Research Article

Effect of Peer-Led Team Learning (PLTL) on Student Achievement, Attitude, and Self-Concept in College General Chemistry in Randomized and Quasi Experimental Designs

Julia Y. K. Chan and Christopher F. Bauer

University of New Hampshire, Durham, New Hampshire

Received 4 March 2014; Accepted 17 November 2014

Abstract: This study investigated exam achievement and affective characteristics of students in general chemistry in a fully-randomized experimental design, contrasting Peer-Led Team Learning (PLTL) participation with a control group balanced for time-on-task and study activity. This study population included two independent first-semester courses with enrollments of about 600. Achievement was measured by scores on exams written by an instructor blind to student participation. Established instruments were used to assess changes in attitude to chemistry and self-concept as a chemistry learner. No differences were found in achievement, attitude, or self-concept for students who participated in PLTL vs. those who participated in documented alternative study activities. Overall, certain aspects of attitude and self-concept showed a slight but significant decline from beginning to end of semester, consistent with previous studies. Males have higher positive attitude and self-concept than females, and first-year students have higher positive attitude, self-concept, and achievement than non first-year students. In a quasi-experimental comparison of 10 other course sections over seven years, students who self-selected into PLTL showed stronger exam achievement than those who did not choose to participate. These findings suggest that past reports of improved student performance with PLTL may in part be a consequence of attracting students who are already motivated to take advantage of its value.

Keywords: PLTL; attitude; self-concept; cooperative learning; random assignment; achievement; college or postsecondary students; chemistry

Numerous studies have shown that traditional methods for teaching chemistry, such as instructor-centered lectures, verification labs, and teaching assistant-led recitations, are not effective for a majority of students (Freeman et al., 2013; Hake, 1998; McKeachie, Pintrich, Lin, & Smith, 1987; Smith, Sheppard, Johnson, & Johnson, 2005; Wood, 2009). These passive learning approaches provide students with minimal opportunity to develop critical and metacognitive thinking skills. In contrast, active processes where students create and construct their own knowledge support development of higher level reasoning and conceptual understanding (Varma-Nelson & Coppola, 2005). Student-centered group learning pedagogies can improve students’ academic achievement and lead to more positive attitudes and self-concepts about their educational experience and the subject area (Johnson, Johnson, & Smith, 1998; Stevens & Slavin, 1995). The characteristics of three student-centered pedagogies are compared and contrasted in...
depth by Eberlein et al.: problem-based learning (PBL), process-oriented guided inquiry learning (POGIL), and peer-led team learning (PLTL) (Eberlein et al., 2008).

The learning outcome claims concerning PLTL are the focus of this article (both content knowledge and affective characteristics). Since implementing PLTL at his institution in 2000, the corresponding author has been curious about where the PLTL advantage comes from. The record demonstrates that students who participate in PLTL tend to persist in courses and have stronger course outcomes as documented by course grades and self-report end-of-course surveys. All published studies take a quasi-experimental approach to assessment: either by historical comparisons across semesters, or side-by-side comparisons in concurrent (parallel) sections of a course (Gosser & Roth, 1998; Hockings, De Angelis, & Frey, 2008; Kampmeier & Varma-Nelson, 2009; Lewis, 2011; Lewis & Lewis, 2005, 2008; Mitchell, Ippolito, & Lewis, 2012; Shields et al., 2012; Smith, Wilson, Banks, Zhu, & Varma-Nelson, 2014; Tien, Roth, & Kampmeier, 2002; Varma-Nelson & Coppola, 2005; Wamser, 2006). The results of these studies supported the establishment and dissemination of PLTL and identified in a qualitative way several critical components for implementation. There is no question that students who engage in PLTL can benefit. Furthermore, there are strong theoretical arguments, described below, for why PLTL might support learning. These studies, however, do not demonstrate definitively that PLTL participation alone causes the stronger outcomes, nor do they elucidate experimentally a mechanism by which PLTL produces those results. As with chemical reactions, it is possible to know that a particular combination of conditions gives a good product yield without knowing exactly how it happens. It is also possible that a better yield will come from a different feedstock of starting materials. If we can understand more about the mechanisms that play a role in PLTL’s success and more about the characteristics of students who participate, and how these factors interact, we will be in a better position to make decisions that strengthen the benefits of PLTL for more students.

This study investigated both exam achievement and affective characteristics in a fully-randomized experimental design, contrasting PLTL participation with a control group balanced for motivation, time-on-task and study activity. Achievement was measured by exam scores on common exams. Also measured via validated instruments were affective characteristics, including dimensions of self-concept and attitude to chemistry. Additionally, exam achievement for PLTL participants and non-participants was compared for 10 other classes in which students were allowed to self-select into PLTL. This latter condition is more directly comparable to previously reported studies. This allows a comparison between study designs (randomized vs. quasi-experimental) in understanding the effects of PLTL. As we will describe, these controls lead to some surprising results.

