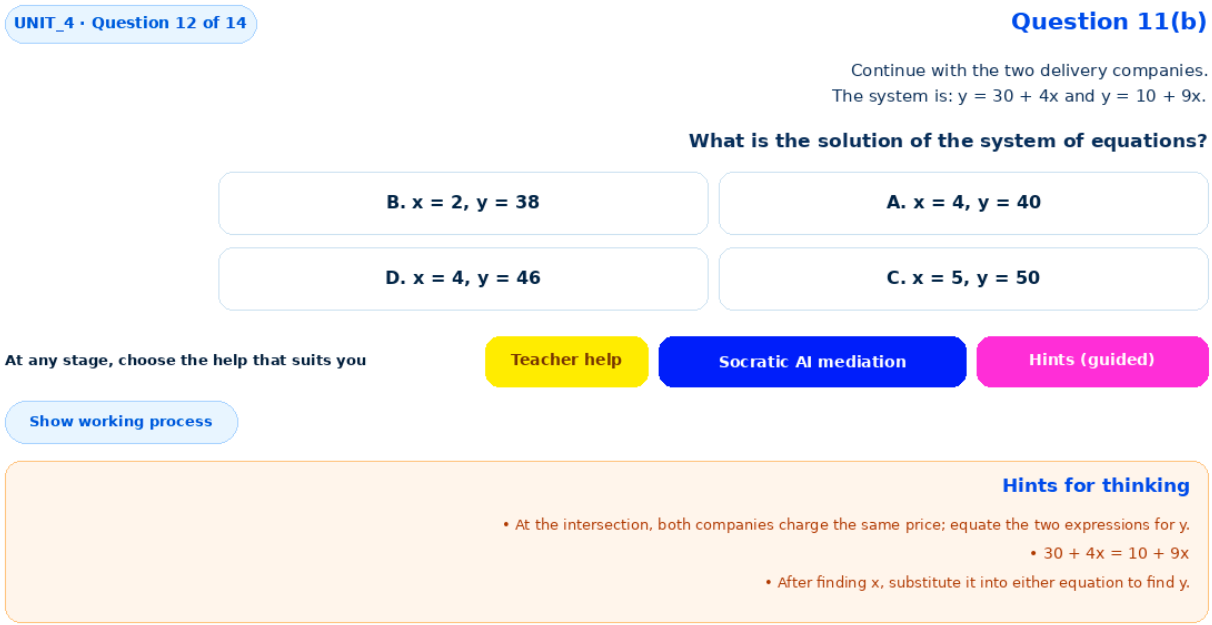
