education sciences

Article

Impact of a Low-Stakes Assessments Model with Retake in General Chemistry: Connecting to Student Attitudes and Self-Concept

Vijay S. Vyas, Llanie Nobile, James R. Gardinier and Scott A. Reid *

Department of Chemistry, Marquette University, Milwaukee, WI 53233, USA; vijay.vyas@marquette.edu (V.S.V.); llanie.nobile@marquette.edu (L.N.); james.gardinier@marquette.edu (J.R.G.)
* Correspondence: scott.reid@marquette.edu

Abstract: Across a variety of fields, the use of low-stakes assessments has led to reductions in achievement gaps and improved student success. Here, we probe the use of a low-stakes assessment model with a retake option for failed quizzes in a two-semester general chemistry sequence. We find that the quiz-retake rate in general chemistry II was significantly higher for students who had completed a retake in a general chemistry I section, and the percentage of students who failed at least one quiz in general chemistry I but passed all quizzes in general chemistry II was significantly higher for students who had retaken at least one quiz in general chemistry I. However, across both semesters only 40% of students who failed a quiz and were offered a retake completed one. To examine this trend, we probed a connection to student attitudes and self-concept. As instruments, we used version 2 of the Attitudes towards Chemistry Subject Inventory (ASCIv2) and the Chemistry Subject Concept Inventory (CSCI), which were administered across all sections of our general chemistry I course in the fall 2021 semester, and the results subjected to confirmatory factor analysis. Two sections employed low-stakes assessments (quizzes), with one section offering a retake option, while the remaining two used a traditional assessment pattern of five exams. The instruments were applied again for the quiz-retake section of general chemistry II, affording a longitudinal comparison of students common to both sections. In a pairwise comparison, we find significant increases in factors corresponding to Intellectual Accessibility and Chemistry Self-Concept for students in the quiz-retake sections across semesters, with the former more pronounced for men and the latter for women. We take these results to provide additional data supporting the benefit of low-stakes assessments with a retake option, that may be particularly impactful for women in chemistry.

Keywords: low-stakes assessments; retrieval processes; introductory chemistry

check for updates

Citation: Vyas, V.S.; Nobile, L.; Gardinier, J.R.; Reid, S.A. Impact of a Low-Stakes Assessments Model with Retake in General Chemistry: Connecting to Student Attitudes and Self-Concept. *Educ. Sci.* **2023**, *13*, 1235. https://doi.org/10.3390/educsci13121235

Academic Editor: Han Reichgelt

Received: 12 September 2023
Revised: 5 December 2023
Accepted: 12 December 2023
Published: 13 December 2023

1. Introduction

The implementation of frequent, low-stakes assessments has been increasingly recognized in the educational literature as a successful strategy to improve student success. Such an assessment strategy can reduce or eliminate gender inequities that are commonly found when employing high-stakes assessments [1–5]. Studies across a range of disciplines including psychology [6,7], biology [8,9], medicine [10], and chemistry [11,12] have shown the benefits of frequent assessment, a strategy that is well grounded in educational theories centered in mechanisms of retrieval [13–17] and growth mindset [18]. In distinguishing between fixed and growth mindsets, Dweck emphasizes that a fixed mindset is not permanent but may evolve into a growth mindset with proper intervention. Thus, longitudinal studies can be particularly important in probing the degree to which this is true.

We have recently examined the impact of low-stakes assessments on student success in general chemistry, focusing particularly on the combined impact of such assessments and instructor interventions on rates of Ds, Fs, and withdrawals (i.e., DFW rates). In particular, we focused on the impact of this assessment strategy on DFW outcomes in large

(~180 students) sections of our general chemistry I/II courses. There, 10 weekly low-stakes assessments (quizzes) accounted for 40% of the overall grade, and students failing any quiz were provided with a retake opportunity after meeting with the instructor. In this meeting, the instructor reviewed the quiz and associated content, and discussed strategies for succeeding in the course. Across both semesters, the quiz-retake rate was around 40%, and the mean improvement in quiz score upon retake was around 30%, which was typically sufficient to push the quiz score into a passing range, as only the highest of the two scores was counted for the overall grade. All quizzes were administered through our university's course management system, incorporating webcam monitoring and browser lockdown, and were constructed via random selection from pools of similar questions, with prominent question types including multiple choice and multi-select.

In this article, we seek to examine the quiz-retake model in more detail, and particularly to probe the connection between quiz retakes and attributes of the affective domain, specifically student attitudes and self-concept. As highlighted in a recent review, affective characteristics such as self-efficacy and self-concept are critical for engagement and performance in STEM (Science, Technology, Engineering, and Mathematic) fields, with women typically having lower self-concept in STEM [19]. Significant literature has focused on the measurement of student attitudes across different disciplines [20–29], with a recent review summarizing studies of affective domain research in the chemistry education literature [30]. These studies emphasized that "attitude" is inherently a multidimensional construct [31], encompassing as it does such varied aspects as beliefs [32], interests and motivations [33,34], values [35], self-efficacy [36], and self-concept and its associated measures [37–39].

Early psychometric measures of attitude and self-concept in chemistry include that of Shrigley, who in 1984 reported the design and evaluation of a Likert-scale instrument [40]. In 2005, Bauer reported a new survey instrument, denoted the Chemistry Self-Concept Inventory or CSCI [41]. Factor analysis of responses to the 40 survey questions identified five distinct factors that included Mathematical Self-Concept (SC), Chemistry SC, Academic SC, Academic Enjoyment SC, and Creativity SC. Reliability values were typically above 0.7, and thus considered to be strong. In recent years, the CSCI instrument has been extensively used in chemistry education research [42–45], alongside other instruments [46,47]. In a particularly notable study, Lewis and co-workers used the CSCI and a cluster analysis to demonstrate a significant disparity in retention for students in a low self-concept group [48].

In 2008, Bauer developed a second instrument to measure student attitudes towards chemistry, the Attitudes towards the Subject of Chemistry Inventory, or ASCI [49]. Factor analysis of the 20 items in the original instrument identified three unique factors, labelled *Interest* and *Utility*, *Anxiety*, and *Intellectual Accessibility*. This instrument was later modified by Lewis and co-workers to a shorter instrument (ASCIv2), with eight items factored into two subscales, labelled *Intellectual Accessibility* and *Emotional Satisfaction* [50]. As in the original instrument, the reliability factors were found to be strong. Over the last decade, the ASCI and ASCIv2 instruments have been arguably the most widely used quantitative instruments across affective domain chemistry education research [47,51–57].

The focus of the present article is to probe our central hypothesis by examining the connection between student attitudes and self-concept and the use of a low-stakes assessment model incorporating voluntary retakes of failed quizzes, and the evolution of these patterns over the course of a general chemistry I/II sequence. As illustrated in Figure 1, in the fall 2021 semester we gave a diagnostic exam across all sections of our general chemistry I course (N = 624). The exam consisted of 10 chemistry-specific questions and the ASCIv2 and SCSI instruments. These data were subjected to confirmatory factor analysis, and subsequently used to examine in detail trends in the general chemistry I section that included a quiz-retake option for failed quizzes. The instruments were applied again at the beginning of the general chemistry II course for the quiz-retake section, providing a parallel view of students across different general chemistry I sections, and a longitudinal view of changes in student attitudes and self-concept for the students (n = 101) common to both quiz-retake sections.