Peer-Led Team Learning in Chemistry

PLTL is a structured form of group learning where a small group of six to eight students meets weekly to discuss and solve problems related to topics covered in lectures. These groups are led by a trained peer who earned at least a B in the same course in the previous year. The role of the peer leader is to engage the group in problem-solving activities, assist students in developing conceptual understanding, and to facilitate discussion of scientific concepts and ideas (Gafney & Varma-Nelson, 2008; Gosser et al., 2001). The theoretical basis for the claimed value of PLTL is social constructivism (Eberlein et al., 2008; Tien et al., 2002): Meaning is negotiated and scaffolded with the assistance of a more experienced and successful learner. Peer leaders participate in a parallel course that introduces them to issues in human learning, cognition and group dynamics, and reviews chemistry content and materials for use in the study groups. The PLTL model has been reported in numerous studies in courses such as general chemistry and

Journal of Research in Science Teaching
organic chemistry to reduce high student attrition rates, enhance student motivation and engagement, and increase achievement in chemistry (Gosser & Roth, 1998; Hockings, De Angelis, & Frey, 2008; Kampmeier & Varma-Nelson, 2009; Lewis, 2011; Lewis & Lewis, 2005, 2008; Mitchell et al., 2012; Shields et al., 2012; Smith et al., 2014; Tien et al., 2002; Varma-Nelson & Coppola, 2005; Wamser, 2006).

Affective Domain in Learning

In addition to enhancement of achievement, collaborative small group learning pedagogies claim to have positive impacts on students’ affective experiences, such as their attitudes toward the subject and perception of themselves as learners. For instance, self-concept can shape the way students develop meaningful understanding, process information, and perform in the course (Nieswandt, 2007). Research on the role of affect has been more limited at the post-secondary level in comparison with the elementary and secondary levels (Cheung, 2009; Helmke & van Aken, 1995; Hofstein, Ben-Zvi, Samuel, & Tamir, 1977; Menis, 1983; Salta & Tzougraki, 2004; Stevens & Slavin, 1995). With reference to chemistry, Lewis, Shaw, Webster, & Heitz (2009) found in general chemistry that students with higher self-concept performed better on exams, as was found in past studies (Helmke & van Aken, 1995; Marsh & Yeung, 1997). Self-concept played a role in student performance even after controlling for content related factors, such as SAT scores. Brandriet, Xu, Breit, & Lewis (2011) looked at attitudinal changes in general chemistry and found a positive relationship between attitude and achievement in chemistry, similar to past studies (Bennett, Rollnick, Green, & White, 2001; Freedman, 1997; House, 1995; Salta & Tzougraki, 2004; Weinburgh, 1995). In particular, high performing students had a more positive attitude towards chemistry throughout the semester while low performing students had a more negative attitude throughout the semester.

There has also been continual interest on sex differences in science achievement (Blickenstaff, 2005; Greenfield, 1996; Osborne, Simon, Collins, 2003; Turner & Lindsay, 2003; Willson, Ackerman, & Malave, 2000). In college chemistry, Turner & Lindsay (2003) looked at sex differences in cognitive factors (ACT, spatial-visualization, introductory chemistry scores) and noncognitive factors (confidence, anxiety, motivation, usefulness) related to organic chemistry achievement. When both factors were considered, noncognitive factors played a more important role in predicting course grades for male than for female students. Other studies have also looked at sex differences specific to attitudes in chemistry (Brandriet et al., 2011; Cheung, 2009). For the most part, males report a more positive attitude than females towards chemistry. The current study, therefore, afforded the opportunity to explore how PLTL participation might be related to achievement and affective outcomes differentially for males and females.

Self-Concept

Based on previous research results, we would expect the PLTL environment to be conducive to improving a student’s self-concept. Self-concept refers to an individual’s persistent perceptions and beliefs about themselves in general or specific areas of knowledge (Bauer, 2005; Marsh & Yeung, 1997; Nieswandt, 2007). The statement “I have trouble understanding anything based on chemistry” is a self-concept judgment (Bauer, 2005). It develops through one’s experiences with various learning environments (Bong & Skaalvik, 2003). This is a more general self-assessment than self-efficacy which refers to individuals’ judgment and confidence in their ability to accomplish a specific task (Pintrich & de Groot, 1990). For instance, a students’ expectation that he will be able to master a set of skills in a course is a self-efficacy judgment. In this study, the Chemistry Self-Concept Inventory (CSCI) is used to assess students’ self-concept as a learner of chemistry (Bauer, 2005).
Attitude

Based on previous research, we would expect PLTL to enhance a student’s attitude toward the subject matter of chemistry. An attitude is defined as a learned tendency to respond in a positive or negative manner with respect to a given attitude object (Oskamp & Schultz, 2005). “Improving student attitudes” is often cited as the goal of individual instructors as well as of those involved in curriculum innovation (Dalgety, Coll, & Jones, 2003; Lawrenz, Wood, Kirchhoff, Kim, & Eisenkraft, 2009; Salta & Tzougraki, 2004). Fortus (2014) and Osborne et al. (2003) have provided comprehensive overviews of research on affect in science education. In this study, the Attitude toward the Subject of Chemistry Inventory (ASCI) is used to assess students’ attitudes regarding chemistry as a “body of knowledge or practices” (Bauer, 2008).