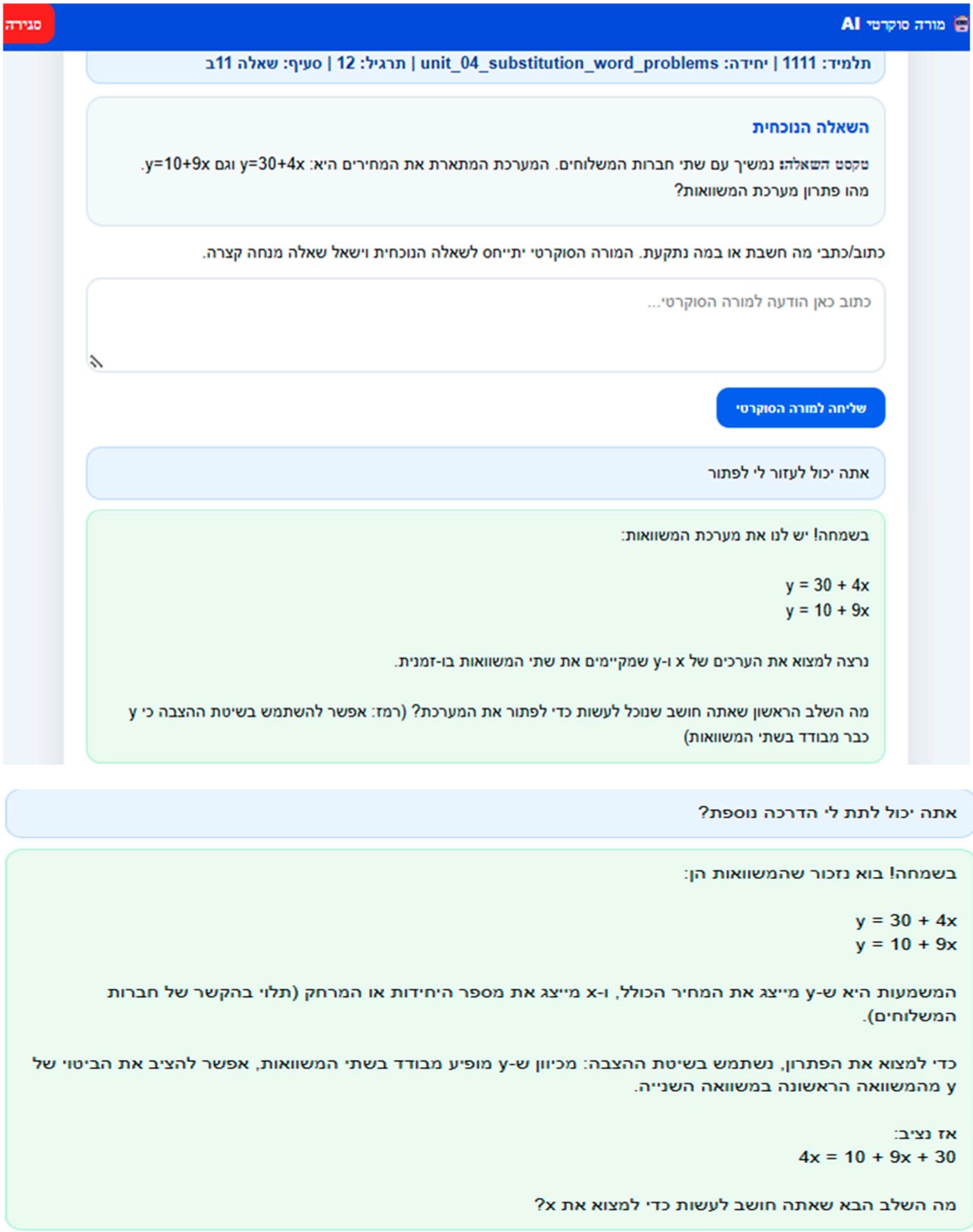