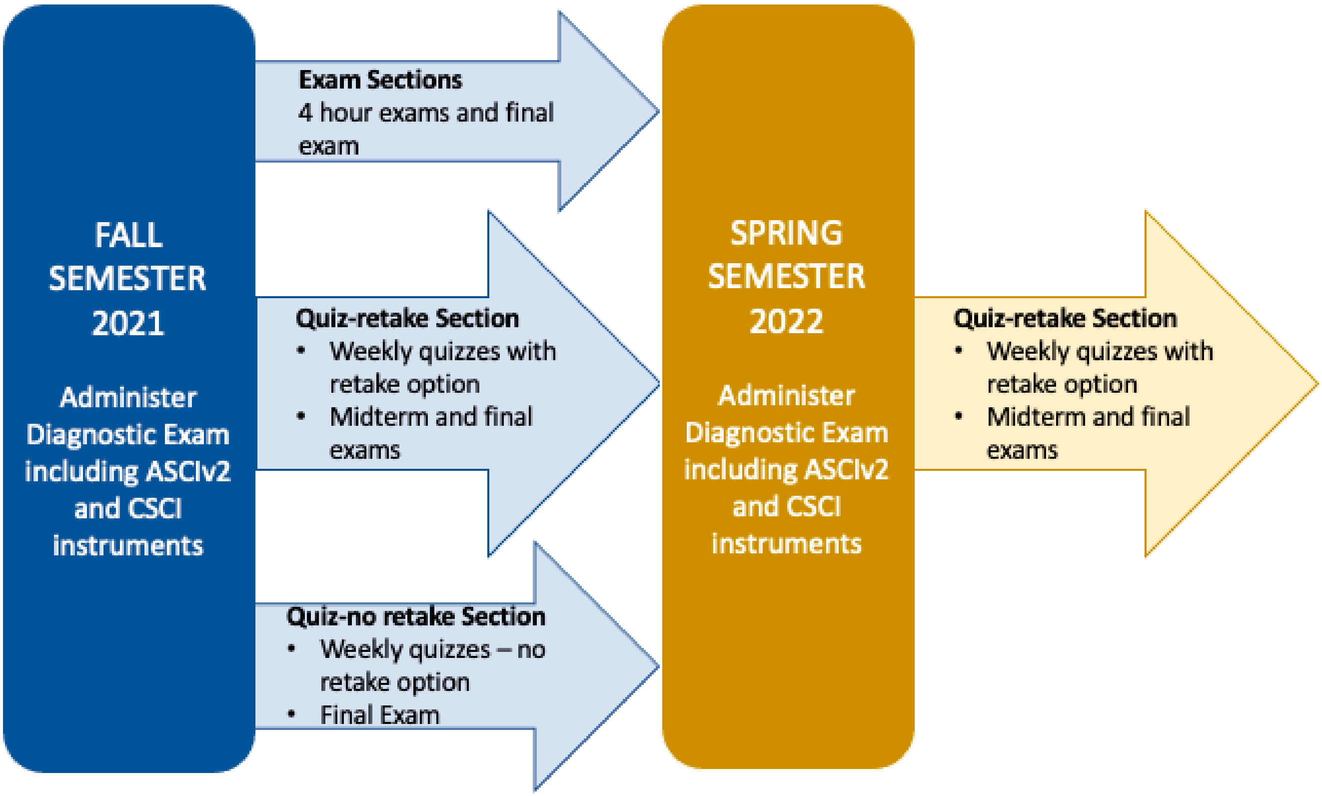

Figure 1. Illustration of the approach to data collection used in this study.

2. Theoretical Frameworks

The role that affective domain characteristics such as attitude and self-concept play in learning is well grounded in a variety of psychological theories [58–60]. As the present study seeks to examine the connection and evolution of student attitudes and self-concept with the retaking of failed quizzes in a low-stakes assessment model, a framework of particular relevance centers on the testing effect and associated retrieval processes, focusing in particular on learning gains via the practice of regular testing, which fosters retrieval mechanism development [13–17]. The testing theory of Bjork and Bjork inherently discriminates *storage strength* (i.e., the degree to which an item is well learned) vs. *retrieval strength* (i.e., the degree to which the item is retrievable) [16], and defines *desirable difficulties* as those exercises or practices that improve long-term retention [61]. Regular testing is certainly one of these practices [15], as is the ability to retake a failed quiz after review. Importantly, this theory predicts that students with weaker initial study skills may benefit more from frequent testing [8,13,16,17,62].

3. Research Questions

The research questions that we aimed to address in this study were: (1) Do student attitudes and self-concept, as measured by quantitative instruments, connect to their decisions regarding offered retakes of failed low-stakes assessments? (2) What is the longitudinal evolution of student attitudes and self-concept, as measured by quantitative instruments, across two semesters of general chemistry in sections using a low-stakes assessment model with retake option for failed quizzes? (3) Given the gender disparities reported for high-stakes assessments [1–5], what differences in the evolution of student attitudes and self-concept (and performance!) are observed by binary gender? To address these questions, we focus on large ($n \sim 180$) sections of general chemistry I/II that employed a quiz-retake model taught on-sequence in fall 2021 and spring 2022.

4. Student Demographics and Section Detail

The students participating in this study were enrolled in a large private Midwestern university in a non-majors general chemistry I/II sequence. A separate sequence for majors is also offered, but as these sections were small (capped at 24 students) and had a different student demographic, they were not included in this study. The demographic breakdown of the student population for the fall general chemistry I sections (N = 624 total, with full demographic information available for N = 596) is shown in Table S1 (the S prefix denotes material in the Supporting Information). We note that the first-generation population (23%) has remained constant over time, while the percentage of White, non-Hispanic students has dropped over time and here was 65%. The fastest-growing demographic in our general chemistry courses is Hispanic/Latino students, who here comprised 18% of the total population. The breakdown by college shows that 76% of students are from two colleges, Engineering and Health Sciences, reflecting the service nature of the courses.

Two of the four fall general chemistry I sections used a traditional assessment pattern centered around five examinations, that together counted for 64% of the grade. The other two sections used a pattern of weekly quizzes, with one offering a retake option. In both quizzing sections, the quizzes counted for 40% of the total grade. Considering the quiz-retake section, which is the focus of this article, students failing a quiz (defined as a score of 50% or below) were invited to meet with the instructor and thereafter were given the option of retaking the quiz, with the higher score counting towards the final quiz grade. As mentioned above, in these meetings the instructor reviewed the quiz and discussed study strategies for the class. While the quizzes counted for 40% of the overall grade, we classify these as "low-stakes assessments", as each carried a fraction of the weight of a traditional high-stakes examination. This is illustrated in Figure S1, which compares the current course assessment pattern with a historical pattern based upon high-stakes examinations.