Purpose of the Study

Unique to this study is an attempt to control or minimize potentially confounding factors that have not been controlled in previous studies of PLTL student outcomes. This includes randomized experimental design, time-on-task control, common assessment standards, and instructor-blind conditions. The following research questions were pursued in this study:

- To what extent is there a difference in students’ academic achievement (i.e., term exams and final exam scores) between students participating in PLTL and those who do not (using both randomized and quasi-experimental designs)?
- To what extent is there a difference in chemistry self-concept and attitude toward chemistry in students enrolled in PLTL versus those who are not?
- To what extent is self-concept and attitude moderated by student characteristics of sex or class year (first-year students vs non-first-year) over the course of semester?

Research Methodology

Randomized Design and Participants

PLTL has been incorporated continuously in general chemistry since 2001 at the University of New Hampshire, a large, 4-year residential, public research university. In two fall semesters (2008, 2009), this course was taught in three lecture sections each meeting weekly in three 50-minute periods. About two-thirds of the class were first-year students, representing a diverse set of majors in engineering, sciences, health and human services, and liberal arts. The lecture instructor was a single, experienced female person (not the authors) and the curriculum, class content, and all examinations were identical for all sections. Lecture style was a hybrid of teacher-centered lecture interspersed with occasional short-duration student peer discussion and problem solving. Course enrollments for 2008 and 2009 were 620 and 682, respectively.

PLTL was implemented according to the recommendations of the model (Gafney & Varma-Nelson, 2008). Leaders were successful students (B grade or better) from the previous year’s course who applied to participate. After initial orientation sessions, they participated in a 2-credit course that met weekly for 90 minutes through the semester. The course divided time between building their understanding of human cognition and science learning, developing their skills as facilitative instructors, and reviewing chemistry content. Leaders reflected on their practice through weekly writing and visits to observe other leaders with their groups. The PLTL Leader Handbook was used as a text along with supplemental readings (Roth, Goldstein, & Marcus, 2001). Materials used in the study groups included activities from the PLTL library (Gosser, Strozak, & Cracolice, 2006) and materials designed by one of the authors (CB) to facilitate
knowledge construction and to support group engagement. No answer keys were provided, recognized critical characteristic to promote discussion and problem solving. The library of PLTL activities was sufficient to allow leaders flexibility in deciding which to use for their group in a particular week. Since groups met nearly every day of the week, this flexibility helped leaders match content to the lecture schedule. Because of the constraints of the academic schedule, PLTL sessions were 80 minutes in 2008 and 50 minutes in 2009.

The two comparison groups, PLTL participants and non-participants, were determined as illustrated by Figure 1. It was a course requirement that all students participate in an out-of-class learning activity, one of these being PLTL. All students were strongly encouraged to sign up for PLTL as the primary option. However, because more students signed up than the planned capacity, only a fraction of the students who elected to be in PLTL could be included (about 60%). To be fair, students were selected randomly for inclusion. Thus, two conditions were created: in or not-in PLTL. Since both groups shared the common desire to be included, they do not differ in motivation to participate. Consequently, there is no self-selection bias between the PLTL and non-PLTL students.

To determine which students to include in PLTL, selection was accomplished by random sampling from the list of students using the statistics program Minitab. We sampled equally from the three lecture sections (roughly equivalent in size), and proportionally in terms of class year (50% first-year students, 38% sophomores, 9% juniors, 3% seniors). This process intended to balance the demographic characteristics of the two comparison groups.

PLTL sessions were scheduled throughout the week. Students could request times that fit their schedules. Based on those selections, PLTL groups were formed by manual sorting. In order to make each PLTL group as similar as possible (and thus try to make the “treatment” conditions as similar as possible), we attempted to balance for lecture time and demographics (sex, class year, and major). However, the small group size and constraints on student schedules made this difficult to achieve fully. The final composition of PLTL groups had the following broad common characteristics. Median size was 6 students per group with most groups in the range of 4–9 members. Two-thirds of groups had more female than male members, with most of the rest having equal female/male numbers. Two-thirds of groups had more first-year than non-first-year students. Finally, 80% of groups were dominated by students majoring in the biological or health sciences,

---

**Figure 1.** Process by which students were included in the study.

*Journal of Research in Science Teaching*
with most of the remainder in the physical sciences or liberal arts. These distributions are similar to the make-up of the entire population of the course.