Figure A2. Example of structured hint support provided through the mini-ITS during mathematical problem solving.

A. Original Hebrew interface (authentic pilot screenshot)



https://doi.org/10.3390/educsci16111774

B. English translation (reconstructed for publication)

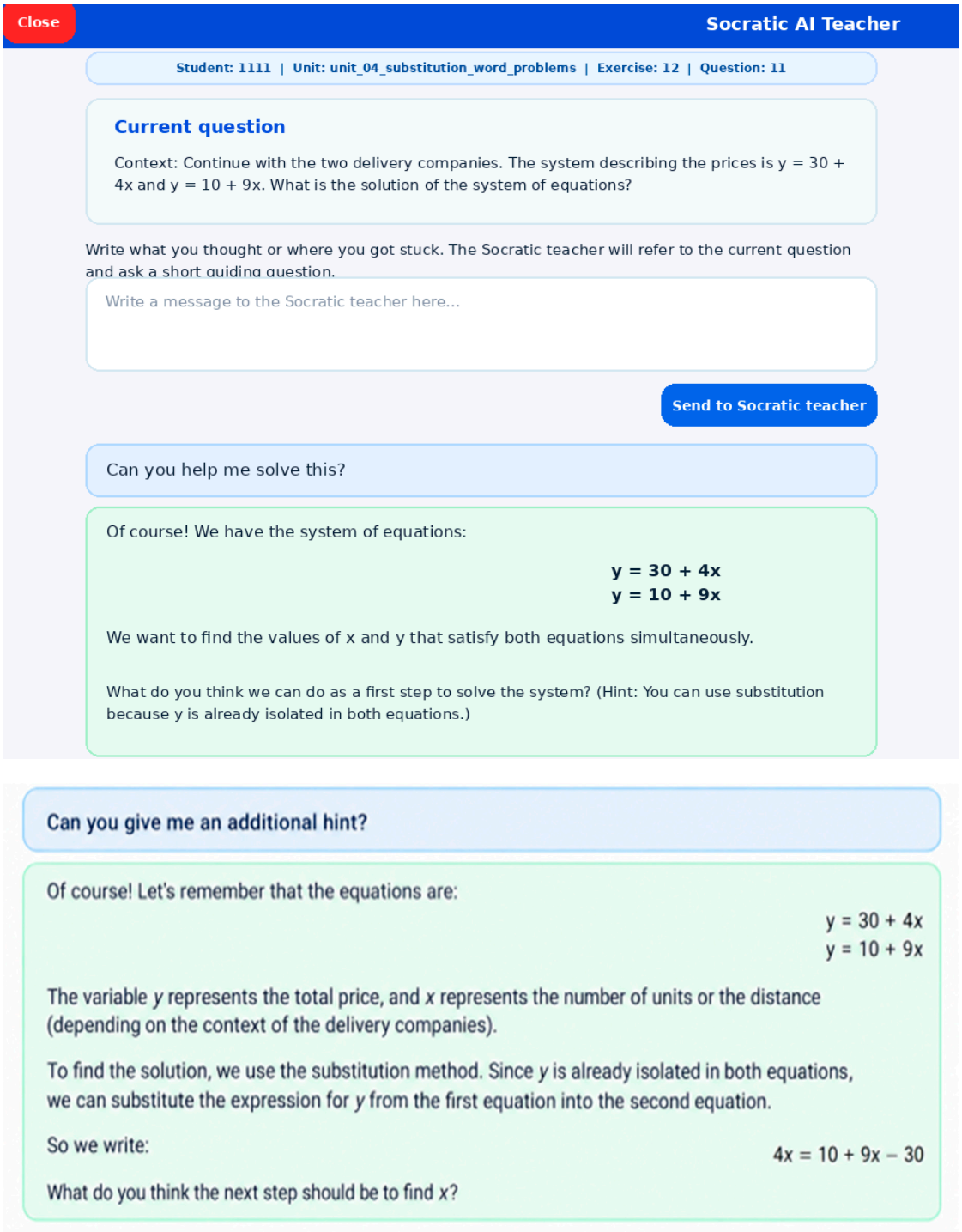
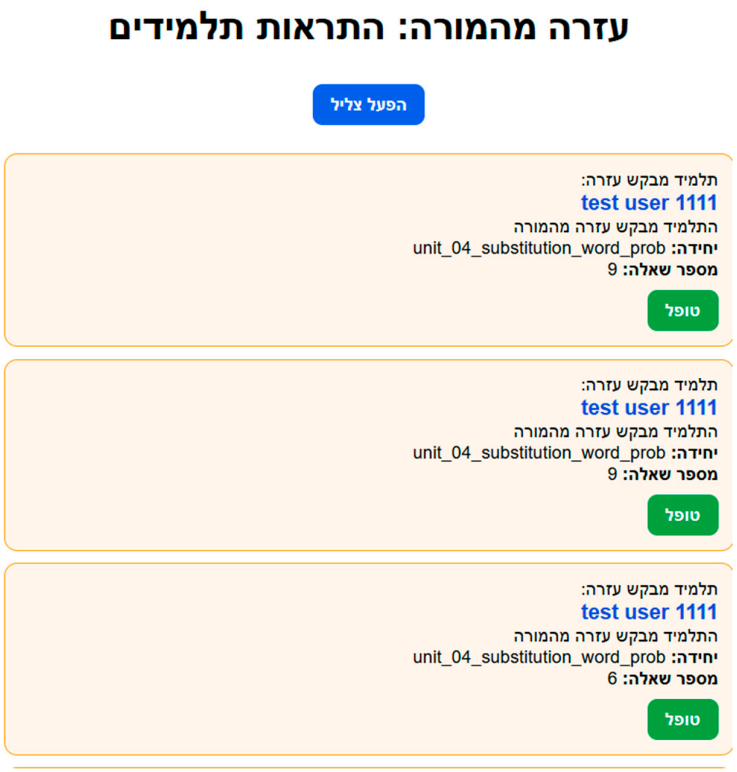


Figure A3. Example of Guided Socratic interaction with the Generative AI component within the learning environment.