5. Methodology

Following IRB review, we obtained deidentified student demographic and grade information for the on-sequence (i.e., fall general chemistry I, spring general chemistry II) sections of general chemistry I and II in the 2021-2022 academic year. As noted above, at the beginning of the year a diagnostic quiz consisting of 10 chemistry-focused questions and the ASCIv2 and CSCI instruments was given across all four sections of general chemistry I (total N = 624). In the quizzing sections, the diagnostic was administered as the first quiz in the class at the end of the first week, to aid in introducing students to the quizzing environment. In the other sections, the quiz was administered during the first week and was counted for participation points. A total of 522 students (83.7%) completed the chemistry component of the diagnostic, while a total of 518 students (83.0%) completed both the chemistry component and the ASCIv2, and a total of 509 students (81.6%) completed all components. All data were stored in Excel workbooks and analyzed using Excel and SPSS Statistics v. 28.0 [63].

Considering results from the chemistry-specific diagnostic questions, we classified this as Measure 0 (abbreviated M0). In addition to this diagnostic, we also used for the quiz-retake section the initial knowledge check (classified as M00) in the ALEKS (Adaptive Learning in Knowledge Spaces) adaptive homework system that was used in the course [64]. Moving to results from the ASCIv2 instrument, after rescoring we performed confirmatory factor analysis (CFA) on the eight ASCIv2 items using AMOS 29 [65]. The CFA was performed first on data from the complete set of students, and again only for students in the quiz-retake section, which is the focus of this article. The fit results, shown in Table S2, validate the two components identified by Lewis (termed here M1 and M2, corresponding to Intellectual Accessibility and Emotional Satisfaction) [50]. A graphical representation of the path diagram that includes standardized estimates from the fit to the data for the complete set of students is shown in Figure S2.

Turning to the CSCI instrument, after rescoring we performed confirmatory factor analysis on the 40 items in the instrument using AMOS 29. The CFA was performed first on data from the complete set of students, and again only for students in the quiz-retake section, which is the focus of this article. The results, shown in Table S3, validate the five factors determined in the original analysis [41], although the fit quality was significantly poorer than found for the ASCIv2 instrument. These five factors are: Mathematical Self-Concept (Measure 3 or M3), Chemistry Self-Concept (M4), Academic Self-Concept (M5), Academic Enjoyment Self-Concept (M6), and Creativity Self-Concept (M7). A graphical representation of the path diagram that includes standardized estimates from the fit to the data for the complete set of students is shown in Figure S3.

As another view of the reliability of the CSCI instrument, we used Cronbach's α to examine the internal consistency of the subscales [66], a measure also used by Bauer in his original article [41]. The results for (a) all students in the fall sections and (b) students only in the quiz-retake section are shown in Table 1, where they are compared with the values derived by Bauer. Overall, the values are comparable, with the exception of the Academic Self-Concept subscale, and thus we will not seek here to draw significant conclusions from this measure.

Table 1. Reliability comparison of the CSCI subscales.

Subscale	Bauer [41]	This Study, All Sections	This Study, Quiz-retake Section
Mathematical Self-Concept	0.90	0.91	0.92
Chemistry Self-Concept	0.91	0.88	0.88
Academic Self-Concept	0.77	0.67	0.67
Academic Enjoyment Self-Concept	0.77	0.81	0.81
Creativity Self-Concept	0.62	0.65	0.71

Summarizing, in this work we thus considered a total of nine measures, abbreviated here for clarity as follows:

- Chemistry diagnostic score (M0);
- Initial ALEKS knowledge check (M00);
- ASCIv2 factor 1: Intellectual Accessibility (M1);
- ASCIv2 factor 2: Emotional Satisfaction (M2);
- CSCI factor 1: Mathematical Self-Concept (M3);
- CSCI factor 2: Chemistry Self-Concept (M4);
- CSCI factor 3: Academic Self-Concept (M5);
- CSCI factor 4: Academic Enjoyment Self-Concept (M6);
- CSCI factor 5: Creativity Self-Concept (M7).

Finally, considering the structure of the quizzes themselves, each quiz consisted of pooled sets of questions drawn from a master question library stored in our course management system, which were randomized for each attempt. The quizzing structure and associated item libraries were developed and refined in summer term courses taught in 2020 and 2021. The quizzes used browser lockdown and webcam monitoring.

6. Results and Discussion

6.1. Item Descriptive Statistics and Correlations

As an overview, shown in Table 2 are descriptive statistics for the measures used in this work (M0–M7) across all sections. Considering the chemistry component of the diagnostic exam, the mean was 6.49 (scale of 10) with a standard deviation of 1.93. For the ASCIv2 and CSCI measures (M1–M7, on a five-point scale), the means varied from 2.61

to 3.98, while the standard deviations varied from 0.59 to 0.84. Apart from M00 (Aleks KC) and M6 (Academic Enjoyment SC), all skewness and kurtosis values were less than 1, indicating good normality. Measure M6 (Academic Enjoyment SC) displayed a mean centered towards the high end of the scale, which may impact on the discernment of differences in a longitudinal study.

Table 2. Descriptive statistics for the measures used in this work.

Item	Values	*N*	Mean (*SD*)	Median	Skewness (*SE*)	Kurtosis (*SE*)
M0	0–10	522	6.60 (1.92)	7.00	−0.27 (0.11)	−0.28 (0.21)
M00 *	0–10	167	2.90 (0.95)	2.77	0.78 (0.19)	1.45 (0.37)
M1	1–5	518	2.48 (0.62)	2.50	−0.07 (0.11)	−0.03 (0.21)
M2	1–5	518	3.33 (0.65)	3.25	−0.18 (0.11)	−0.08 (0.21)
M3	1–5	509	3.73 (0.78)	3.82	−0.64 (0.11)	0.11 (0.22)
M4	1–5	509	3.20 (0.71)	3.20	−0.28 (0.11)	0.03 (0.22)
M5	1–5	509	3.83 (0.44)	3.83	−0.21 (0.11)	−0.02 (0.22)
M6	1–5	509	4.14 (0.59)	4.14	−0.96 (0.11)	1.89 (0.22)
M7	1–5	509	3.35 (0.78)	3.25	−0.05 (0.11)	−0.55 (0.22)

* Only recorded for quiz-retake section.

Considering the quiz-retake general chemistry I section, here 167 of the 183 students completed the diagnostic (91.3%). To probe further the validity of the measures, we examined: (1) initial quiz average, (2) final quiz average (i.e., after retake if applicable), (3) difference in quiz average, (4) number of quiz retakes, (5) midterm exam, (6) final exam, (7) overall final score. Correlations amongst these variables were calculated using the Pearson statistic, with range between −1 and 1 [67]. The results are shown in Table S4. Considering the midterm and final exam scores, the correlations with measures M1–M7 were weak, as expected, since measures associated with attitude and self-concept should not correlate strongly with achievement, as pointed out by Bauer. The number of quiz retakes was negatively correlated with M3 (Mathematical SC) and positively correlated with M7 (Creativity SC). Similar trends were observed in the quiz-retake general chemistry II section, as shown in Table S5. Here the midterm and final exam scores also displayed weak correlations with measures M1–M7, but again both scores were strongly correlated with overall quiz average. The number of quiz retakes was negatively correlated with M1 (Intellectual Accessibility) and M3 (Mathematical SC).