Students who were not included in PLTL were provided several other options for alternative learning activities. These included self-organized study groups (2–6 students who agreed to meet regularly, no instructional leader present); instructor-led review sessions (course instructor interacting with 5–30 students in a weekly question/answer session); drop-in tutorials led by graduate or undergraduate chemistry majors (primarily one-on-one assistance); and instructor office hours (individual ad hoc visits by individuals). Students could do one or more of these options to earn their hours. Thus, the key comparison groups are PLTL students and non-PLTL students (doing one or more of the alternate activities). Some students did not respond to the PLTL sign-up instruction in a timely fashion, and were not included in either of the comparison groups (Figure 1).

PLTL sessions were 50 (Fall 2009) to 80 (Fall 2008) minutes long and occurred nearly every week of the semester (12 sessions scheduled). As a means for controlling time-on-task, all non-PLTL participants (those requesting PLTL but not being chosen) were required to document the same number of hours of “alternative learning” activities outside of class time, equivalent to the PLTL time commitment. Of the non-PLTL students, 45% participated in self-organized study groups, 22% in instructor-led review sessions, 23% in graduate or undergraduate student-led tutorials, and 10% in instructor office hours and individual tutoring. Students documented hours on a standard form which had to be signed by the instructional leader for that session. In the case of self-organized groups, one group member reported hours by each exam date. This process does not provide strict rigor on reporting and is likely biased high, but students frequently expressed concern about records being correctly recorded, so we believe they were honoring the spirit of the requirement. Both PLTL participants and non-PLTL students earned credit (10% of their overall grade) for their study hours in proportion to their attendance. In 2008, 86% of the students who participated in PLTL attended 7 or more 80-minutes sessions, while in 2009, attendance was 98% for 50-minute sessions. In both years, most non-PLTL students reported at least 10 hours of extra activity. The number of hours students spent in study outside of these reported times was not recorded.

PLTL was managed by the corresponding author. The course lecturer was not involved in the student selection process and was blind to student participation in PLTL and the alternative activities until grades were assigned. Students might have revealed participation to the instructor through conversations and presence at review session or office hours, but any potential effect of this revelation would be diluted by the large number of students.

Human subject consent was obtained from participants. Of the 600 plus students in the course, about 25% were diverted to another academic support program and are not included in this study. Of the 297 students who consented during fall 2008, 158 students (PLTL, \(N = 107\); non-PLTL, \(N = 51\)) completed both the CSCI pre/post while 166 students (PLTL, \(N = 113\); non-PLTL, \(N = 53\)) responded to both the ASCI pre/post. In fall 2009, of the 150 consenting, 80 students responded (PLTL, \(N = 50\); non-PLTL, \(N = 30\)) to the CSCI pre/post while 91 (PLTL, \(N = 59\); non-PLTL, \(N = 32\)) students responded to the ASCI pre/post.

**Affective Instruments and Exam Achievement**

Two established instruments were used for affective data collection in a pre/post design. These instruments were validated with a student population that is identical to those included in the current study. The Attitude to Subject of Chemistry Inventory (ASCI) was used to assess students’ attitude toward chemistry (Bauer, 2008) and the Chemistry Self-Concept Inventory (CSCI) was used to measure students’ self-concept towards chemistry (Bauer, 2005). Both
surveys were given online to students twice, once at the beginning (first 2 weeks of September) and once at the end of semester (second week of December). The ASCI has a semantic differential format in which students position themselves between two polar adjectives on a seven point Likert scale. It contains five variables: interest and utility, anxiety, intellectual accessibility, emotional satisfaction, and fear. The other instrument, CSCI, is a 40 item Likert style assessment instrument for measuring student’s self-concept toward chemistry. It also consists of five variables: chemistry self-concept, mathematics self-concept, academic self-concept, academic enjoyment self-concept, and creativity self-concept. Students respond to statements that are (1) “very inaccurate of me” to (7) “very accurate of me”. Cronbach’s $\alpha$ values were computed to check for the internal consistency of the instrument subscales at both time points (pre/post) in both years (Supplementary Table S1 and S2). For the ASCI, all subscales had Cronbach’s $\alpha$ values that were consistently above 0.8. For the CSCI, most variables were above 0.8 except for creativity (0.71 (pre); 0.66 (post)) and academic self-concept (0.78 (pre)). These results match or exceed those reported previously.

Student achievement was simply the raw score in points achieved for each exam. Three progress exams were given at one month intervals. These exams tested the chemistry content within that month’s period, were written fresh by the course instructor each year, and consisted of about 70% written answers and 30% multiple choice. About three weeks after the third exam, the final exam was held which consisted entirely of multiple choice questions providing comprehensive coverage of the semester’s topics. Half of the exam included questions from the General Chemistry Special First Term Exam (1997), developed and nationally normed by the American Chemical Society Examinations Institute. The other half consisted of instructor-written questions intended to complement the ACS exam question topics and match the course curriculum at this university. Nearly all of these questions had been tested by repeated use over several years. All final exam problems are secure, unreleased items. Considering all tests together, the mix of thinking required (verbal argument, problem solving) and format (written response, multiple choice) was intended to avoid providing an advantage to either the PLTL or non-PLTL groups.

