Chemistry Education Research and Practice

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d3rp00148b

A lack of impact of pedagogy (peer-led team learning compared with didactic instruction) on long-term student knowledge of chemical equilibrium

Jessica D. Young,[a] Betül Demirdöğen,[b] Christopher F. Bauer[c] and Scott E. Lewis*[a]

Received 15th June 2023,
Accepted 20th August 2024

DOI: 10.1039/d3rp00148b

rsc.li/cerp

Peer-led team learning is a socially mediated pedagogy where trained peer leaders, students who have completed a course, return to lead students in groups within a targeted course. The effect of peer-led team learning to improve student success in chemistry has been extensively documented but it is unclear if it is just as effective at facilitating retention of knowledge across time. This paper describes two studies designed to examine this possibility, each focusing on the impact of peer-led team learning in second-semester general chemistry on students' long-term knowledge of chemical equilibrium. The first study measured student knowledge at three time points for one year following enrollment in general chemistry. The second study measured student knowledge while enrolled in analytical chemistry. Both studies used a repeated measures design and found no demonstrable effect of pedagogy on the long-term retention of knowledge. This finding indicates that concepts students hold in first-year chemistry remain long-standing throughout their undergraduate training, conceptual understanding of equilibrium shows ample room for improvement across both pedagogies, and peer-led team learning supports knowledge retention comparable to didactic instruction.

Introduction

The effect of active learning pedagogies to improve student success has been extensively documented in the science education and chemistry education research literature (Freeman *et al.*, 2014; Rahman and Lewis, 2020). One nuance that stands out in these analyses though is the diminished impact of active learning observed on cumulative end-of-term exams relative to the impact on single topic in-term exams (Apugliese and Lewis, 2017; Rahman and Lewis, 2020). One possible explanation is that active learning facilitated learning in the moment but was less effective at facilitating retention of knowledge across time. This paper describes two studies that examine the impact of active learning on retention of knowledge across time. The active learning pedagogy of interest was peer-led team learning (PLTL) enacted in second-semester general chemistry. Data was collected on students' knowledge of chemical equilibrium, a foundational topic in the target class, at differing points in time and at two institutional settings.

Background literature

Peer-led team learning

Peer-led team learning (PLTL) is a pedagogy designed to promote socially mediated learning in large classes. PLTL relies on trained peer leaders, students who have successfully completed a course and then return to the course to lead the current cohort of students as they work in groups on content related tasks. PLTL readily scales with large class sizes as peer leaders can be recruited to maintain a desired student to peer leader ratio. The PLTL model describes six critical criteria for implementation (Varma-Nelson *et al.*, 2004). First, PLTL is integrated into a course, which can be either compulsory or voluntary for students, where the content students work on is directly applicable to the course setting. Second, the course instructor is involved with program implementation and peer leader training. Third, peer leaders are trained to promote active learning and supervised. Fourth, the group activities are sufficiently challenging to promote collaboration. Fifth, class logistics are set up to promote group work. Finally, the department and institution demonstrate support for the program.

---

[a] *University of South Florida, Chemistry, Tampa, Florida, USA. E-mail: slewis@usf.edu*

[b] *Zonguldak Bülent Ecevit Universitesi, Secondary Science and Mathematics Education, Zonguldak, Turkey*

[c] *University of New Hampshire, Chemistry, Durham, New Hampshire, USA*

A considerable literature base has been developed on the evaluation of PLTL on post-secondary chemistry students' academic performance. Rahman and Lewis (2020) conducted a meta-analysis that located seven quasi-experimental studies that compared student performance with PLTL instruction to student performance in a teaching-as-normal condition in chemistry. Across these studies, students with PLTL averaged a half standard-deviation higher on assessments than students with primarily didactic instruction. Similarly, Leontyev and colleagues (2017) conducted a meta-analysis of sixteen studies comparing PLTL in chemistry to traditional instruction and found PLTL averaged 0.36 of a standard deviation higher. Some studies in these meta-analyses did not observe a difference in academic performance but did observe a higher rate of students completing the course (Lewis, 2011; Mitchell *et al.*, 2012). More recently, a study found that PLTL integrated into entirely online instruction led to gains of approximately a fifth of a standard deviation relative to online, didactic instruction (Young and Lewis, 2022).

Theoretical framework

Multiple theories plausibly describe how PLTL leads to academic gains, as recently reviewed in Frey and Lewis (2023). The most often cited theory in the PLTL literature invokes Vygotsky's Zone of Proximal Development. The zone of proximal development describes skills that learners can perform when the skill is modeled, and it is thought that instruction in this zone would promote the learners' skill set. Peer leaders may be well placed for situating instruction within a learner's zone of proximal development, as students are more similar in conceptual preparation to peer leaders than students are to expert course instructors. PLTL also facilitates cooperative learning, as students in PLTL work together on a common problem set. Multiple theories exist for the efficacy of cooperative learning to promote learning gains (Slavin, 1996). Cooperative learning promotes students generating explanations which requires a mental reorganizing of concepts which may lead to greater learning. Within PLTL students may generate explanations more often than students experiencing didactic instruction, which may explain the impact of PLTL. Alternatively, the PLTL structure may lead to students normalizing challenges they experience in chemistry or see learning as a shared goal instead of an individual effort. As a result, students with PLTL instruction may experience more motivation to learn chemistry and exhibit greater persistence in their studies than students with didactic instruction. The academic gains observed with PLTL in the literature could be accounted for by any of these theories. Although multiple possible mechanisms for PLTL to impact learning are plausible, it is assumed within this study that PLTL does impact how students learn chemistry which is supported by the academic gains observed with PLTL in the literature.

How students learn and understand content is expected to impact their memory and the retention of knowledge from that content. As reviewed by Wang (2021), memory is sustained by the meaning making generated by an individual. The act of remembering is influenced by an individual's interpretation and organization of information learned, cultural background including social expectations, experiences after learning the material, and the mental, physical and social context where remembering takes place (Wang, 2021). It has been posited that teaching methods impact students' knowledge retention (Engelbrecht *et al.*, 2007), although it is important to acknowledge that other factors will influence knowledge retention. This supposition is supported by studies in mathematics education (Narli, 2011; Förster *et al.*, 2022) and chemistry education (Underwood *et al.*, 2016) that have shown the impact of pedagogy on retention of knowledge post-course.

Past work on knowledge retention in chemistry

Multiple research studies that evaluate pedagogical reform in chemistry education have called for longitudinal studies to examine knowledge retention over time (Bodé *et al.*, 2016; Çalık and Cobern, 2017; Chase *et al.*, 2017) however such studies are infrequent. Longitudinal studies can adopt a variety of research designs to explore knowledge retention. First, a within-subjects design has the same group of students complete the same instrument at multiple time points. By comparing student performance from one time point to the other, the extent of knowledge retention can be ascertained. One challenge with this design is determining what extent of knowledge retention is to be expected, that is, what extent of knowledge retention is needed to be deemed satisfactory. A second research design is a between-subjects design. In this case, students from two methods of instruction are measured on some construct at a later point in time and the comparison between groups can provide context for setting an expectation of knowledge retention. Third, researchers can invoke a between-subjects and within-subjects design concurrently. In this design, students from two methods of instruction are each given the same instrument repeated over time. This design allows for the comparison of one group across time (within-subjects) and the comparison between the two groups at any time point (between-subjects).

Two examples of a within-subjects longitudinal design in chemistry education are from Coştu and colleagues (2010) and Shah and colleagues (2018). Each study sought to evaluate a pedagogical intervention by administering an instrument to students with a pretest, posttest, and delayed posttest design. In each case, the instrument was topic specific and matched to the topic focus of the pedagogical intervention. Both studies found a gain from the pretest to the post-test and no evidence of statistical significance in the differences from the posttest to the delayed post-test. These findings led to each study providing support for pedagogical effectiveness to promote knowledge owing to the gain from pretest to posttest, and to retain knowledge owing to the lack of difference from posttest to delayed posttest.

Two examples of a between-subjects design have been conducted to evaluate the efficacy of PLTL. Mitchell and colleagues (2012) and Lewis (2014) compared students who took general chemistry with either PLTL pedagogy or a didactic-based pedagogy. Students were measured based on their enrollment rates and course grades in follow-on courses. Both studies found an increase in enrollment rates among students with PLTL-based pedagogy but no statistically significant difference in course grades. However, course grades in follow-on courses are broad metrics that are partially dependent on learning from prior coursework but also dependent on students' actions and motivations within the course for which the grades are being assigned. Thus, they may not serve as a sufficiently specific metric to investigate retention of knowledge.