6.2. Group Comparisons

Addressing our first research question, we examined trends across groups of students in the fall quiz-retake section, defining three quiz groups (Q0, Q1nr, Q1r) within that section, as follows. Group Q0 consisted of students who did not fail a quiz. Group Q1nr included students who failed at least one quiz and were invited to meet with the instructor but declined. Group Q1r included students who failed at least one quiz and completed at least one retake after meeting with the instructor.

Comparing groups Q1nr and Q1r, or discriminating between students who failed at least one quiz on the basis of retake completion, the statistics for these groups (Table S6) showed small differences in midterm and final exam score that were not statistically significant. Note that the mean number of retakes in the Q1r group was 1.6 (Table S6). The results of an independent samples t-test comparing these groups are shown in Table 3. Between these groups, none of the measured differences were significant. Note that in this table and those below, the n values for different measures are slightly different, as not all students completed all components of the diagnostic exam. Also, due to the multiple hypothesis tests that will be employed on these data, we applied the Bonferroni correction to calculate an adjusted α value of 0.01 for significance [68], which will be used throughout. Non-significant differences are identified in each table (n.s.).

Table 3. Independent *t*-test results for general chemistry I quiz-retake groups Q1nr vs. Q1r as described in the text.

Item	Group	*n*	Mean (*SD*)	*t* Statistic	*p* Value (Two-Tailed)	Cohen's *d*
M0	Q1nr	27	6.04 (1.56)	1.83	0.07 (n.s.)	...
	Q1r	36	5.17 (2.08)			
M00	Q1nr	27	2.76 (0.78)	1.49	0.14 (n.s.)	...
	Q1r	36	2.46 (0.78)			
M1	Q1nr	26	2.31 (0.54)	−0.30	0.77 (n.s.)	...
	Q1r	36	2.36 (0.67)			
M2	Q1nr	26	3.31 (0.62)	0.07	0.94 (n.s.)	...
	Q1r	36	3.30 (0.73)			
M3	Q1nr	25	3.73 (0.73)	1.52	0.13 (n.s.)	...
	Q1r	36	3.41 (0.84)			
M4	Q1nr	25	2.95 (0.78)	−0.52	0.60 (n.s.)	...
	Q1r	36	3.05 (0.68)			
M5	Q1nr	25	3.50 (0.55)	−0.95	0.34 (n.s.)	...
	Q1r	36	3.65 (0.61)			
M6	Q1nr	25	3.88 (0.59)	−1.12	0.27 (n.s.)	...
	Q1r	36	4.08 (0.78)			
M7	Q1nr	25	3.21 (1.00)	−2.59	0.02 (n.s.)	...
	Q1r	36	3.79 (0.74)			

The data for the spring quiz-retake general chemistry II section were also examined. Here the diagnostic exam was administered again at the beginning of the course, and while a different set of (10) chemistry-focused questions (i.e., M0) was employed and the ALEKS initial knowledge check (M00) was also different, the same ASCIv2 and CSCI instruments were administered. We further note that 101 students from the fall quiz-retake section continued into the spring quiz-retake section, and we will examine in detail below data for students common to both courses. Comparing here the groups Q1nr and Q1r for the general chemistry II course, the statistics for these groups (Table S7) showed a small difference in midterm score, but a marked difference (+7.3 for Q1r group) in final exam score, which was statistically significant with a medium effect size (Table 4). Again, the mean number of retakes in the Q1r group was around 1.5 (Table S7). The results of an independent samples *t*-test comparing these groups are shown in Table 4. Here with larger overall group sizes the difference in M6 (i.e., *Academic Enjoyment SC*) was significant, at a medium effect size, with the retake group showing a higher score.

Table 4. Independent *t*-test results for general chemistry II quiz-retake groups Q1nr vs. Q1r as described in the text.

Item	Group	*n*	Mean (*SD*)	*t* Statistic	*p* Value (Two-Tailed)	Cohen's *d*
M0	Q1nr	42	4.57 (1.75)	−1.07	0.29 (n.s.)	...
	Q1r	39	5.00 (1.85)			
M00	Q1nr	42	0.46 (0.18)	−0.54	0.59 (n.s.)	...
	Q1r	39	0.48 (0.19)			
M1	Q1nr	42	2.43 (0.59)	0.95	0.34 (n.s.)	...
	Q1r	36	2.29 (0.67)			
M2	Q1nr	42	3.19 (0.74)	−0.68	0.50 (n.s.)	...
	Q1r	36	3.30 (0.59)			
M3	Q1nr	41	3.45 (0.72)	0.61	0.55 (n.s.)	...
	Q1r	36	3.35 (0.63)			
M4	Q1nr	41	3.04 (0.55)	−1.43	0.16 (n.s.)	...
	Q1r	36	3.24 (0.67)			
M5	Q1nr	41	3.52 (0.59)	−2.01	0.048 (n.s.)	...
	Q1r	36	3.76 (0.44)			
M6	Q1nr	41	3.48 (0.54)	−3.63	<0.001	0.83
	Q1r	36	3.88 (0.41)			
M7	Q1nr	41	3.27 (0.65)	−0.80	0.42 (n.s.)	...
	Q1r	36	3.40 (0.75)			
Final	Q1nr	42	59.4 (18.4)	−2.01	0.048 (n.s.)	...
	Q1r	39	66.7 (14.1)			

6.3. Comparsons across General Chemistry I/II

To better understand the trends in quizzing retakes and address our second research question, we examined retakes across semesters within the various quiz groups in the cohort of students (n = 101) common to both quiz-retake sections. The results are shown in Table 5, which summarizes the cohort by quiz group in GCI and GCII. Beginning with students who failed at least one quiz in general chemistry I and completed a retake (GCI Group Q1r), we find that 41% (7 of 17) passed all quizzes in general chemistry II (i.e., were in GCII group Q0). Of the students in this group who failed at least one quiz in general chemistry II, roughly 60% (6 of 10) completed a retake. Considering next those students who failed at least one quiz and did not complete a retake in general chemistry I (group Q1nr), we find that only 13% (2 of 13) passed all quizzes in general chemistry II, and 38% (5 of 13) of the students failing at least one quiz in general chemistry II completed a retake. Summarizing, we thus find that the retake rate in general chemistry II was higher for students who had completed a retake in general chemistry I, and the percentage of students who failed at this one quiz in general chemistry I but passed all quizzes in general chemistry II was higher for students who had retaken at least one quiz in general chemistry I. Finally, considering students in Group Q0 in general chemistry I, those who passed all quizzes, 70% also passed all the quizzes in general chemistry II. For those failing a quiz, 43% (9 of 21) used the retake option, which was similar to the overall retake rate across general chemistry I and II.

Table 5. Statistics for quiz retakes across general chemistry I/II for the cohort of 101 students common to both sections.