Quasi-Experimental Achievement Comparisons

As described above, the 2008 and 2009 years used a randomized design to control for self-selection bias. What happens when this control is not exercised? We compared PLTL participants and non-participants in ten sections of general chemistry in Falls of 2004, 2005, 2006, 2007, 2011, 2012, and 2013, all of which were taught by the corresponding author. The course content, expectations, and final exam were closely similar between the two instructors and across these years. Each section enrollment was about 150–200 students. In all of these years, students had the option of joining PLTL. (By joining, students earned attendance credit and lowered the weight of computer-based homework or in-class quizzes). Thus, the years before and after the randomized study provide conditions in which motivation for participation is not controlled.

Over these years, the PLTL leader recruitment, training course, group sizes, and support structures were the same. PLTL group session durations were 80 minutes, except for 2009 and 2011 where the sessions were 50 minutes. Starting in 2011, new leaders had individual veteran leaders as mentors who helped the new leaders initiate their groups and engage in weekly debriefing and planning meetings. Furthermore, in 2007 only (two course sections), students who did not choose to join PLTL were required to engage in an equivalent number of hours of out-of-class study, similar to the requirement for the randomized design in 2008 and 2009. This was a happenstance situation which provides some control of time-on-task.
Results

Nonparticipants and outliers

Students who did not respond to the instruction to sign up for PLTL were excluded from the analysis. Students who did not submit a written record of their alternative learning activities hours were eliminated from the analysis since we had no information regarding how they spent their time. Descriptive statistics and normality tests for skewness and kurtosis on all variables from both ASCI and CSCL were checked and found to be close to normal. For most of the statistical tests reported here, the critical significance level was $p < 0.001$ (Bonferroni correction) to control for Type I error because a large number of statistical comparisons were considered.

Retention

Student persistence was determined by the number of students who dropped or withdrew from the course after the first 2 weeks. Students who dropped before 2 weeks would have had little to no time in PLTL. In each year, 45 students out of 550–600 dropped (less than 10%). Of those who dropped, only 4–6 had signed up for the chance to participate in PLTL.

Exam Achievement in Randomized Design

Descriptive statistics were performed using SPSS 18.0–20.0. (Many statistical tables, given an S label, are included in the supplementary material.) A three-way between-groups repeated measures multivariate analysis of variance (MANOVA) was performed. This was done for each year’s data (2008, 2009) with exam number as the within subject variable and sex, class year, and learning activity as between subject variables. The exam scores are independent of each other and do not have a uniform scale so there is no meaning associated with comparing one exam score to the others. The purpose of the repeated measures was simply to control for the individual student differences in the comparison among the other variables.

In terms of the primary manipulation of study activity, in 2008 there are no contrasts in which PLTL students have distinctly different results from non-PLTL students. Figure 2 shows the mean exam scores (percent) for students disaggregated by class year, sex, and study activity. Supplementary Table S3 lists the disaggregated and pooled means. Class year showed a strong effect size, with first-year students outperforming non-first-year students by 6–12% [$p < 0.001$, $F(4,193) = 10.3$, partial $\eta^2 = 0.18$]. A weaker secondary effect appears in an interaction of year and sex [$p \approx 0.05$, $\eta^2 = 0.021$]; females exhibit a smaller first-year advantage (8%) than do males (15%). The apparent better performance of first-year PLTL females (Figure 2) does not reach a level of significance and the fact that there is no change over the semester’s sequential exams suggests that PLTL is not causing the difference. In 2009 (Figure 3; Supplementary Table S4), all main factors show significant differences: non-PLTL over PLTL by 3%, first-year over non-first-year by 8%, and males over females by 7%. These results are, however, misleading.

Although the learning activity groupings were determined by random assignment, we also checked for group equivalence on mathematical facility by using Math SAT scores as a covariate in the analysis. When Math SAT is included as a covariate, all of the main factors for 2009 become non-significant. For 2008 data, including the SAT Math covariate reduced but did not eliminate the significant effects of class year and its interaction with sex. Math SAT has frequently been shown to account for a substantial amount of variance in general chemistry performance (Andrews & Andrews, 1979; Ozsogomonyan & Loftus, 1979; Lewis & Lewis, 2007; Spencer, 1996; Wagner, Sasser, & DiBiase, 2002). For our populations, $t$-test comparison for the 2009 class indicates that the mean Math SAT scores are somewhat different for males vs females: [$M = 593$, $SD = 192$ (males); $M = 532$, $SD = 150$ (females), $p = 0.08$]. Given that Math SAT often has a strong
relationship with chemistry achievement scores (Lewis & Lewis, 2007), a small unaccounted for difference between the comparison groups on this factor can confound interpretation. In principle, random assignment should have eliminated this potential problem. One might argue that the random selection was not really random, but for the other demographic variables of sex, class year, and major, no significant differences were found between the PLTL and non-PLTL groups.