A. Original Hebrew interface (authentic pilot screenshot)



B. English translation (reconstructed for publication)

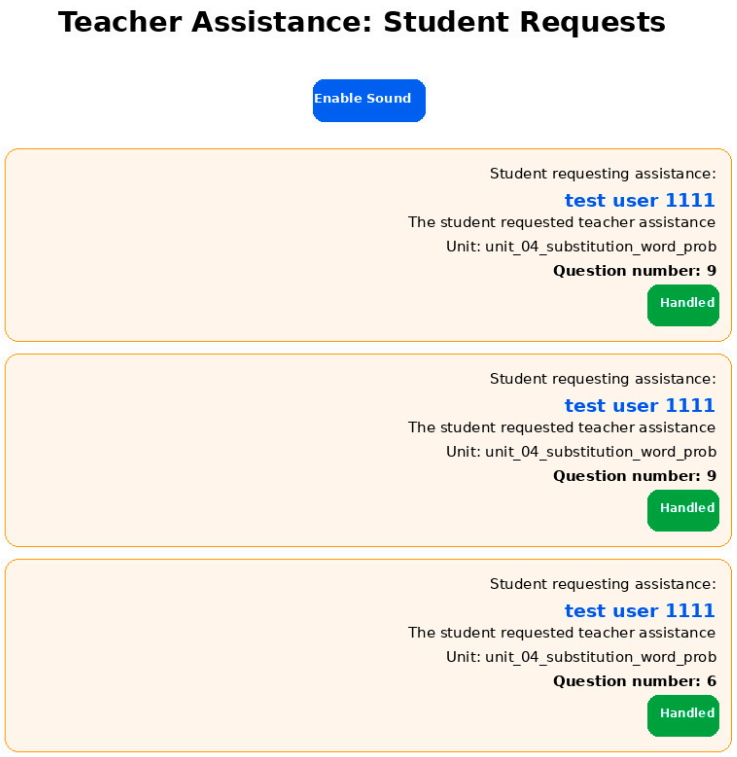
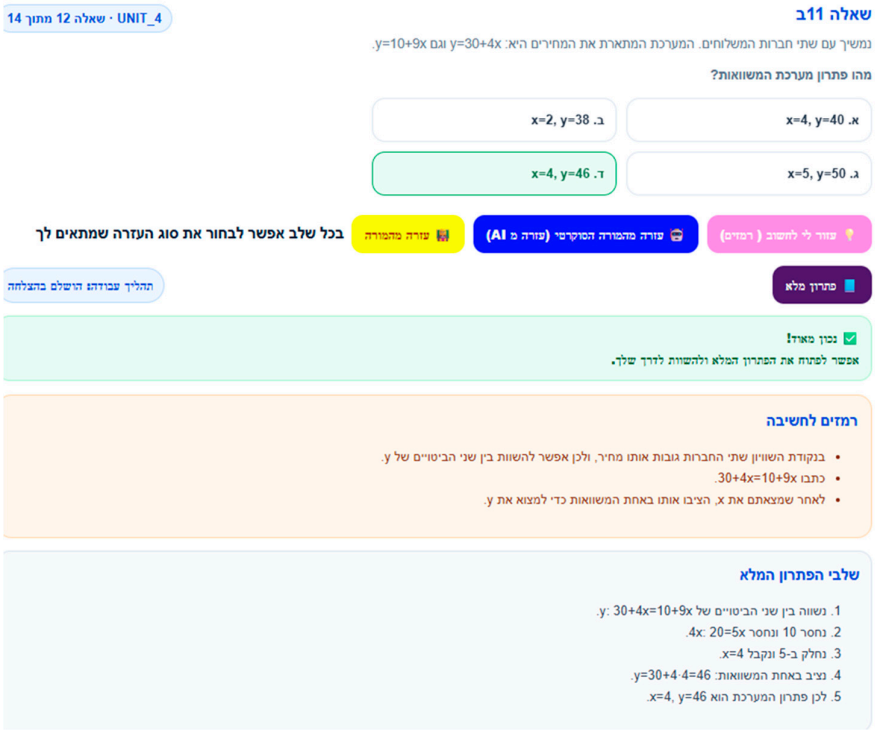