GC I Group *	*n*	GC II Group	*n*	GC II Group %
Q1r	17	Q1r	6	35%
		Q1nr	4	24%
		Q0	7	41%
Q1nr	15	Q1r	5	33%
		Q1nr	8	53%
		Q0	2	13%
Q0	69	Q1r	9	13%
		Q1nr	12	17%
		Q0	48	70%

* Key: Q1r = students who failed at least one quiz and retook at least one, Q1nr = students who failed at least one quiz but did not complete a retake, Q0 = students who did not fail a quiz.

The data shown in Table 5 are informative regarding the impact of a low-stakes assessment model with retakes on the metacognition strategies of students [69]. Pazicni showed that students who reviewed their exam answers as a means to improve their grade not only gained in metacognition but also improved on subsequent exams [70]. This is consistent with the trend observed here, where within the limitations of the sample sizes it would appear that students who review failed quizzes are less likely to fail future quizzes and more likely to complete a retake if they do fail.

Returning to the affective domain measures, we compared measures M1–M7 using paired *t*-tests for all students common to both quiz sections, and the results are shown in Table 6. We find significant increases in M1 and M4 (Intellectual Accessibility and Chemistry SC) and a decrease in M6 (Academic Enjoyment SC), with effect sizes ranging from small to large. While naturally our chemistry course was not the only course taken by students, the opposite trend for chemistry self-concept vs. academic enjoyment self-concept is intriguing. In a longitudinal study of the restructuring of a one-semester chemistry course for nursing students using the CSCI, a statistically significant increase in chemistry self-concept was observed, with no statistically significant differences for the other subscales [71]. This was attributed to modifications in the learning environment that increased the relevance of course material and student motivation. Lewis and co-workers used the CSCI in a longitudinal (pre-test/post-test) study across a first term general chemistry course ($n = 83$) that employed a process-oriented, guided-inquiry learning framework, and similarly found a statistically significant increase only in chemistry self-concept [48]. This was taken as evidence that a student-centered pedagogy can lead to improvement in student self-concept, and led the authors to call for additional studies, like that reported here, which assess longitudinal changes in self-concept.

Addressing our third research question, we also examined the paired results sorted by binary gender, which is particularly desirable given the known impact of low-stakes assessments on the performance of women in STEM courses. The results from these comparisons are presented in Tables S8 and S9. Considering women, Table S8, M4 (Chemistry SC) showed a significant increase with medium effect size, while M6 (Academic Enjoyment SC) showed a significant decrease and a large effect size. Considering men, Table S9, only M1 (Intellectual Accessibility) showed a significant increase with medium effect size, while M3 (Mathematical SC) and M6 (Academic Enjoyment SC) showed significant decreases, with medium to large effect sizes.

Table 6. Paired *t*-test results across general chemistry I and II courses for students common to both quiz-retake sections.

Item	Comparison Group	*n*	Mean Difference (*SD*)	*t* Statistic	*p* Value (Two-Tailed)	Cohen's *d*
M1	GC II -GC I	94	0.22 (0.59)	3.54	<0.001	0.37
M2	GC II -GC I	94	0.12 (0.65)	1.81	0.07 (n.s.)	...
M3	GC II -GC I	93	−0.12 (0.47)	−2.47	0.015 (n.s)	...
M4	GC II -GC I	93	0.20 (0.58)	3.36	0.001	0.35
M5	GC II -GC I	93	−0.05 (0.44)	−1.18	0.24 (n.s.)	...
M6	GC II -GC I	93	−0.49 (0.49)	−9.67	<0.001	1.00
M7	GC II -GC I	93	−0.09 (0.64)	−1.32	0.19 (n.s.)	...

Unpacking these trends, we find that, across general chemistry I, chemistry self-concept increased for all students in the quiz-retake section, with a larger increase for women that reached statistical significance. The change in academic enjoyment self-concept of both men and women significantly decreased, but the change was larger for men. Finally, mathematical self-concept of men showed a significant decrease, in comparison to the negligible change observed for women.

Considering the further impact of this assessment strategy on women, across both semesters of the quiz-retake section we note that there were no statistically significant differences by gender in overall quiz average, midterm exam average, or final exam average. In general chemistry I, women completed more quiz retakes than men, while this trend reversed in general chemistry II. Taken together, these results suggest that an assessment strategy centered around low-stakes regular quizzing with a retake option may be effective in reducing or eliminating well-known gender disparities that occur with high-stakes assessments [1–5].

Finally, paired *t*-tests were used to compare measures M1–M7 across the general chemistry I quiz groups, focusing on the retake groups Q1r and Q1nr, and the results are shown in Tables S10 and S11. Given the small sample sizes in the two groups, these results should be considered tentative. For both groups we find statistically significant decreases in M6 (Academic Enjoyment SC), with a large effect size. This is consistent with the trend observed for all students, where a decrease in academic enjoyment self-concept across the first semester was observed. As the items in the CSCI coded to academic enjoyment inherently ask about multiple (i.e., most) academic subjects, it is hard to draw larger conclusions from this observation beyond the obvious one that, after their first semester, students in this cohort report on average a reduced enjoyment of their academic subjects.

7. Limitations

Two primary limitations of this work are that (1) it was carried out at a single institution and (2) results were not replicated due to scheduling issues. Considering the specific research goals of this study, particularly the desire to examine the evolution of student attitudes and self-concept across two semesters of general chemistry in sections that employed a quiz-retake model, a further limitation is the number of students common to both sections. This was caused in part by an unanticipated schedule change. Our general chemistry sections are traditionally offered on MWF at 8 a.m., 10 a.m. and 1 p.m., and the schedule change moved the primary instructor from the 1 pm general chemistry I section to the 8 am general chemistry II section. This change impacted the number of students following from the general chemistry I to II quiz-retake sections. The limited number of common students also prevented a more detailed analysis by student demographic (e.g., race/ethnicity, first generation status, etc.) that would be desirable.

A further limitation of this study is its reliance on quantitative measures of student affect. As pointed out by Flaherty [30], studies of affect in the chemistry education literature are dominantly (~90%) quantitative, and many have used the instruments employed here or other constructs. Given the complexity and subjectivity of affect, the premise of using numerical scales to measure it can be questioned, thus motivating the need for qualitative studies.

Considering the instruments used in this work, we have employed confirmatory factor analysis; however, ideally measurement invariance testing [72] would be applied to validate the use of these instruments across the groups in this study. Such an analysis unfortunately requires large data sets. As the authors of a recent primer on this approach note: "*For a practical approach, if measurement invariance testing is not feasible, we suggest a careful review of the literature for instruments which have been appropriately tested with diverse populations, to support appropriate data collection and analyses that lead to meaningful conclusions*" [72]. Given that both the ASCIv2 and CSCI instruments have been widely used and tested with diverse populations, we believe that their use is supported here; however, the lack of confirmation through measurement invariance testing is a limitation.

A final limitation of this study is that it was conducted in the COVID-19 era. All classes, recitation sections, and laboratory meetings were held in person, albeit with masking requirements that were lifted in the middle of the spring 2022 semester. Nonetheless, it is impossible to quantify the possible effects of the pandemic on this study.