Exam Achievement in Quasi-Experimental Comparison

Other comparative studies of PLTL have often used course grade as the dependent outcome variable. We chose not to use grades for three reasons. First, the grade incorporates PLTL attendance credit, thus creating the complication of an implicit correlation. Second, there was no systematic control over how grade cutoffs were determined year-to-year. Lastly, a letter grade scale reduces a continuous variable (course points) to an ordinal variable, which sacrifices statistical power. Instead of grades, we used total exam achievement as the dependent variable. This is justifiable since exams represent a major contribution to course grades in this study as well as in other PLTL studies. Furthermore, the exam frequency, content, structure, and expectations were very similar across the comparison years. An exam achievement score was computed for each student by summing the raw scores across the three monthly exams and one final exam.
students who attended more than 50% of the PLTL sessions and other alternative learning activities were included in the analysis. Students who did not complete all four exams were dropped out of the analysis.

Students’ exam achievement scores were compared across ten sections of general chemistry (Fall 2004, 2005, 2006, 2007, 2011, 2012, and 2013). PLTL participation was optional in these years (Figure 4). Overall, PLTL students achieved higher exam scores than non-PLTL students. Significant differences between the two groups were found in the following years: 2006, 2007A, 2007B, 2011A, and 2012 (A and B identify different course sections in the same year.) Supporting statistical detail can be found in the supplementary material (Supplementary Table S5). This result should be contrasted with the 2008 and 2009 finding of “no difference” when there was no possibility of self-selection into PLTL (also shown in Figure 4).

Attitudes to Chemistry

To determine differences in attitudes between the PLTL and non-PLTL group, we conducted a mixed repeated measures analysis of variance (ANOVA) with time (pre and post) as a within-subject factor and participation in the learning activities group (PLTL or non-PLTL) as the between-subject factor. These findings are summarized in Figure 5 visually, with the supporting statistical detail in the supplementary material (Supplementary Table S6). Each variable is represented in a different color; solid error bars (95% confidence interval for mean) indicate

Journal of Research in Science Teaching

Figure 3. Comparison for 2009 exam results across the semester for students in PLTL (green, right side of pair) or not (blue, left side of pair) disaggregated by sex and class year (error bars show 95% confidence interval).
pre-time points and dashed error bars indicate post-time points. Recall that the ASCI response scale is a one-to-seven point Likert scale where a higher value indicates a stronger feeling that chemistry is “interesting and useful,” “anxiety producing,” and “intellectually accessible.” A score of four indicates a neutral attitude towards chemistry. Overall, attitudes changed in the negative direction throughout the semester \((p < 0.001)\). However, no difference was found between PLTL and non-PLTL students and no significant interactions were found between learning activity and time. In terms of specific statistical comparisons \((p < 0.001)\), interest and utility, intellectual accessibility, and emotional satisfaction toward chemistry all decreased significantly throughout the semester \([M = -0.95, SD = 1.63, t(165) = -7.48, d = 0.58 \text{ (interest & utility)}; M = -0.38, SD = 1.37, t(165) = -3.56, d = 0.28 \text{ (intellectual accessibility)}; M = -0.82, SD = 1.52, t(165) = -6.93, d = 0.54 \text{ (emotional satisfaction)})\]. On the other hand, students’ anxiety and fear of chemistry increased slightly yet significantly throughout the semester \([M = 0.52, SD = 1.63, t(165) = 4.87, d = 0.37 \text{ (anxiety)}; M = 0.51, SD = 1.82, t(165) = 3.59, d = 0.28 \text{ (fear)})\]. The scale of Cohen’s \(d\) effect size ranges from 0 to 2 where 0.2 is considered small, 0.5 is medium, and 0.8 is large (Warner, 2008). Thus, the observed effect sizes are small to medium, indicating there is a modest meaningful change in attitude across the semester. Brandriet et al. (2011) reported small effect sizes in attitudinal differences across the semester as well using ASCIv2, a modified version of the ASCI (Xu & Lewis, 2011).

**Self-Concept**

Similarly, to evaluate self-concept change throughout the semester between PLTL and non-PLTL groups, we ran a mixed repeated measures ANOVA. These findings are summarized in

*Figure 4.* Exam achievement between PLTL(left) vs. non-PLTL(right) students for randomized study years* (2008,2009) and quasi-experimental years.
Figure 6, with the supporting statistical detail in the supplementary material (Supplementary Table S7). Like attitudes toward chemistry, students’ self-concept regarding chemistry did not differ between PLTL and non-PLTL groups and there were no significant interactions between group learning activity and time. In only one instance was there a change: students’ academic enjoyment self-concept, though remaining on positive side of scale, decreased significantly throughout the semester \[ M = -0.39, SD = 0.83, t(157) = -5.96, d = 0.46 \]. The same data analysis procedures were carried out for the fall 2009 class. Results were similar to the fall 2008 class. Again, no significant differences were found in terms of self-concept and attitudes toward chemistry between the PLTL and non-PLTL group over the course of the semester.