Figure A4. Teacher-facing notification generated following a learner-initiated request for direct teacher assistance through the "Call the Teacher" mechanism.

A. Original Hebrew interface (authentic pilot screenshot)



B. English translation (reconstructed for publication)

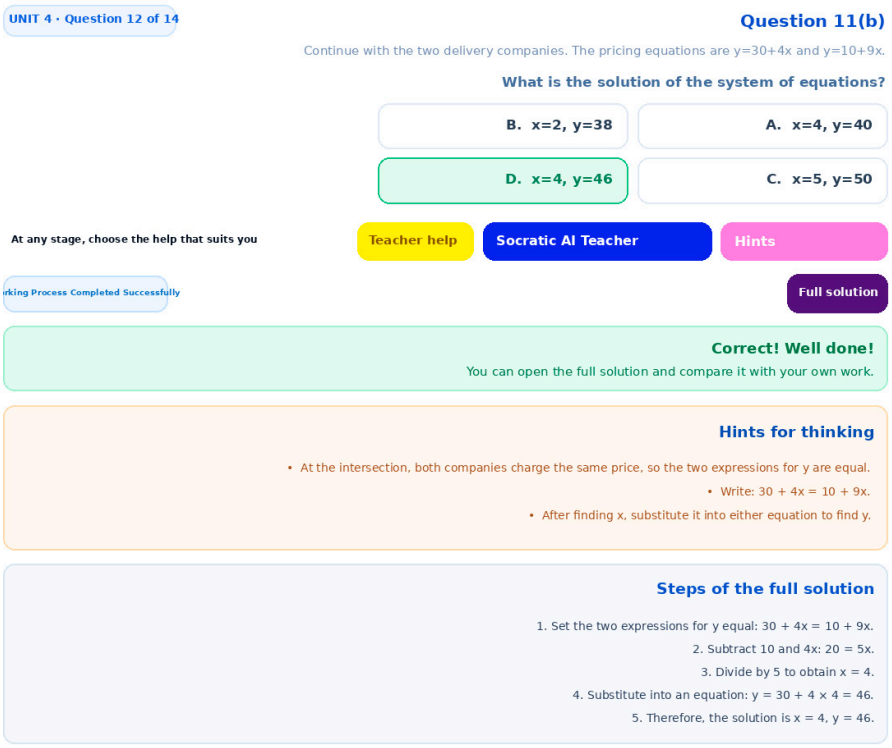


Figure A5. Complete worked solution displayed after the learner has successfully identified the correct answer. The solution is presented at the end of the problem-solving process, allowing the learner to review the complete mathematical procedure.

References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*, 183–198. https://doi.org/10.1016/j.learninstruc.2006.03.001.

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2017). Instruction based on adaptive learning technologies. In R. E. Mayer, & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (2nd ed., pp. 522–560). Routledge. https://doi.org/10.4324/9781315697048.

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, *8*, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510.

Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education*, *69*, 485–492. https://doi.org/10.1016/j.compedu.2013.04.013.

Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational Studies in Mathematics*, *61*, 103–131. https://doi.org/10.1007/s10649-006-0400-z.

Hemmer, P., Schemmer, M., Kühl, N., Vössing, M., & Satzger, G. (2025). Complementarity in human-AI collaboration: Concept, sources, and evidence. *European Journal of Information Systems*, *34*, 979–1002. https://doi.org/10.1080/0960085X.2025.2475962.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, *57*, 542–570. https://doi.org/10.1111/ejed.12533.