To determine long-term impact of a pedagogy on student knowledge it is likely necessary to measure academic performance using an instrument tailored to the content that was learned with the pedagogy as done in the within-measures design by Coştu and colleagues (2010) and Shah and colleagues (2018). Further, to contextualize the observed trends, it is beneficial to have the same measures over time administered to students from multiple pedagogies, that is a between-subjects and within-subjects design. Underwood and colleagues (2016) offer one example of this design in chemistry; they found that students in the chemistry, life, the universe, and everything (CLUE) curriculum connected structure and property earlier than students in the traditional curriculum and the differences remain after four semesters. The current work seeks to enact a between-subjects and within-subjects design to evaluate the impact of PLTL in second-semester general chemistry on long-term student knowledge of a particular topic. The topic chosen was chemical equilibrium owing to its central role in the targeted course.

Teaching and learning chemical equilibrium

The challenging topic of chemical equilibrium is fundamental to beginner students' subsequent learning of other chemistry topics. Meaningful learning of chemical equilibrium requires adequate understanding of preceding chemistry topics (*e.g.*, structure of matter, kinetics, thermodynamics, and stoichiometry) (Pedrosa and Dias, 2000; Greenbowe *et al.*, 2007). Chemical equilibrium itself has sub-concepts classified as 'matter' or 'things' (*e.g.*, substances, open and closed system) and 'process' or 'constraint-based interactions' (*e.g.*, all entities that exist at equilibrium, dynamic nature of equilibrium) (Chiu *et al.*, 2002). Connecting ideas across the various preceding topics and within the sub-concepts makes chemical equilibrium an integrated topic in chemistry (Ganaras *et al.*, 2008; Quílez, 2009). Moreover, student understanding of equilibrium requires conceptualizing the topic across the macroscopic, submicroscopic, and symbolic domains (Johnstone, 2000a, b; Ganaras *et al.*, 2008; Talanquer, 2011). For example, how a system reaches equilibrium could be observed macroscopically through observation (*e.g.*, color change) or instruments (*e.g.*, pressure change), can be described through theoretical models that describe submicroscopic entities, and *via* various symbolic representations including mathematical equations, drawings, and graphs. The complex characteristics of knowledge in chemical equilibrium make this topic a challenging one for chemistry students to learn (Ganaras *et al.*, 2008).

Chemical equilibrium has been described as an anchoring concept within the content of general chemistry (Holme *et al.*, 2015). Moreover, students need to utilize chemical equilibrium concepts when learning about content that is covered in organic chemistry, inorganic chemistry, physical chemistry, and analytical chemistry. Research has also provided evidence for the central role of chemical equilibrium on students' success in thermodynamics (Bain *et al.*, 2014), biochemistry (Wolfson *et al.*, 2014), phase transitions (Azizoğlu *et al.*, 2006), organic chemistry (Cartrette and Mayo, 2011), and analytical chemistry (Nyachwaya, 2016). Ideas and concepts in chemical equilibrium help to observe experiments (*e.g.*, identification tests, titration, synthesis of chemical substances) and interpret the data obtained from experiments as well (Ganaras *et al.*, 2008). That is, chemical equilibrium is a central chemistry concept that underlies different topics across various chemistry disciplines and enables students to explain a wide variety of experiments (Ganaras *et al.*, 2008).

A substantial literature base describes the complex and abstract nature of learning chemical equilibrium and student challenges exhibited when learning equilibrium (Chiu *et al.*, 2002; Kousathana and Tsaparlis, 2002; Ganaras *et al.*, 2008; Quílez, 2009). Integrating an understanding of chemical kinetics into chemical equilibrium poses a particular challenge (Turányi and Tóth, 2013). Language used in textbooks (Pedrosa and Dias, 2000) and concepts of stoichiometry (Greenbowe *et al.*, 2007) may also explain varying alternative conceptions related to chemical equilibrium that have been observed. Alternative conceptions include that the concentrations of reactants and products are equal at equilibrium or that the reaction has stopped at chemical equilibrium (Thomas and Schwenz, 1998; Özmen, 2008; Demircioğlu *et al.*, 2013), the equilibrium constant changes with pressure (Thomas and Schwenz, 1998) and concentration (Özmen, 2008), the constant increases with an increase in temperature (Özmen, 2008), and the constant is independent of the change in entropy and enthalpy of a chemical reaction (Thomas and Schwenz, 1998). In terms of Le Chatelier's principle, alternative conceptions include that a catalyst would lead to a change in the production of reactants or products, that a change in the temperature will not affect the equilibrium, and that Le Chatelier's principle is applicable to compounds that are in the solid or liquid state of matter (Thomas and Schwenz, 1998; Özmen, 2008; Demircioğlu *et al.*, 2013).

Considering the essential nature of chemical equilibrium in learning various chemistry disciplines and prevalence of alternative conceptions about chemical equilibrium among university students, exploring student knowledge of this topic over time can be impactful to curriculum design. Literature has provided evidence that instruction as normal has not resulted in students' mastery of chemical equilibrium concepts (Ganaras *et al.*, 2008; Quílez, 2009; Wolfson *et al.*, 2014). Thomas and Schwenz (1998) proposed that active learning supplementing a lecture format may promote student success on this topic. PLTL is an example of active learning that may contribute to students' understanding of chemical equilibrium.

Rationale and research question

This study seeks to examine the longitudinal benefits of PLTL by employing a within-subjects and between-subjects study design. In designing this study, an assessment on chemical equilibrium was developed and given at varying time points following enrollment in the class enabling a within-subjects comparison. The participants in this study are students who were previously enrolled in second-semester general chemistry classes that used either PLTL or didactic-based instruction. By sampling from both pedagogies, a between-subject comparison is enabled. To triangulate the observed results, a similar research design was conducted in two separate studies, herein referred to as Study 1 and Study 2. Study 1 surveys student volunteers following the completion of general chemistry, while Study 2 surveys students enrolled in an Analytical Chemistry class, which utilizes an understanding of chemical equilibrium. Each study was guided by the same over-arching research question: what is the impact of PLTL pedagogy in second-semester general chemistry on students' knowledge of equilibrium over time?

Methods

Setting

This research takes place at two research settings. The first research setting is a large, research-intensive university in the southeast United States. The second research setting is a smaller research-intensive university in the northeast United States. At both settings, general chemistry is a two-semester course that is required for students majoring in any of the natural sciences or planning to pursue most health professions. The second-semester course covers the following topic sequence: intermolecular forces, colligative properties, chemical kinetics, equilibrium, weak acids and bases, buffer solutions, solubility, thermodynamics (introduction of entropy and free energy), and electrochemistry. At both settings, multiple classes of the course are offered each semester. At the first research setting, the classes are coordinated using a common syllabus, common exams at the same time, and class meetings are twice weekly for 75 minutes. At the second research setting class meetings comprise 150 to 160 minutes per week in two or three class meetings per week.

The implementation of PLTL at the two settings was designed to follow the six critical criteria articulated in the PLTL literature (Varma-Nelson *et al.*, 2004).

First, at both research settings, PLTL is integrated into the course. At the first setting, PLTL is compulsory in weekly meetings that account for approximately one-half of structured class time. PLTL is enacted in two classes of second-semester general chemistry each semester, while anywhere from one to five additional classes of general chemistry each semester do not use PLTL. At the second setting three classes are offered and PLTL is an option with students opting in to weekly meetings outside of lecture time. At both settings, the content is selected based on applicability to the course objectives and in similar scope to the course assessments.

Second, the course instructors are involved in implementation and peer leader training. At both research settings, instructors play an active role in designing and selecting problems that students will engage in during the PLTL sessions. At the first research setting, instructors are present during the PLTL sessions which occur within a lecture hall and the peer leaders are trained by one of the course instructors that regularly teaches general chemistry. The instructor monitored student progress throughout the session, gave feedback to the peer leaders, and on rare occasion gave brief (less than 5 minute) presentations to clarify a question or particular topic. At the second research setting, the leaders meet on their own once a week with their student teams. Leaders also meet weekly with each other and one of the course lecture instructors, who is also the PLTL course trainer, to review topics to include in PLTL sessions.