8. Implications for Research and Practice

The change in assessment strategy in our general chemistry classes evidenced in Figure S1 shows a move from strictly summative assessments to a blend of formative and summative assessments. While the quizzes themselves are summative, students failing a quiz are afforded a retake after meeting with the course instructor. Thus, low-stakes quizzing is itself an effective early warning system that provides early feedback to students on their performance in the course and opportunities for remediation. Moreover, the use of this strategy demonstrates a growth vs. fixed mindset on the part of the instructor. This is important, as it has been demonstrated that faculty communicating a fixed mindset regarding ability undermine the performance of women in STEM fields [73].

Our longitudinal study shows that students who complete a retake in general chemistry I were more likely to do so in general chemistry II, and yet the percentage of students failing at least one quiz in general chemistry I but passing all quizzes in general chemistry II was markedly higher in the retake group, providing evidence of growth. Results from the applied instruments suggest that this strategy improves chemistry self-concept, particularly for women, and given the known gender disparities for high-stakes assessments, can reduce or eliminate such disparities. To that end, we note that an assessment strategy incorporating low-stakes assessments has now been widely adopted at our institution across a variety of chemistry courses.

9. Conclusions

This study has examined trends in student attitudes and self-concept, as measured using the ASCIv2 and CSCI instruments, across two semesters of general chemistry in sections employing a low-stakes assessment (quiz-retake) model. This model incorporated weekly quizzes as a major component of assessment, and offered retakes for students who failed a quiz following a one-on-one meeting with the instructor. This meeting consisted of a quiz review and discussion with suggestions for improvement in the course.

We gave the ASCIv2 and CSCI instruments across all sections (N = 624) of our fall 2021 semester general chemistry I course, and used confirmatory factor analysis to validate the previously derived factors. This led to a total of seven affective domain measures. The ASCIv2 and CSCI instruments were applied again at the beginning of the general chemistry II course in the quiz-retake section, affording a longitudinal evaluation of changes in student attitudes and self-concept for those students common to both sections. Comparing those students who were invited to retake a quiz and did vs. those who did not, in the fall cohort we find a significant difference in Creativity Self-Concept, with the retake group showing a higher score, while in the spring cohort these groups differed in Academic Enjoyment Self-Concept, again with the retake group showing a higher score.

Turning to the longitudinal comparison, we first examined the quiz-retake rate across both semesters, finding that in general chemistry II this rate was significantly higher for students who had completed at least one retake in general chemistry I. Secondly, we found that the percentage of students who failed at least one quiz in general chemistry I but passed all quizzes in general chemistry II was significantly higher for students who had completed at least one retake in general chemistry I. Considering the affective domain measures, by employing a paired analysis for students common to both quiz-retake sections, we find significant increases in factors corresponding to Intellectual Accessibility and Chemistry Self-Concept for students in the quiz-retake sections across semesters, with the former more pronounced for men and the latter for women. We also observe a significant decrease, for both men and women, in Academic Enjoyment Self-Concept, which is not necessarily reflective of their experience in chemistry.

Overall, we take these results to provide additional data supporting the benefit of low-stakes assessments with a retake option for development of a growth mindset, that may be particularly impactful for women in chemistry.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/educsci13121235/s1, Table S1: Demographic breakdown of the Fall 2021 sections; Table S2: Results of Confirmatory Factor Analysis of ASCIv2 items; Table S3: Results of Confirmatory Factor Analysis of CSCI items; Table S4: Correlations among measures in the quiz-retake section of general chemistry I; Table S5: Correlations among measures in the quiz-retake section of general chemistry II; Table S6: Group Statistics for the general chemistry I retake comparison; Table S7: Group Statistics for the general chemistry II retake comparison; Table S8: Paired *t*-test results for female students across general chemistry I and II courses for the quiz-retake sections; Table S9: Paired *t*-test results for male students across general chemistry I and II courses for the quiz-retake sections; Table S10: Paired *t*-test results for general chemistry I group Q1r students across general chemistry I and II courses for the quiz-retake sections; Table S11: Paired *t*-test results for general chemistry I group Q1nr students across general chemistry I and II courses for the quiz-retake sections; Figure S1: Evolution of assessment strategy; Figure S2: AMOS v29 path diagram with standardized estimates for the ASCIv2 CFA, using data from all sections; Figure S3: AMOS v29 path diagram with standardized estimates for the CSCI CFA, using data from all sections.

Author Contributions: Conceptualization, S.A.R.; methodology, S.A.R., J.R.G., L.N. and V.S.V.; formal analysis, S.A.R.; data curation, S.A.R.; writing—original draft preparation, S.A.R.; writing—review and editing, S.A.R., J.R.G., L.N. and V.S.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was submitted to the Institutional Review Board at our institution (IRB #3562) and was classified as exempt from further review.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Cantley, I.; McAllister, J. The Gender Similarities Hypothesis: Insights From A Multilevel Analysis of High-Stakes Examination Results in Mathematics. *Sex Roles* **2021**, *85*, 481–496. [CrossRef]
2. Cotner, S.; Ballen, C.J. Can mixed assessment methods make biology classes more equitable? *PLoS ONE* **2017**, *12*, e0189610. [CrossRef]