**Difference in Attitude by Sex**

Figures 7 and 8 show mean scores on the attitudinal variables across time (pre/post) for females and males in 2008 and 2009, respectively. At the beginning of the semester in both years, none of the attitudinal variables were significantly different between females and males. At the end of the fall 2008 semester, fear and intellectual accessibility were significantly different between males and females with medium effect sizes. Females reported they had higher fear while males reported they had higher intellectual accessibility towards chemistry. At the end of fall 2009 semester, emotional satisfaction differed between the males and females \[ M = 4.41, SD = 1.25 \] (males), \[ M = 3.38, SD = 1.35 \] (females), \( t(89) = 3.39, d = 0.78 \). With a sample size half as large
as that of 2008, these effect sizes are on the borderline of medium/large, suggesting considerable meaningful differences in attitudes between males and females.

Differences in Attitudes by Class Year

Figures 9 and 10 shows mean scores on the attitudinal variables across time for first-year and non-first-year in 2008 and 2009, respectively. Overall, non first-year students showed a more negative attitude toward chemistry than first-year students. Throughout the semester non first-year students reported they were more anxious \([M = 4.22, SD = 1.02 \text{ (first-year)}, M = 4.79, SD = 0.94 \text{ (non first-year)}, t(164) = -3.72, d = 0.58 \text{ (anxiety; pre)}; M = 4.57, SD = 1.07 \text{ (first-year)}, M = 5.26, SD = 1.09 \text{ (non first-year)}, t(164) = -4.08, d = 0.63 \text{ (anxiety; post)}]\) and found chemistry to be less intellectually accessible than first-year students \([M = 3.35, SD = 1.01 \text{ (first-year)}, M = 2.76, SD = 0.95 \text{ (non first-year)}, t(164) = 3.83, d = 0.60 \text{ (intellectual accessibility; pre)}; M = 3.18, SD = 1.19 \text{ (first-year)}, M = 2.44, SD = 0.97 \text{ (non first-year)}, t(164) = 4.34, d = 0.67 \text{ (intellectual accessibility; post)}]\). Medium effect sizes were found suggesting a considerable difference in attitudes between first-year and non-first-year students.

Similar results for the fall 2009 class were found (Figure 10). Non first-year students reported negative attitudes consistently throughout the semester on anxiety, intellectual accessibility, and emotional satisfaction (Supplementary Table S8 and S9). Even though there was a smaller sample size in fall 2009, the effect sizes were larger (between 0.55–1.00 in fall 2009 vs. 0.58–0.67 in fall 2008), indicating a strong and meaningful attitudinal difference in terms of class year.
Difference in Self-Concept by Sex

Figures 11 and 12 show mean scores on the self-concept variables between males and females across pre/post time points in 2008 and 2009, respectively. Males did not differ from females in terms of their self-concepts towards chemistry throughout the semester in both years.

Difference in Self-Concept by Class Year

Figures 13 and 14 present results for the five self-concept variables across pre/post time points according to class year (first-year, non first-year) for 2008 and 2009, respectively. At the beginning of the semester, academic, chemistry, and math self-concept significantly differed between first-year and non first-year. First-year students reported more positive academic self-concept \( M = 5.09, \ SD = 0.87 \) (first-year), \( M = 4.64, \ SD = 0.83 \) (non first-year), \( t(156) = 3.27, \ d = 0.62 \), chemistry self-concept \( M = 4.20, \ SD = 1.28 \) (first-year), \( M = 3.55, \ SD = 1.09 \) (non first-year), \( t(156) = 3.41, \ d = 0.54 \), and math self-concept \( M = 5.05, \ SD = 1.13 \) (first-year), \( M = 4.37, \ SD = 1.19 \) (non first-year), \( t(156) = 3.68, \ d = 0.59 \). At the end of the semester, only chemistry self-concept was different. First-year students reported more positive chemistry self-concept again \( M = 4.24, \ SD = 1.26 \) (first-year), \( M = 3.41, \ SD = 1.13 \) (non first-year), \( t(156) = 4.28, \ d = 0.69 \).

In fall 2009, no self-concept variables significantly differed for class year groups, however, at the end of the semester, only chemistry self-concept significantly differed between first-year and non first-year. First-year students reported higher chemistry self-concept than non first-year.
Medium to large effect sizes were found, suggesting there is a considerable difference in self-concept between first-year and non first-year students.

Discussion

This study looked at exam achievement measures and affective student outcomes (self-concept and attitude) associated with PLTL participation. The hypothesis that students participating in PLTL would achieve at a higher level and have enhanced attitudes and self-concept than students not participating was not confirmed. This negative result was surprising, and it contrasts with a number of other studies which have shown significant positive achievement differences for students engaged in PLTL.