Third, at both research settings peer leaders are trained in active learning and supervised. At the first research setting, training takes place weekly where the trainer models a peer-led session and the peer leaders take the role of the students. During these sessions, peer leaders work through the same problem set their students will see later that week. This approach facilitates a discussion of pedagogical content knowledge, where peer leaders and the trainer discuss instructional resources (*e.g.* follow-up questions, relevant text) that are tailored to the content and support active learning. As a result of the training peer leaders become aware of the canonical solutions to the problem sets, but are instructed to prompt students to describe the process for solving problems and provide feedback on students' problem solving processes. At the second research setting peer leaders for this course had one or more prior semesters of experience as leaders, and had two training courses focused on cognitive development, STEM pedagogical content knowledge, metacognition, and group dynamics. In addition to modeling active-learning during training and engaging in mock sessions, leaders visit and are visited by veteran leaders to obtain formative feedback on their facilitation skills.

Fourth, at both research settings, instructors design problem sets that are meant to challenge students, promote collaboration, and match the course objectives. Examples of problem sets used during the PLTL sessions are included in the Appendix Part 1. At the first research setting, each peer leader was assigned approximately twelve students and the leaders were instructed to promote collaboration within groups of two to four students. This structure deviates from the literature which describes a peer leader working with a single group of six to eight students (Varma-Nelson *et al.*, 2004) which was necessitated by the course logistics and matches other reports of PLTL in the literature (Frey and Lewis, 2023). At the second research setting, peer leaders meet with a group of five to ten students for 80 minutes.

Fifth, the class logistics at the first research setting was a lecture hall with fixed seats. Students were assigned to sit in a particular row in the lecture hall and each row was assigned to one peer leader. The peer leader encouraged students to work with those seated directly next to them and communicated with students as a group. At the second research setting peer leaders met with students weekly outside of lecture.

Sixth, both peer leading programs have a long-standing presence of greater than five years within the department and institution at each research setting. The authors each have used PLTL in chemistry at each setting to contribute to the existing literature base on PLTL effectiveness (Chan and Bauer 2015, Robert *et al.* 2016).

At both study settings, equilibrium was introduced over approximately one week of class time, following a sequence of topics that mirrors many general chemistry textbooks. Building on chemical kinetics, students are introduced to the concept of the reverse reaction and that equilibrium will inevitably occur when the forward and reverse rates equal each other. Equilibrium is described macroscopically in terms of stable concentration values. Students are then introduced to the equilibrium constant and its formulation based on the chemical species within a chemical reaction, and how manipulations of a chemical reaction can lead to predictable manipulations of the equilibrium constant. Next students are introduced to Le Chatelier's principle and predictions that result from disturbances to a chemical equilibrium. Subsequent instruction, lasting two to three weeks, revisited these concepts when presenting equilibrium calculations and equilibrium applications in acid–base or solubility systems.

Didactic instruction presented these topics primarily through lecture and presenting worked examples of typical problems associated with these topics. In the first research setting, PLTL replaced approximately one half of lecture time with students working through problems related to the topics in PLTL sessions. In the second research setting, students with PLTL worked related problems in addition to the lecture instruction.

Instrument

A thirteen-item, multiple choice instrument to measure students' knowledge of chemical equilibrium was developed to be administered online *via* Qualtrics. The instrument development began with a set of nine free-response items and four multiple choice items created by two of the authors who each had experience teaching chemical equilibrium. This set of items was administered to students enrolled in Analytical Chemistry at four institutions including the two institutions participating in this study. Student responses to the open-ended questions were coded and the most frequent responses were used to develop multiple choice distractors. Five students from the sample participated in think-aloud interviews to determine if the participants interpreted the items as the instrument authors intended. The written responses and interview responses were used to modify the items to improve item clarity. For example, an original free-response item was:

A chemical reaction $B_2H_6(g) \rightleftarrows 2B(s) + 3H_2(g)$ is at equilibrium in a container. Then additional $B_2H_6(g)$ is added to the container. Describe what happens next to the concentrations of all three chemical substances.

Student responses to the items showed some students described the concentration of B_2H_6 relative to the concentration before the addition and others relative to the concentration after the addition. To avoid a misreading of the question, the final multiple-choice item was crafted as follows:

A chemical reaction $B_2H_6(g) \rightleftarrows 2B(s) + 3H_2(g)$ is at equilibrium in a container. Next, additional $B_2H_6(g)$ is added to the container, increasing its concentration. **After** the addition of $B_2H_6(g)$ describe what happens to the concentrations of all three chemical substances.

	Increases	Decreases	Does not change
$B_2H_6(g)$			
$B(s)$			
$H_2(g)$			

Response process validity was obtained for the revised instrument in the form of think-aloud interviews with nine students (three Analytical Chemistry students and six peer leaders) and no further revisions to the instrument were made.

The items are presented in the Appendix Part 2 and were designed to measure specific learning objectives (Table 1). The learning objectives were developed by the instructors at the first research setting and confirmed as applicable by an instructor at the second research setting. Of note, the first and last learning objectives describe the dynamics of a non-equilibrium situation, while the middle three objectives pertain to the meaning of the equilibrium constant. Since the intended administration was an online survey, the instrument design was focused on concepts of equilibrium and avoided lengthier calculations (*e.g.*, solving for equilibrium concentrations) owing to a concern that research participants may not persist through the calculations.

Given the sample sizes of Study 1 (N = 82) and Study 2 (N = 49), confirmatory factor analyses could reasonably be done only with the Study 1 sample. The confirmatory factor analyses on the responses to time points 1, 2, and 3 in Study 1 (see

Table 1 Learning objectives and corresponding items in the instrument

Learning objectives	Item/s
Define dynamic equilibrium and make predictions regarding relative concentrations and rates of change for a system at dynamic equilibrium	1–5
Describe conceptually the significance of very small or very large values for K	6
Given a chemical reaction and concentrations at equilibrium solve for the value of K	7–8
Indicate how modifying a chemical reaction (reversing, multiplying by a constant or combining reactions) impacts the value for K	9
Use Le Chatelier's principle to predict the direction the reaction will proceed	11–13

Appendix Part 3) modeling the thirteen items onto a single latent dimension showed acceptable goodness of fit indices (CFI = 0.96, 0.96, 0.96; RMSEA = 0.04, 0.04, 0.04). However, factor loadings indicated items (7 and 11) with problematic loadings less than 0.32 (Brown, 2015). These items were removed leaving eleven items. The subsequent confirmatory factor analyses on the responses to time points 1, 2, and 3 in Study 1 modeling the eleven items onto a single latent dimension showed acceptable goodness of fit indices (CFI = 1.00, 0.95, 0.97; RMSEA = 0.00, 0.06, 0.05; McDonald's ω = 0.76, 0.75, 0.74). As well, the factor loadings indicated a tenable model. The instrument was thus scored by the overall percent correct obtained from the eleven items. The following analyses were also conducted retaining all thirteen items and no substantive differences in conclusions were found. Two of the items are multi-part questions and each part was weighed as partial credit for the item.

Statement on ethical considerations

Study 1 and 2 were approved by the Institutional Review Boards at research setting #1 (Pro00031819) and research setting #2 (IRB 7002). In Study 1, students were offered a \$20 gift card for completing a survey at each of the three time points. In Study 2, students were offered a small portion of extra credit toward their final grade for each survey that was completed. In both studies, recruitment was conducted by a research team member that was not affiliated with the instruction in the course to minimize potential perceptions of coercion. As part of the informed consent process, students were given written instructions describing the procedures asked of students and voluntary nature of participation. Following receiving the instructions, students consented to participate.

Study 1

Methods specific to study 1

Study 1 sought to measure students' knowledge over time beginning with the successful completion of second-semester general chemistry. From the population of students who completed general chemistry, 120 students were recruited to participate in the study, 90 from the first research setting and 30 from the second setting. The students were given the equilibrium survey online the week following completion of general chemistry (time point 1), approximately six months later (time point 2), and approximately twelve months after the completion of general chemistry (time point 3) (Fig. 1). Students had approximately two weeks to complete the survey at each implementation.