3. Montolio, D.; Taberner, P.A. Gender differences under test pressure and their impact on academic performance: A quasi-experimental design. *J. Econ. Behav. Organ.* **2021**, *191*, 1065–1090. [CrossRef]
4. O'Reilly, T.; McNamara, D.S. The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high-stakes" measures of high school students' science achievement. *Am. Educ. Res. J.* **2007**, *44*, 161–196. [CrossRef]
5. Salehi, S.; Cotner, S.; Azarin, S.M.; Carlson, E.E.; Driessen, M.; Ferry, V.E.; Harcombe, W.; McGaugh, S.; Wassenberg, D.; Yonas, A.; et al. Gender Performance Gaps Across Different Assessment Methods and the Underlying Mechanisms: The Case of Incoming Preparation and Test Anxiety. *Front. Educ.* **2019**, *4*, 107. [CrossRef]
6. Leeming, F.C. The exam-a-day procedure improves performance in psychology classes. *Teach. Psychol.* **2002**, *29*, 210–212. [CrossRef]
7. Pennebaker, J.W.; Gosling, S.D.; Ferrell, J.D. Daily Online Testing in Large Classes: Boosting College Performance while Reducing Achievement Gaps. *PLoS ONE* **2013**, *8*, e79774. [CrossRef]
8. Orr, R.; Foster, S. Increasing student success using online quizzing in introductory (majors) biology. *CBE Life Sci. Educ.* **2013**, *12*, 509–514. [CrossRef]
9. Walck-Shannon, E.M.; Cahill, M.J.; McDaniel, M.A.; Frey, R.F. Participation in Voluntary Re-quizzing Is Predictive of Increased Performance on Cumulative Assessments in Introductory Biology. *CBE Life Sci. Educ.* **2019**, *18*, ar15. [CrossRef]
10. Sullivan, S.G.B.; Hoiriis, K.T.; Paolucci, L. Description of a change in teaching methods and comparison of quizzes versus midterms scores in a research methods course. *J. Chiropr. Educ.* **2018**, *32*, 84–89. [CrossRef]
11. Casselman, B.L.; Atwood, C.H. Improving General Chemistry Course Performance through Online Homework-Based Metacognitive Training. *J. Chem. Educ.* **2017**, *94*, 1811–1821. [CrossRef]
12. Crucho, C.I.C.; Avo, J.; Diniz, A.M.; Gomes, M.J.S. Challenges in Teaching Organic Chemistry Remotely. *J. Chem. Educ.* **2020**, *97*, 3211–3216. [CrossRef]
13. Dempster, F.N. Distributing and managing the conditions of encoding and practice. In *Human Memory*; Bjork, E.L., Bjork, R.A., Eds.; Academic Press: Cambridge, MA, USA, 1996; pp. 197–236.
14. Dempster, F.N. Using tests to promote classroom learning. In *Handbook on Testing*; Dillon, R.F., Ed.; Greenwood Press: Westport, CT, USA, 1997; pp. 332–346.
15. Roediger, H.L.; Karpicke, J.D. The Power of Testing Memory Basic Research and Implications for Educational Practice. *Perspect. Psychol. Sci.* **2006**, *1*, 181–210. [CrossRef] [PubMed]
16. Bjork, R.A.; Bjork, E.L. A new theory of disuse and an old theory of stimulus fluctuation. In *From Learning Processes to Cognitive Processes: Essays in Honor of William K. Estes*; Healy, A., Kosslyn, S., Shiffrin, R., Eds.; Erlbaum: Mahwah, NJ, USA, 1992; Volume 2, pp. 35–67.
17. Bjork, R.A. Retrieval practice and the maintenance of knowledge. In *Practical Aspects of Memory: Current Research and Issues*; Gruneberg, M.M., Morris, P.E., Sykes, R.N., Eds.; Wiley: Hoboken, NJ, USA, 1988; Volume 1, pp. 396–401.
18. Dweck, C.S. *Mindset: The New Psychology of Success*; Random House: New York, NY, USA, 2016.
19. Murphy, S.; MacDonald, A.; Wang, C.A.; Danaia, L. Towards an Understanding of STEM Engagement: A Review of the Literature on Motivation and Academic Emotions. *Can. J. Sci. Math. Technol. Educ.* **2019**, *19*, 304–320. [CrossRef]
20. Richmond, J.M.; Baumgart, N. A Hierarchical Analysis of Environmental Attitudes. *J. Environ. Educ.* **1981**, *13*, 31–37. [CrossRef]
21. Haladyna, T.; Olsen, R.; Shaughnessy, J. Correlates of Class Attitude toward Science. *J. Res. Sci. Teach.* **1983**, *20*, 311–324. [CrossRef]
22. Shrigley, R.L. *Test of Science-Related Attitudes*; Wiley: Hoboken, NJ, USA, 1983; Volume 20, pp. 87–89. [CrossRef]
23. Schibeci, R.A. Students, Teachers, and the Assessment of Attitudes to School. *Aust. J. Educ.* **1984**, *28*, 17–24. [CrossRef]
24. Shrigley, R.L.; Koballa, T.R. Attitude Measurement—Judging the Emotional Intensity of Likert-Type Science Attitude Statements. *J. Res. Sci. Teach.* **1984**, *21*, 111–118. [CrossRef]
25. Schibeci, R.A.; Riley, J.P. Influence of Students Background and Perceptions on Science Attitudes and Achievement. *J. Res. Sci. Teach.* **1986**, *23*, 177–187. [CrossRef]
26. Shrigley, R.L.; Koballa, T.R.; Simpson, R.D. Defining Attitude for Science Educators. *J. Res. Sci. Teach.* **1988**, *25*, 659–678. [CrossRef]
27. Schibeci, R.A. Home, School, and Peer Group Influences on Student-Attitudes and Achievement in Science. *Sci. Educ.* **1989**, *73*, 13–24. [CrossRef]
28. Shrigley, R.L.; Koballa, T.R. A Decade of Attitude Research Based on Hovland Learning-Theory Model. *Sci. Educ.* **1992**, *76*, 17–42. [CrossRef]
29. Osborne, J.; Simon, S.; Collins, S. Attitudes towards science: A review of the literature and its implications. *Int. J. Sci. Educ.* **2003**, *25*, 1049–1079. [CrossRef]
30. Flaherty, A.A. A review of affective chemistry education research and its implications for future research. *Chem. Educ. Res. Pract.* **2020**, *21*, 698–713. [CrossRef]
31. Mager, R.F. *Developing Attitude toward Learning*; Fearon Publishers: Belmont, CA, USA, 1968.
32. Ajzen, I.; Madden, T.J. Prediction of Goal-Directed Behavior—Attitudes, Intentions, and Perceived Behavioral-Control. *J. Exp. Soc. Psychol.* **1986**, *22*, 453–474. [CrossRef]
33. Schunk, D.H. Self-Efficacy and Academic Motivation. *Educ. Psychol.* **1991**, *26*, 207–231. [CrossRef]
34. Schunk, D.H. Self-Efficacy, Motivation, and Performance. *J. Appl. Sport Psychol.* **1995**, *7*, 112–137. [CrossRef]