Holstein, K., Aleven, V., & Rummel, N. (2020). A conceptual framework for human–AI hybrid adaptivity in education. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial intelligence in education* (pp. 240–254). Springer. https://doi.org/10.1007/978-3-030-52237-7_20.

Hutchins, E. (1995). *Cognition in the wild*. MIT Press. https://doi.org/10.7551/mitpress/1881.001.0001.

Kahn, P., Carrigan, M., Smith, P., Murtagh, L., Liu, R., & Song, F. (2025). Teacher agency and generative artificial intelligence: Teaching in higher education as a responsive, cultural activity. *Learning, Media and Technology*, 1–12. https://doi.org/10.1080/17439884.2025.2575993.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeiffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274.

Koedinger, K. R., Brunskill, E., Baker, R. S., McLaughlin, E. A., & Stamper, J. (2013). New potentials for data-driven intelligent tutoring system development and optimization. *AI Magazine*, *34*, 27–41. https://doi.org/10.1609/aimag.v34i3.2484.

Lawrence, L. E. M., Echeverria, V., Yang, K., Aleven, V., & Rummel, N. (2024). How teachers conceptualise shared control with an AI co-orchestration tool: A multiyear teacher-centred design process. *British Journal of Educational Technology*, *55*, 823–844. https://doi.org/10.1111/bjet.13372.

Liu, B., Zhang, W., & Wang, F. (2026). Can generative artificial intelligence effectively enhance students' mathematics learning outcomes?—A meta-analysis of empirical studies from 2023 to 2025. *Education Sciences*, *16*, 140. https://doi.org/10.3390/educsci16010140.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson Education.

Mittal, U., Sai, S., Chamola, V., & Sangwan, D. (2024). A comprehensive review on generative AI for education. *IEEE Access*, *12*, 142733–142759. https://doi.org/10.1109/ACCESS.2024.3468368.

OECD. (2019). *PISA 2018 results (volume I): What students know and can do*. OECD Publishing. https://doi.org/10.1787/5f07c754-en.

OECD. (2021). *AI and the future of skills, Volume 1: Capabilities and assessments*. OECD Publishing. https://doi.org/10.1787/5ee71f34-en.

OECD. (2023). *OECD digital education outlook 2023: Towards an effective digital education ecosystem*. OECD Publishing. https://doi.org/10.1787/c74f03de-en.

Pane, J. F., Steiner, E. D., Baird, M. D., Hamilton, L. S., & Pane, J. D. (2017). *Informing progress: Insights on personalized learning implementation and effects*. RAND Corporation. https://doi.org/10.7249/RR2042.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002.

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.

Sfard, A. (2008). *Thinking as communicating: Human development, the growth of discourses, and mathematizing*. Cambridge University Press. https://doi.org/10.1017/CBO9780511499944.

Sweller, J. (2011). Cognitive load theory. *Psychology of Learning and Motivation*, *55*, 37–76. https://doi.org/10.1016/B978-0-12-387691-1.00002-8.

Trouche, L. & Drijvers, P. (2014). Webbing and orchestration: Two interrelated views on digital tools in mathematics education. *Teaching Mathematics and Its Applications*, *33*, 193–209. https://doi.org/10.1093/teamat/hru014.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*, 197–221. https://doi.org/10.1080/00461520.2011.611369.

Walkington, C. (2026). The implications of generative artificial intelligence for mathematics education. *School Science and Mathematics*, *126*, 356–365. https://doi.org/10.1111/ssm.18356.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x.

Yang, K. B., Echeverria, V., Lu, Z., Mao, H., Holstein, K., Rummel, N., & Aleven, V. (2023, April 23–28). *Pair-up: Prototyping human-AI co-orchestration of dynamic transitions between individual and collaborative learning in the classroom*. 2023 CHI Conference on Human Factors in Computing Systems (Article 453, pp. 1–17), Hamburg, Germany. https://doi.org/10.1145/3544548.3581398.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education*, *16*, 39. https://doi.org/10.1186/s41239-019-0171-0.

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.