Only responses from 95 students who completed all three surveys were considered for the analysis. Responses were further screened with respect to response duration. The distribution of response durations and graphs of survey scores plotted against response duration were inspected. A decision was made to exclude any survey response that was completed in less than 270 seconds out of concern that the survey was not given serious consideration. This led to complete data from 82 students, with 62 from the first research setting and 20 from the second. To determine the impact of the screening decision on the findings, the same analyses were conducted on the 95 students. The results of this analysis are detailed in Appendix Part 4 with no substantive differences in the findings. Students' pedagogy in second-semester general chemistry, PLTL or didactic-based instruction, was determined by enrollment records.

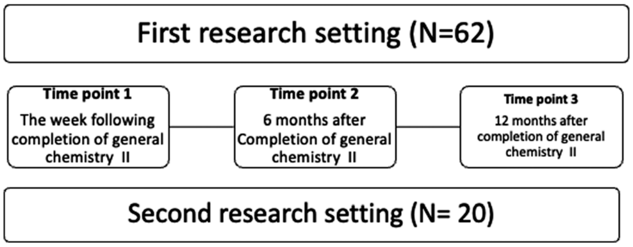


Fig. 1 Timeline of study 1 at both research settings.

Results of study 1

The survey was scored as a percentage of correct responses out of 11 possible items. The survey scores at three time points correlated between 0.72 and 0.74 indicating some consistency in student placement relative to the cohort average at each time point. Descriptive statistics for the survey at each time point are presented in Table 2.

Given the skewness and kurtosis values were mostly within two error terms of zero, the decision was made to treat the data as normally distributed at each time point. The time point 1 and time point 3 kurtosis values were slightly above 1, however this violation to normality is not severe and the statistical tests used in data analysis are robust to violations of normality. To determine the impact of instructional method on students' knowledge, the average survey score for students who had PLTL in second-semester general chemistry was compared to the average score for students with didactic-based instruction at each time point. To describe the difference, an independent sample *t*-test was conducted, and the effect size (Cohen's d) was reported where 0.2 indicates a small effect and 0.5 a medium effect (Table 3).

At each time point, students with PLTL have comparable survey scores to students with didactic instruction. The observed differences of 1 to 2% represent less than 1/10 of a standard deviation and may be a result of unexplained variance inherent in measures of student learning. These results support an interpretation that these groups did not substantively differ at any time point.

Table 2 Descriptive statistics in study 1

Statistic	Time point 1	Time point 2	Time point 3
Mean	0.68	0.62	0.63
Std. deviation	0.23	0.22	0.23
Minimum	0.24	0.18	0.12
Maximum	1.00	1.00	1.00
Skewness (error)	−0.36 (0.27)	0.04 (0.27)	−0.063 (0.27)
Kurtosis (error)	−1.10 (0.53)	−0.96 (0.53)	−1.01 (0.53)

Table 3 Survey score comparison by pedagogy

Time point	Pedagogy	Average	Standard deviation	*t*-Test	Significance	Effect size (Cohen's *d*)
Time point 1	PLTL	0.68	0.25	0.031	0.98	−0.0069
	Didactic	0.69	0.22			
Time point 2	PLTL	0.61	0.26	0.37	0.71	−0.082
	Didactic	0.63	0.19			
Time point 3	PLTL	0.62	0.25	0.44	0.66	−0.098
	Didactic	0.64	0.22			

To determine if the groups changed over time, a general linear model (repeated measures ANOVA) was conducted treating time as a within measure (time point = 1, 2 or 3) and pedagogy as a categorial between measure (PLTL or didactic instruction). A test for sphericity (consistency in variance across time points) showed that sphericity has not been violated ($\chi^2 = 2.13$, df = 2, $p = 0.345$, Greenhouse-Geisser epsilon = 0.97). The results are presented in Table 4 and include a measure of effect size (Partial η^2) where 0.01 is small, 0.06 medium, and 0.14 large. The model indicated a significant impact across time, matching the drop in average observed from time point 1 to time point 2, but no evidence of significant impact from pedagogy overall or pedagogy across time.

Study 2

Methods specific to study 2

Study 2 took place at the first research setting. The equilibrium survey was given to the population of 150 students enrolled in Elementary Analytical Chemistry at two time points. The first administration (time point 1) was during the second week of the semester and had 100 students respond and consent to participate in the study (Fig. 2). The second administration (time point 2) was during the penultimate week of the semester and 72 students from the original 100 completed it. The concept of equilibrium was reviewed in the fourth week of the semester and throughout the rest of the semester applications of equilibrium, in the form of calculations or Le Chatelier's principle, were relied on in most of the analytical methodologies. Students had approximately ten days to complete the survey at each implementation. One student left six of the thirteen items blank in time point 1 and twelve of the thirteen items blank in time point 2 and this student's data was omitted as incomplete. The remaining 71 students had no more than one item blank on the thirteen items and an item blank was treated as an incorrect answer in scoring the survey. The following analyses will rely on the completed surveys by these 71 students.

Table 4 Repeated measures ANOVA results

Variable	*F*-Statistic	Significance	Partial η^2
Within measures comparisons			
Time	7.57	< 0.001	0.086
Time * pedagogy	0.193	0.825	0.0024
Between measures comparisons			
Pedagogy	0.093	0.761	0.0012

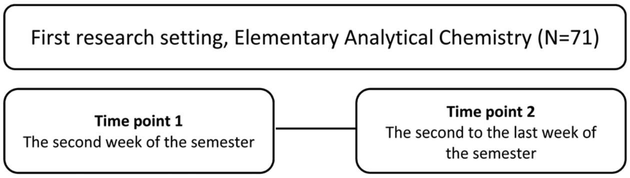


Fig. 2 Timeline of study 2 at research setting 1.

Students in Analytical Chemistry had taken second-semester general chemistry in an earlier semester and may have taken the general chemistry with or without the PLTL pedagogy. Second semester general chemistry pedagogy was determined by using rosters at the home institution from the three years prior to data collection to determine if their second semester general chemistry class used PLTL. The roster data was triangulated by a survey question that asked students about their classroom experiences when taking second semester general chemistry. Of the 71 students, 26 had PLTL, 23 had primarily didactic instruction, and 22 were not identified. Of the 22 students that were not identified, 20 reported in the survey they took second semester general chemistry at another college, one reported having credit for the class from coursework at a secondary institution, and one took the class at the research setting but could not be located in roster files. Since the strong majority of those not identified in the roster took general chemistry at another institution, the general chemistry pedagogy for this group was unknown and this group was not analyzed further. The remaining analysis will focus only on those identified as taking second semester general chemistry with PLTL or didactic instruction.

Descriptive statistics for time point 1 and time point 2 are presented in Table 5. As with Study 1, the normality tests led to a decision to treat the data as normally distributed.

To compare the pedagogical conditions (PLTL and didactic) an independent sample *t*-test was conducted for time points 1 and 2 separately with the results presented in Table 6. At time point 1, students with PLTL in GC2 scored 7% higher than students with didactic in GC2, representing approximately one-third of a standard deviation. The comparison between the two

Table 5 Descriptive statistics in study 2

Statistic	Time point 1	Time point 2
Mean	0.58	0.54
Std. deviation	0.24	0.24
Minimum	0.09	0.09
Maximum	1.00	1.00
Skewness (error)	−0.08 (0.34)	0.17 (0.34)
Kurtosis (error)	−1.03 (0.67)	−0.73 (0.67)

Table 6 Independent sample *t*-test at each time point

Survey	Group	Average	Standard deviation	*t*-Test	Significance	Effect size (Cohen's *d*)
Time point 1	PLTL	0.61	0.24	1.01	0.316	0.29
	Didactic	0.54	0.25			
Time point 2	PLTL	0.58	0.21	1.15	0.256	0.33
	Didactic	0.50	0.26			

pedagogies found no evidence of a statistically significant difference and the effect size of this comparison showed a small effect at both time points 1 and 2.

As with Study 1, to determine if the pedagogy influenced changes over time, a general linear model (repeated measures ANOVA) was conducted that treated time as a within measure and pedagogy as a categorial measure (PLTL or didactic). The outcome of the test, presented in Table 7, indicated no evidence of a statistically significant effect for the time comparison, the GC2 pedagogy, or the interaction between time and pedagogy. The small effect size for pedagogy favors PLTL as the average across time for this group at 59% is higher than didactic at 52% but this difference could be attributed to chance. The small effect size for time describes the overall decrease from the time point 1 (58%) to time point 2 (54%) as shown in Table 5.