35. Gable, R.K. The Measurement of Attitudes toward People with Disabilities—Methods, Psychometrics and Scales. *Contemp. Psychol.* **1990**, *35*, 81. [CrossRef]
36. Krumrei-Mancuso, E.J.; Newton, F.B.; Kim, E.; Wilcox, D. Psychosocial Factors Predicting First-Year College Student Success. *J. Coll. Stud. Dev.* **2013**, *54*, 247–266. [CrossRef]
37. Beane, J.A.; Lipka, R.P.; Ludewig, J.W. Synthesis of Research on Self-Concept. *Educ. Leadersh.* **1980**, *38*, 84–89.
38. Beane, J.A.; Lipka, R.P. Self-Concept and Self-Esteem—A Construct Differentiation. *Child Study J.* **1980**, *10*, 1–6.
39. Marsh, H.W.; Walker, R.; Debus, R. Subject-Specific Components of Academic Self-Concept and Self-Efficacy. *Contemp. Educ. Psychol.* **1991**, *16*, 331–345. [CrossRef]
40. Hassan, A.M.A.; Shrigley, R.L. Designing a Likert scale to measure chemistry attitudes. *Sch. Sci. Math.* **1984**, *84*, 659–669. [CrossRef]
41. Bauer, C.F. Beyond "student attitudes": Chemistry self-concept inventory for assessment of the affective component of student learning. *J. Chem. Educ.* **2005**, *82*, 1864–1870. [CrossRef]
42. Nielsen, S.E.; Yezierski, E. Exploring the Structure and Function of the Chemistry Self-Concept Inventory with High School Chemistry Students. *J. Chem. Educ.* **2015**, *92*, 1782–1789. [CrossRef]
43. Nielsen, S.E.; Yezierski, E.J. Beyond academic tracking: Using cluster analysis and self-organizing maps to investigate secondary students' chemistry self-concept. *Chem. Educ. Res. Pract.* **2016**, *17*, 711–722. [CrossRef]
44. Temel, S.; Sen, S.; Yilmaz, A. Validity and Reliability Analyses for Chemistry Self-Concept Inventory. *J. Balt. Sci. Educ.* **2015**, *14*, 599–606. [CrossRef]
45. Werner, S.M.; Chen, Y.; Stieff, M. Examining the Psychometric Properties of the Chemistry Self-Concept Inventory Using Rasch Modeling. *J. Chem. Educ.* **2021**, *98*, 3412–3420. [CrossRef]
46. Grove, N.; Bretz, S.L. CHEMX: An instrument to assess students' cognitive expectations for learning chemistry. *J. Chem. Educ.* **2007**, *84*, 1524–1529. [CrossRef]
47. Dalgety, J.; Coll, R.K.; Jones, A. Development of Chemistry Attitudes and Experiences Questionnaire (CAEQ). *J. Res. Sci. Teach.* **2003**, *40*, 649–668. [CrossRef]
48. Lewis, S.E.; Shaw, J.L.; Heitz, J.O.; Webster, G.H. Attitude Counts: Self-Concept and Success in General Chemistry. *J. Chem. Educ.* **2009**, *86*, 744–749. [CrossRef]
49. Bauer, C.F. Attitude towards chemistry: A semantic differential instrument for assessing curriculum impacts. *J. Chem. Educ.* **2008**, *85*, 1440–1445. [CrossRef]
50. Xu, X.Y.; Lewis, J.E. Refinement of a Chemistry Attitude Measure for College Students. *J. Chem. Educ.* **2011**, *88*, 561–568. [CrossRef]
51. Brandriet, A.R.; Xu, X.Y.; Bretz, S.L.; Lewis, J.E. Diagnosing changes in attitude in first-year college chemistry students with a shortened version of Bauer's semantic differential. *Chem. Educ. Res. Pract.* **2011**, *12*, 271–278. [CrossRef]
52. Brandriet, A.R.; Ward, R.M.; Bretz, S.L. Modeling meaningful learning in chemistry using structural equation modeling. *Chem. Educ. Res. Pract.* **2013**, *14*, 421–430. [CrossRef]
53. Montes, L.H.; Ferreira, R.A.; Rodriguez, C. Explaining secondary school students' attitudes towards chemistry in Chile. *Chem. Educ. Res. Pract.* **2018**, *19*, 533–542. [CrossRef]
54. Kahveci, A. Assessing high school students' attitudes toward chemistry with a shortened semantic differential. *Chem. Educ. Res. Pract.* **2015**, *16*, 283–292. [CrossRef]
55. Chadwick, S.; Baker, A.; Alexander, J. Implementing ASCIv2 to measure students' attitudes towards chemistry in large enrolment chemistry subjects at an Australian university. *Abstr. Pap. Am. Chem. Soc.* **2015**, *249*, 1155.
56. Cha, J.; Kan, S.Y.; Wahab, N.H.A.; Aziz, A.N.; Chia, P.W. Incorporation of Brainteaser Game in Basic Organic Chemistry Course to Enhance Students' Attitude and Academic Achievement. *J. Korean Chem. Soc.* **2017**, *61*, 218–222. [CrossRef]
57. An, J.; Loppnow, G.R.; Holme, T.A. Measuring the impact of incorporating systems thinking into general chemistry on affective components of student learning. *Can. J. Chem.* **2021**, *99*, 698–705. [CrossRef]
58. Morshead, R.W.; Krathwohl, D.R.; Bloom, B.S.; Masia, B.B. Taxonomy of Educational-Objectives Handbook Ii—Affective Domain. *Stud. Philos. Educ.* **1965**, *4*, 164–170. [CrossRef]
59. Haertel, G.D.; Walberg, H.J.; Weinstein, T. Psychological Models of Educational Performance—A Theoretical Synthesis of Constructs. *Rev. Educ. Res.* **1983**, *53*, 75–91. [CrossRef]
60. Gable, R.K.; Ludlow, L.H.; Wolf, M.B. The Use of Classical and Rasch Latent Trait Models to Enhance the Validity of Affective Measures. *Educ. Psychol. Meas.* **1990**, *50*, 869–878. [CrossRef]
61. Anderson, M.C.; Bjork, R.A.; Bjork, E.L. Remembering Can Cause Forgetting—Retrieval Dynamics in Long-Term-Memory. *J. Exp. Psychol. Learn.* **1994**, *20*, 1063–1087. [CrossRef] [PubMed]
62. Bjork, E.L.; Little, J.L.; Storm, B.C. Multiple-choice testing as a desirable difficulty in the classroom. *J. Appl. Res. Mem. Cogn.* **2014**, *3*, 165–170. [CrossRef]
63. IBM. *IBM SPSS Statistics*; IBM: Armonk, NY, USA, 2021.
64. McGraw Hill. *ALEKS Corporation*; McGraw Hill: New York, NY, USA, 2021; Available online: https://www.aleks.com/ (accessed on 1 January 2023).
65. Amos Development Corporation. *IBM SPSS Amos*; Amos Development Corporation: McLean, VA, USA, 2022.
66. Cronbach, L.J. My current thoughts on coefficient alpha and successor procedures. *Educ. Psychol. Meas.* **2004**, *64*, 391–418. [CrossRef]

67. Williams, S. Pearson's correlation coefficient. *N. Z. Med. J.* **1996**, *109*, 38.
68. Kaltenbach, H.-M. *A Concise Guide to Statistics*; Springer: Berlin/Heidelberg, Germany, 2012.
69. Chan, J.Y.K.; Bauer, C.F. Identifying At-Risk Students in General Chemistry via Cluster Analysis of Affective Characteristics. *J. Chem. Educ.* **2014**, *91*, 1417–1425. [CrossRef]
70. Pazicni, S. Mitigating Students' Illusions of Competence with Self-Assessment. In Proceedings of the 247th ACS National Meeting, Dallas, TX, USA, 16–20 March 2014. Paper 1529.
71. Smith, A.L.; Paddock, J.R.; Vaughan, J.M.; Parkin, D.W. Promoting Nursing Students' Chemistry Success in a Collegiate Active Learning Environment: "If I Have Hope, I Will Try Harder". *J. Chem. Educ.* **2018**, *95*, 1929–1938. [CrossRef]
72. Rocabado, G.A.; Komperda, R.; Lewis, J.E.; Barbera, J. Addressing diversity and inclusion through group comparisons: A primer on measurement invariance testing. *Chem. Educ. Res. Pract.* **2020**, *21*, 969–988. [CrossRef]
73. Canning, E.A.; Ozier, E.; Williams, H.E.; AlRasheed, R.; Murphy, M.C. Professors Who Signal a Fixed Mindset About Ability Undermine Women's Performance in STEM. *Soc. Psychol. Personal. Sci.* **2022**, *13*, 927–937. [CrossRef]