Discussion

Prior to interpreting specific results, it is helpful to note the commonality of the results between studies 1 and 2. At the first time point, Study 1 had Didactic higher by 1% while in Study 2 PLTL was higher by 7%. The differences in performance by pedagogy at the first time point results from the sampling technique which did not generate a representative sample of the GC2 population. As a result, the first time point should not be interpreted as an indication of the efficacy of PLTL in GC2. Instead, the first time point serves as a baseline to investigate the change in scores over time. In both studies, a decline in scores from the first to the second time point was observed. The relative average scores of the pedagogies were also consistent across the time points; in Study 1 Didactic was higher by 1% to 2% at each time point, in Study 2 PLTL was higher by 7% to 8% at each time point. The consistency in differences over time was responsible for the near zero effect sizes observed for the interaction of time by pedagogy in Tables 4 and 7. Seeing these similar patterns in the data across both studies lends greater confidence to these findings.

The decline in average score over time suggests that students' conceptions of equilibrium that are built from their general chemistry experiences do not improve following general chemistry. In Study 1, participants were recruited based on completing general chemistry and as a result had a variety of experiences in subsequent chemistry coursework between time points 1 and 3. In Study 2, all participants were enrolled in Analytical Chemistry, an upper-level course that explicitly covers equilibrium concepts and the utilization of these concepts. Despite these differences in experiences following general chemistry, both observed a decline over time. This finding matches a related prior study that qualitatively analyzed general chemistry students' conceptions of ionic and covalent bonding following general chemistry and found their conceptions relatively stable across time (Bowe *et al.*, 2022). Additionally, other studies have shown students in upper-level chemistry courses encountering challenges with general chemistry concepts (Xu *et al.*, 2017; Wang and Lewis, 2020). Combined, the current results and past studies call for careful reconsideration for how students encounter general chemistry concepts across the undergraduate curriculum, to support more meaningful and longer-lasting learning.

Table 7 Repeated measures ANOVA results

Variable	*F*-Statistic	Significance	Partial η^2
Within measures comparisons			
Time	1.83	0.183	0.037
Time * pedagogy	0.019	0.890	<0.001
Between measures comparisons			
Pedagogy	1.39	0.245	0.029

The central research question of this work was to investigate whether the PLTL pedagogy impacted students' knowledge of chemical equilibrium over time. To address this question, the interactions between time and pedagogy in Tables 4 and 7 are evaluated. In each study, the interaction effects were non-significant with near-zero effect sizes. The near-zero effect sizes describes that the pedagogical approach (PLTL or didactic) did not influence the change over time within each sample. Thus, the claim that students' conceptions of equilibrium do not improve following general chemistry would be advanced irrespective of the PLTL or didactic pedagogy used in general chemistry. Framed differently, for skeptical faculty that may argue PLTL within introductory courses does not adequately prepare students for upper-level courses, no evidence was found in support of this argument. It is important to note that this evaluation does not speak toward the efficacy of PLTL in promoting student success *within* the target course. It is possible that PLTL supports a greater proportion of students in developing a conceptual understanding of equilibrium, but following the course students' changes in their conceptions were not impacted by the pedagogy.

The results over time can also be placed in the context of other longitudinal studies conducted in chemistry education. As reviewed earlier, Underwood and colleagues (2016) conducted the only identified chemistry education investigation using a between-subjects and within-subjects research design. They found that students with the CLUE curriculum had higher scores on connecting chemical structure to chemical properties compared to students with the traditional curriculum. An explanation for the contrast of Underwood and colleagues' work with the work presented herein may be the nature of the independent variable investigated. The CLUE curriculum was designed to reconceptualize the nature of the chemistry content introduced, emphasizing structure to property relationships (Underwood *et al.*, 2023). In contrast, PLTL is flexible to the nature of the content presented (Gosser, 2015). In this study, the equilibrium problem sets that were provided in the

PLTL sessions were analogous to the problem sets provided to students in the didactic instruction. Thus, it may be that to make a difference in students' understanding of chemistry content over time, it is required to restructure the presentation and focus of that content within the curriculum. Restructuring the curriculum may better support students through distributed practice than changing pedagogy. Distributed practice describes spreading learning across a longer time, which cognitive science has shown to improve retention (Dunlosky *et al.*, 2013). The flexibility of PLTL is a strength in that it can work with any content, and the literature supports the concept that it improves overall student success within the target class (Frey and Lewis, 2023), but PLTL may not change students' long-term conceptual understanding without a reconceptualization of the nature of the chemistry content introduced.

Longitudinal studies are important as they inform the extent that students carry forward knowledge beyond the class where the knowledge was introduced. Curriculum design rests on a presumption that students who succeed in one course will remain proficient with the knowledge base from that course. Longitudinal studies, therefore, inform the plausibility of this presumption. The current study found that students' proficiencies with equilibrium after general chemistry was characterized by a wide range in performance (standard deviations of approximately 25%) and average scores below 70%. Thus, for a substantial portion of the sample, fluency with chemical equilibrium should not be presumed. Further, it is also worth noting that performance on the survey decreased from time point 1 to 2 in both studies, indicating that curriculum experiences subsequent to general chemistry were not effective at further developing knowledge with this topic. This point is particularly salient in study 2 for students enrolled in Analytical Chemistry, a course which relies heavily on chemical equilibrium principles.

An instructional implication from this work applies to the teaching of upper-level coursework that relies on student understanding of chemical equilibrium, such as Analytical Chemistry and courses on chemical instrumentation. An assumption of proficiency with equilibrium following completion of general chemistry is not supported. One suggestion to try is implementing active pedagogy, including PLTL, in advanced courses to build and refine knowledge of this challenging topic. A second suggestion for how to proceed is presented in Xu and colleagues' (Xu *et al.*, 2017) work. They created a formative quiz to provide students feedback on key concepts in a biochemistry course that were covered in prior classes (*e.g.*, hydrogen bonding). Instructors gave the quiz using a classroom response system at the beginning and end of the semester in biochemistry classes and used the response selection to identify the concepts students held. The responses provided guidance for iterative improvement in instruction that integrated key concepts that were covered in prior classes with the biochemistry topics. A similar structure may prove fruitful regarding students' understanding of chemical equilibrium within advanced courses that rely on this understanding.

Limitations

The finding of the pedagogy having no impact on performance over time invites scrutiny on two characteristics of the study: fidelity of implementation of PLTL, and the nature of the assessment instrument used. Fidelity of implementation describes how closely the PLTL enacted herein adheres to the intention of those who designed PLTL or the current enactment of PLTL. An effort was made to detail the general enactment of PLTL in both settings as they align with the critical criteria of PLTL. More specific to the topic equilibrium, example problems on this topic that students worked with in PLTL are presented in part 1 of the appendix. It is also worth noting that the enactment of PLTL at each setting was informed by the authors past experiences with using PLTL. Each author has prior experience implementing and evaluating PLTL at each research setting (Robert *et al.* 2016, Chan and Bauer 2015). PLTL is a complex pedagogical practice taking place in a naturalistic setting which makes it challenging to summarily judge the implementation, however it is argued here that the implementation of PLTL is in line with how PLTL is implemented within the research literature.

Second, the results presented are dependent on the nature of the assessment instrument used. In this study, the assessment was purposefully designed to measure students' understanding of foundational concepts related to chemical equilibrium as described in Table 1. Past literature has demonstrated a stronger impact of pedagogical reform on single-topic assessments than cumulative assessments (Rahman and Lewis, 2020), so a single-topic was selected here to maximize the opportunity for the pedagogy to distinguish performance over time. Further, the focus of the assessment was on foundational concepts related to defining equilibrium, interpreting the equilibrium constant K, and Le Chatelier's principle. These concepts were chosen as they were considered essential to meaning making of applied equilibrium problems such as the solving for concentrations at equilibrium. Owing to the nature of the assessment, it is important to understand the scope of this work as directed toward answering the research question and not as a comprehensive evaluation of the impact of PLTL. PLTL can have a multitude of impacts on student experiences including students' self-efficacy, communication skills, cooperation, task management, problem solving, and understanding of other topics throughout the course, which can lead to potential long-term impacts on academic persistence and accomplishments. The current study is not designed to investigate these possibilities and does not make claims related to them.

Conclusions

This study found that differences in pedagogy in general chemistry, PLTL compared to didactic instruction, did not influence long-term student retention of chemical equilibrium. It is important to note that this evaluation does not speak toward the efficacy of PLTL in promoting student success *within* the target course. It is possible that PLTL supports a greater proportion of students in developing a conceptual understanding of equilibrium, but after the course is complete there is no residual effect. Improving long-term retention of equilibrium ideas may require more changes to pedagogy (*e.g.* to the lecture or curriculum structure) than the implementation of PLTL described herein. Further, instructors of subsequent courses in which student understanding of chemical equilibrium is important should be prepared to adjust their approaches and expectations to match students' incoming knowledge.

Data availability

The data are not publicly available as approval for this study did not include permission for sharing data publicly.

Conflicts of interest

SEL receives funding from the Royal Society of Chemistry (RSC). The RSC played no role in the data collection, data analysis or manuscript preparation in this work.

Appendices

Part 1: example problems on chemical equilibrium used during peer-led sessions

The following are excerpts of example problems that have been modified for clarity as they were embedded within larger problem sets at research setting 1. Research setting 2 used similar problems that were modified from published textbook questions and problems chosen from Moog and Farrell (2017).

Learning objective 1: define dynamic equilibrium and make predictions regarding relative concentrations and rates of change for a system at dynamic equilibrium.

When a chemical reaction $A \rightleftarrows 2B$ is **at equilibrium**, evaluate each statement as true or false. For statements you indicate it is false, describe how the statement could be rephrased to be true.

(a) The rate of $A \rightarrow 2B$ is equal to the rate of $2B \rightarrow A$
(b) The concentration of A decreases over time
(c) The concentration of B stays constant over time
(d) The concentration of A equals the amount of B
(e) The reaction has stopped

Learning objective 2: describe conceptually the significance of very small or very large values for K

Write a mathematical expression for K for the following reaction:

$$N_2O_5(g) \rightleftharpoons N_2O_3(g) + O_2(g) \quad K_c = 7.75 \times 10^{-4}$$

At equilibrium in the above reaction, are products or reactants favored?

Learning objective 3: given a chemical reaction and concentrations at equilibrium solve for the value of K.

Given the chemical reaction $2CH_3OH(g) \rightleftharpoons 2CO(g) + 4H_2(g)$ where $K_P = 1.95 \times 10^{-9}$ if there is 0.100 atm of CH_3OH and 0.250 atm of CO **at equilibrium**, what is the concentration of H_2 **at equilibrium** in atm?

Learning objective 4: indicate how modifying a chemical reaction (reversing, multiplying by a constant or combining reactions) impacts the value for K.

Given the reaction:

$$A(aq) + 2B(aq) \rightleftharpoons C(aq) + 3D(aq) \quad K = 7.4 \times 10^{-3}$$

Solve for the K for this reaction:

$$3A(aq) + 6B(aq) \rightleftharpoons 3C(aq) + 9D(aq) \quad K =$$

Solve for the K for this reaction:

$$C(aq) + 3D(aq) \rightleftharpoons A(aq) + 2B(aq) \quad K =$$

Learning objective 5: use LeChatelier's principle to predict the direction the reaction will proceed.

The reaction $2SO_2(g) + O_2(g) \rightleftarrows 2SO_3(g)$ is exothermic. Once this reaction is at equilibrium, predict the direction the reaction would shift when each of the following occurs:

(a) Add $O_2(g)$
(b) Add Xe(g)
(c) Decrease temperature
(d) Increase volume
(e) Increase pressure
(f) Remove $SO_2(g)$
(g) Add Pt(s) as a catalyst
(h) When O_2 is added in part (a) describe what happens to Q. Does your description match your predicted shift in part (a)?
(i) In part (c) describe what happens to K and does it match your predicted shift?

Part 2: equilibrium survey in study I and II

1.A. Think about a chemical reaction at equilibrium. Do you expect there to be more products, more reactants, or an equal mixture of both?

- More products
- More reactants
- An equal mixture of both
- There is not enough information to tell

1.B. *Displayed only if students answered "There is not enough information to tell"

What information would tell you if there were more products, more reactants, or an equal mixture of both at equilibrium?

- The value of K.
- The value of Q.
- The enthalpy change for the reaction.

2. NO_2 is a dark red gas and N_2O_4 is a colorless gas. If NO_2 is placed in a reaction chamber the following reaction takes place. $2NO_2(g) \rightarrow N_2O_4(g)$ Which observation would provide evidence that this reaction has reached equilibrium if K is unknown?

- The gas becomes colorless.
- The gas becomes dark red.
- The gas becomes light red.
- The color of the gas stops changing.
- The gas is half dark red and half colorless.

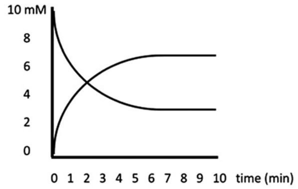
3. Identify which one of the following statements is true:

- All reactions will eventually reach equilibrium.
- Only reactions that increase entropy will reach equilibrium.
- Only reactions that are exothermic will reach equilibrium.
- Only reactions that are spontaneous will reach equilibrium.

- Only reactions in aqueous solution will reach equilibrium.

4. When a chemical reaction is at equilibrium, which of the following is true?
   - Equilibrium represents the highest energy.
   - The forward reaction rate equals the reverse reaction rate.
   - Equilibrium represents the lowest entropy.
   - The mixture contains reactants and products at equal concentrations.

5. The graph shows the concentrations of a reactant and product for a reaction over time. At what time has the reaction first reached equilibrium?



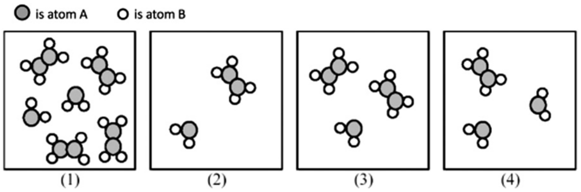
   - 0 minutes
   - 2 minutes
   - 6 minutes
   - 10 minutes
   - The reaction has not reached equilibrium during this time period.

6. What can you conclude if the reaction $O_2(g) + N_2(g) \rightleftarrows 2NO(g)$ has a small K value like 0.0000001?
   - The reaction is slow.
   - The reaction is endothermic.
   - The reaction has just started.
   - The reaction has reached equilibrium.
   - The reaction has more reactants at equilibrium.

7. This reaction $N_2O_4(g) \rightleftarrows 2NO_2(g)$ reaches equilibrium when $[N_2O_4] = 0.10$ M and $[NO_2] = 0.05$ M.

   What is the equilibrium constant for this reaction?
   - 0.025
   - 0.50
   - 1.0
   - 2.0
   - 40
   - The answer is not shown.

8. The following pictures represent mixtures of A_2B_4 molecules and AB_2 molecules, which interconvert according to the equation: $A_2B_4 \rightleftharpoons 2AB_2$. If mixture (1) is at equilibrium, which of the other mixtures are also at equilibrium? All mixtures are at the same temperature.



   - Mixture (2)
   - Mixture (3)
   - Mixture (4)
   - Mixtures (2), (3), and (4) are at equilibrium.
   - Neither Mixture (2), nor (3), nor (4) is at equilibrium.

9. The chemical reaction $H_2(g) \rightleftarrows 2H(g)$ has a value for K smaller than 1.

   When the reaction is written as $2H(g) \rightleftarrows H_2(g)$ what would be true for this rewritten reaction?
   - The value for K is larger than 1 and the reaction favors H_2
   - The value for K is larger than 1 and the reaction favors H
   - The value for K is smaller than 1 and the reaction favors H_2
   - The value for K is smaller than 1 and the reaction favors H

10. A reaction $B_2H_6\ (g) \rightleftarrows 2B(s) + 3H_2(g)$ is exothermic and at equilibrium. What direction would the reaction shift if heat were added?
    - The reaction will shift to the left.
    - The reaction will shift to the right.
    - The reaction will not shift at all.
    - The reaction shifts to the side with more gas molecules.

11. What does it mean to say that a chemical reaction in solution has shifted to the left?
    - The concentration of reactants **increases** while the concentration of products **remains constant**.
    - The concentration of reactants **increases** while the concentration of products **decreases**.
    - At equilibrium, there are more reactants than products.
    - The concentration of reactants **increases** while the concentration of products **increases**.
    - The concentration of reactants **decreases** while the concentration of products **remains constant**.

12. A chemical reaction $B_2H_6(g) \rightleftarrows 2B(s) + 3H_2(g)$ is at equilibrium in a container. Next, additional $B_2H_6(g)$ is added to the container, increasing its concentration. **After** the addition of $B_2H_6(g)$ describe what happens to the concentrations of all three chemical substances.

	Increases	Decreases	Does not change
$B_2H_6(g)$			
$B(s)$			
$H_2(g)$			

13. Given the chemical reaction: $2S_2O_3(g) + O_2(g) \rightleftarrows 4SO_2(g)$ that initially has 0.100 atm of S_2O_3, 0.200 atm of SO_2 and no O_2 present. What would be the expression for the final amount of SO_2 in the equilibrium table below, if the equilibrium quantity of O_2 is X?

	S_2O_3	O_2	SO_2
Initial	0.100	0	0.200
Change			
Final			?

   - $0.200 - 4x$
   - $0.200 - x$
   - $0.200 + 4x$
   - $4x$
   - x

Table 8 Item statistics

	Study 1				Study 2		
		Time point 1	Time point 2	Time point 3		Time point 1	Time point 2
Item	*N*	Proportion correct	Proportion correct	Proportion correct	*N*	Proportion correct	Proportion correct
Q1	82	0.66	0.63	0.60	49	0.43	0.41
Q2	82	0.70	0.68	0.68	49	0.69	0.71
Q3	82	0.90	0.91	0.91	49	0.94	0.78
Q4	82	0.72	0.72	0.67	49	0.63	0.57
Q5	82	0.70	0.59	0.63	49	0.63	0.47
Q6	82	0.73	0.48	0.49	49	0.47	0.47
Q7	82	0.30	0.13	0.15	49	0.02	0.16
Q8	82	0.66	0.66	0.57	49	0.47	0.51
Q9	82	0.70	0.50	0.57	49	0.55	0.49
Q10	82	0.72	0.70	0.76	49	0.69	0.76
Q11	82	0.43	0.40	0.45	49	0.32	0.31
Q12	82	0.49	0.43	0.51	49	0.33	0.32
Q13	82	0.58	0.52	0.56	49	0.52	0.47

Table 9 Survey fit statistics

	Time point 1	Time point 2	Time point 3
χ^2	36.7	57.4	52.9
RMSEA	0.00	0.06	0.05
CFI	1.00	0.95	0.97
TLI	1.03	0.94	0.96
WRMR	0.59	0.79	0.76
Factor loading range	0.29–0.80	0.34–0.83	0.37–0.80
Cronbach's α	0.76	0.75	0.75
McDonald's ω	0.76	0.75	0.74

Table 10 Descriptive statistics in study 1

Statistic	Time point 1	Time point 2	Time point 3
Mean	0.65	0.57	0.59
Std. deviation	0.24	0.25	0.25
Minimum	0.14	0	0
Maximum	1.00	1.00	1.00
Skewness (error)	−0.24 (0.25)	−0.04 (0.25)	−0.02 (0.25)
Kurtosis (error)	−1.14 (0.49)	−0.80 (0.49)	−0.85 (0.549)

Part 3: survey item and fit statistics

MPlus version 7 was used to run all CFA models with a WLSMV estimator since data was not continuous. SPSS was used to calculate Cronbach's α and McDonald's *ω*. These CFA models were only run on Study 1, and not Study 2 since the sample size was too small to have enough power to run a model. Therefore, CFAs were only run for Study 1 at all the three time points with the sample of 95 participants. The model fit statistics criterion for models with dichotomous variables (Hu and Bentler, 1999; Yu, 2002) were used as benchmarks for these model fit statistics. For factor loadings, there is no particular cutoff, however a threshold of 0.32 or greater was chosen to indicate that an item loaded into a particular factor (Brown, 2015). Three separate 13-item unidimensional CFA models were run for Study 1-time points 1, 2, and 3. All fit statistics and Cronbach's alpha values indicated excellent fit. However, since item loadings had a cutoff of 0.32 or greater, this meant there were two items (7 and 11) with low factor loadings for at least two of the time points. These items were removed since they were not contributing to measuring the latent construct of students' knowledge of equilibrium. With the new 11-item unidimensional model, the chi-square test of model fit for all baseline models were statistically significantly at a cutoff value of <0.01, indicating excellent fit. The CFI were all above 0.95 indicating excellent fit. At all time points, the RMSEA values are at or below 0.06 indicating excellent fit. Further, all WRMR values were less than 1 indicating excellent fit. One factor loading in time point 1 was 0.29 which could be considered an unacceptably low factor loading. However, this was the only unacceptably low factor loading. This item was not removed since this is only an issue on one time point, and within that study all other fit indices are excellent. Ultimately, all three factor models were deemed tenable (Tables 8 and 9).

Table 11 Survey score comparison by pedagogy

Time point	Pedagogy	Average	Standard deviation	*t*-Test	Significance	Effect size (Cohen's *d*)
Time point 1	PLTL	0.65	0.26	0.014	0.99	−0.003
	Didactic	0.65	0.23			
Time point 2	PLTL	0.56	0.28	0.58	0.57	−0.12
	Didactic	0.59	0.22			
Time point 3	PLTL	0.58	0.25	0.24	0.81	−0.050
	Didactic	0.59	0.25			

Sphericity test: χ^2 = 2.56, df = 2, p = 0.278, (Greenhouse-Geisser epsilon = 0.97).

Table 12 Repeated measures ANOVA results

Variable	*F*-Statistic	Significance	Partial η^2
Within measures comparisons			
Time	11.48	<0.001	0.110
Time * pedagogy	0.337	0.714	0.004
Between measures comparisons			
Pedagogy	0.093	0.761	<0.001

Part 4: results in study 1 from full sample

The above tables present the same analyses in Study 1 for the entire sample of 95 students, including those who completed the survey in under 270 seconds (Tables 10–12).

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1712230 and 1712164. We would also like to thank the students at both institutions who participated for their time.

References

Apugliese A. and Lewis S. E., (2017), Impact of instructional decisions on the effectiveness of cooperative learning in chemistry through meta-analysis, *Chem. Educ. Res. Pract.*, **18**, 271–278.

Azizoğlu N., Alkan M. and Geban Ö., (2006), Undergraduate pre-service teachers' understandings and misconceptions of phase equilibrium, *J. Chem. Educ.*, **83**, 947–953.

Bain K., Moon A., Mack M. R. and Towns M. H., (2014), A review of research on the teaching and learning of thermodynamics at the university level, *Chem. Educ. Res. Pract.*, **15**, 320–335.

Bodé N. E., Caron J. and Flynn A. B., (2016), Evaluating students' learning gains and experiences from using nomenclature, *Chem. Educ. Res. Pract.*, **17**, 1156–1173.

Bowe K. A., Bauer C. F., Wang Y. and Lewis S. E., (2022), When all you have is a covalent model of bonding, every substance is a molecule: a longitudinal study of student enactment of covalent and ionic bonding models, *J. Chem. Educ.*, **99**, 2808–2820.

Brown T. A., (2015), *Confirmatory factor analysis for applied research*, New York, NY: Guilford publications.

Çalık M. and Cobern W. W., (2017), A cross-cultural study of CKCM efficacy in an undergraduate chemistry classroom, *Chem. Educ. Res. Pract.*, **18**, 691–709.

Cartrette D. P. and Mayo P. M., (2011), Students' understanding of acids/bases in organic chemistry contexts, *Chem. Educ. Res. Pract.*, **12**, 29–39.

Chan J. Y. and Bauer C. F., (2015), Effect of peer-led team learning (PLTL) on student achievement, attitude, and self-concept in college general chemistry in randomized and quasi experimental designs, *J. Res. Sci. Teach.*, **52**(3), 319–346.

Chase A. M., Clancy H. A., Lachance R. P., Mathison B. M., Chiu M. M. and Weaver G. C., (2017), Improving critical thinking via authenticity: the CASPiE research experience in a military academy chemistry course, *Chem. Educ. Res. Pract.*, **18**, 55–63.

Chiu M. H., Chou C. C. and Liu C. J., (2002), Dynamic processes of conceptual change: Analysis of constructing mental models of chemical equilibrium, *J. Res. Sci. Teach.*, **39**, 688–712.

Coştu B., Ayas A. and Niaz M., (2010), Promoting conceptual change in first year students' understanding of evaporation, *Chem. Educ. Res. Pract.*, **11**, 5–16.

Demircioğlu G., Demircioğlu H. and Yadigaroğlu M., (2013), An investigation of chemistry student teachers' understanding of chemical equilibrium, *Int. J. New Trends Educ. Their Implications*, **4**, 192–199.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J. and Willingham, D. T., (2013), Improving Students' Learning With Effective Learning Techniques: Promising Directions from Cognitive and Educational Psychology, *Psychol. Sci. Public Interest*, **14**, 4–58.

Engelbrecht J., Harding A. and Du Preez J., (2007), Long-term retention of basic mathematical knowledge and skills with engineering students, *Eur. J. Eng. Educ.*, **32**, 735–744.

Förster M., Maur A., Weiser C. and Winkel K., (2022), Pre-class video watching fosters achievement and knowledge retention in a flipped classroom, *Comput. Educ.*, **179**, 104399.

Freeman S., Eddy S. L., McDonough M., Smith M. K., Okoroafor N., Jordt H. and Wenderoth M. P., (2014), Active learning increases student performance in science, engineering, and mathematics, *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 8410–8415.

Frey R. F. and Lewis S. E., (2023), Implementing Cooperative Learning in Large Undergraduate Classes via PLTL, in *In Their Own Words: What Scholars and Teachers Want You to Know About Why and How to Apply the Science of Learning in Your Academic Setting*, C. E. Overson, C. M. Hakala, L. L. Kordonowy and V. A. Benassi (ed.), Washington, DC: American Psychological Association, pp. 239–251.

Ganaras K., Dumon A. and Larcher C., (2008), Conceptual integration of chemical equilibrium by prospective physical sciences teachers, *Chem. Educ. Res. Pract.*, **9**, 240–249.

Gosser D. K., (2015), *Peer-Led Team Learning: Origins, Research and Practice*, Ronkonkoma, NY: Linus Learning.

Greenbowe T. J., Rudd J. A. and Hand B. M., (2007), Using the science writing heuristic to improve students' understanding of general equilibrium, *J. Chem. Educ.*, **84**, 1371–1379.

Holme T., Luxford C. and Murphy K., (2015), Updating the general chemistry anchoring concepts content map, *J. Chem. Educ.*, **92**, 1115–1116.

Hu L. and Bentler P. M., (1999), Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives, *Struct. Equation Modeling*, **6**, 1–55.

Johnstone A. H., (2000a), Chemical education research: Where from Here? *Univ. Chem. Educ.*, **4**, 34–38.

Johnstone A. H., (2000b), Teaching of chemistry-logical or psychological? *Chem. Educ. Res. Pract.*, **1**, 9–15.

Kousathana M. and Tsaparlis G., (2002), Students' errors in solving numerical chemical-equilibrium problems, *Chem. Educ. Res. Pract.*, **3**, 5–17.

Leontyev A., Chase A., Pulos S. and Varma-Nelson P., (2017), in *Computer-Aided Data Analysis in Chemical Education Research (CADACER): Advances and Avenues*, ed. T. Gupta, Washington, DC: American Chemical Society, pp. 117–132.

Lewis S. E., (2011), Retention and reform: An evaluation of peer-led team learning, *J. Chem. Educ.*, **88**, 703–707.

Lewis S. E., (2014), Investigating the longitudinal impact of a successful reform in general chemistry on student enrollment and academic performance, *J. Chem. Educ.*, **91**, 2037–2044.

Mitchell Y. D., Ippolito J. and Lewis S. E., (2012), Evaluating Peer-Led Team Learning across the two semester General Chemistry sequence, *Chem. Educ. Res. Pract.* **13**, 378–383.

Moog R. S. and Farrell J. J., (2017), *Chemistry: A Guided Inquiry*, Hoboken, NJ: John Wiley and Sons.

Narli S., (2011), Is constructivist learning environment really effective on learning and long-term knowledge retention in mathematics? Example of the infinity concept, *Educ. Res. Rev.*, **6**, 36–49.

Nyachwaya J. M., (2016), General chemistry students' conceptual understanding and language fluency: Acid–base neutralization and conductometry, *Chem. Educ. Res. Pract.*, **17**, 509–522.

Özmen H., (2008), Determination of students' alternative conceptions about chemical equilibrium: a review of research and the case of Turkey, *Chem. Educ. Res. Pract.*, **9**, 225–233.

Pedrosa M. A. and Dias M. H., (2000), Chemistry textbook approaches to chemical equilibrium and student alternative conceptions, *Chem. Educ. Res. Pract.*, **1**, 227–236.

Quílez J., (2009), From chemical forces to chemical rates: a historical/philosophical foundation for the teaching of chemical equilibrium, *Sci. Educ.*, **18**, 1203–1251.

Rahman M. T. and Lewis S. E., (2020), Evaluating the evidence base for evidence-based instructional practices in chemistry through meta-analysis, *J. Res. Sci. Teach.*, **57**, 765–793.

Robert J., Lewis S. E., Oueini R. and Mapugay A., (2016), Coordinated Implementation and Evaluation of Flipped Classes and Peer-Led Team Learning in General Chemistry, *J. Chem. Educ.*, **93**, 1993–1998.

Shah L., Rodriguez C. A., Bartoli M. and Rushton G. T., (2018), Analysing the impact of a discussion-oriented curriculum on first-year general chemistry students' conceptions of relative acidity, *Chem. Educ. Res. Pract.*, **19**, 543–557.

Slavin R. E., (1996), Research on cooperative learning and achievement: What we know, what we need to know, *Contemp. Educ. Psychol.*, **21**, 43–69.

Talanquer V., (2011), Macro, submicro, and symbolic: The many faces of the chemistry ''triplet'', *Int. J. Sci. Educ.*, **33**, 179–195.

Thomas P. L. and Schwenz R. W., (1998), College physical chemistry students' conceptions of equilibrium and fundamental thermodynamics, *J. Res. Sci. Teach.*, **35**, 1151–1160.

Turányi T. and Tóth Z., (2013), Hungarian university students' misunderstandings in thermodynamics and chemical kinetics, *Chem. Educ. Res. Pract.*, **14**, 105–116.

Underwood S. M., Kararo A. T., Posey L. A., Pollock A. M., Herrington D. G., Stowe R. L., Carmel J. H., Klymkowsky M. W. and Cooper M. M., (2023), Components Critical to Successful Adoption and Adaptation of CLUE, a Transformed General Chemistry Curriculum, *J. Chem. Educ.*, **100**, 3374–3385.

Underwood S. M., Reyes-Gastelum D. and Cooper M. M., (2016), When do students recognize relationships between molecular structure and properties? A longitudinal comparison of the impact of traditional and transformed curricula, *Chem. Educ. Res. Pract.*, **17**, 365–380.

Varma-Nelson P., Cracolice M. S. and Gosser D. K., (2004), in *Invention and impact: Building in undergraduate science, technology, engineering, and mathematics education*, Cunningham S. and George Y. (ed.), Washington, DC: American Association for the Advancement of Science, pp. 43–48.

Wang Q., (2021), The cultural foundation of human memory, *Annu. Rev. Psychol.*, **72**, 151–179.

Wang Y. and Lewis S. E., (2020), Analytical chemistry students' explanatory statements in the context of their corresponding lecture, *Chem. Educ. Res. Pract.*, **21**, 1183–1198.

Wolfson A. J., Rowland S. L., Lawrie G. A. and Wright A. H., (2014), Student conceptions about energy transformations: Progression from general chemistry to biochemistry, *Chem. Educ. Res. Pract.*, **15**, 168–183.

Xu X., Lewis J. E., Loertscher J., Minderhout V. and Tienson H. L., (2017), Small Changes: Using Assessment to Direct Instructional Practices in Large-Enrollment Biochemistry Courses, *CBE-Life Sci. Educ.*, **16**, 1–13.

Young J. D. and Lewis S. E., (2022), Evaluating peer-led team learning integrated into online instruction in promoting general chemistry student success, *J. Chem. Educ.*, **99**, 1392–1399.

Yu C. Y., (2002), *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*, Doctoral dissertation, University of California.