We suggest several possible reasons for these different results. First, perhaps we were not really implementing PLTL properly and thus our claim that students experienced PLTL is not valid. Second, perhaps the controls we put in place for time-on-task created some competitive advantage for the students not participating in PLTL. A third reason could be that in previous studies, the “PLTL treatment” embeds some other factor, like quizzes, which may provide a learning advantage through the “testing effect” (Pyburn, Pazicni, Benassi, & Tappin, 2014). Thus, it may not be social knowledge construction that confers the outcome benefit, as others have suggested. Fourth, our experimental conditions and assessments, being more uniform and controlled than those in other studies, may have neutralized unknown confounding factors.
Finally, our study design provided explicit control for self-selection between the students included in PLTL and those who were not. This control is missing from all other studies of PLTL effectiveness. We will address these potential alternative explanations in turn.

**Was PLTL Properly Implemented?**

Since, a possible interpretation for our results lies in the fidelity of implementation of the PLTL instructional model (as the “treatment”), it is important to describe how closely our implementation compares. The originators of the PLTL model claim that six components are critical for achieving positive student outcomes (Gafney & Varma-Nelson, 2008): faculty involvement, integration with the course, leader training, appropriate materials, appropriate setting (room, group size, duration), and administrative support.

The course instructor was sufficiently involved. This instructor was supportive of PLTL and met with the corresponding author (the instructor for the PLTL leaders) at least weekly to discuss course content. The course instructor and author were extensively experienced with the content and goals of this general chemistry course, having taught and coordinated together many times. The PLTL leaders had been students of both of these instructors in previous years and thus familiar with teaching approaches and course assessments. Because of the randomized study design, the course instructor was blind to student participation in PLTL and the PLTL instructor was not involved in course exams or grade assignments.

---

**Figure 9.** Error bar graphs showing 95% confidence interval for mean scores on ASCI subscales for first-year and non-first-year groups for pre(solid)/post(dash) time points (fall 2008). Top to bottom in legend is left to right on graph. Top to bottom in legend is left to right on graph.
PLTL was well integrated. At the time of this study, we had seven years of experience with PLTL in this course. Students were informed of PLTL by the course instructor or a PLTL student coordinator, and sign-up occurred within the first 2 weeks of the semester. Regular attendance and participation with the assigned group were expected, and participation contributed explicitly to the course grade, with loss of points for missed sessions. Communication about PLTL occurred through the Blackboard management site for the course.

Leaders were selected based on prior course performance and were provided “training” via the two-credit course described in detail above. The corresponding author was the instructor, who contributed both content expertise and learning specialist expertise. We cannot claim that PLTL was implemented as a single uniform experience for all students. In the years of this study, PLTL leaders included 67% new leaders with the balance having at least one year’s experience with PLTL. We know via leader written reflections, student evaluations, and peer observation that different leaders had different levels of success in attaining the goal of using facilitated discussion rather than tutoring. Although one could argue that this means that some students in the study population might not have experienced the value of PLTL in full, this same leader-preparation routine led to apparent PLTL advantages in other years (Figure 4). We looked for student differences as a function of leader identity and experience, but small N values led to lack of sensitivity for these comparisons. We note that few studies speak to the issue of leader variability or the internal workings of groups (Black & Deci, 2000; Kulatunga, Moog, & Lewis, 2014). Our leaders had the freedom to select from a set of common PLTL-style instructional materials or to
bring their own ideas to the study sessions. Some other implementations of PLTL have constrained leaders to cover a common activity or set of activities that may have been tied more explicitly to test content, so it is possible that a closer linkage of materials to assessments provided an advantage in those instances. We note again (Figure 4) that in the years surrounding the 2008–2009 cohorts, with the same set of materials, PLTL participants often had stronger course achievement. This suggests that the materials and their use was not the cause of the failure to find a PLTL advantage in the 2008 and 2009 years.

Considering setting, size, and duration of group sessions, we described group sizes in detail above as being aligned with the recommended size for PLTL groups. Groups met in a range of smaller class and conference rooms with movable chairs and sufficient board space to support group work. Leaders provided “room reviews” so that we could avoid choosing a room that compromised goals. We provided leaders whatever materials they needed to work in the space they had. In terms of duration, we could be rightly challenged in that our scheduled contact time was 50 or 80 minutes weekly. The suggested PLTL duration (Gafney & Varma-Nelson, 2008) is 120 minutes. Perhaps our shorter sessions yielded less effective outcomes. In Figure 4, the years 2009 and 2011 were 50 minute implementations and all other years, including 2008, were 80 minutes. Although 2011 had shorter sessions, one of the sections showed a significant PLTL advantage similar in size to the advantage in some of the years with 80-minute sessions. Consequently, the exam score differences observed do not seem related to session duration. This

*Figure 11.* Error bar graphs showing 95% confidence interval for mean scores on CSCI subscales for males and females for pre(solid)/post time points (fall 2008). Top to bottom in legend is left to right on graph.
suggests that the 2008 and 2009 study years were not special in disadvantaging PLTL because of contact time.

Lastly regarding administrative support, coordinating PLTL has been part of the corresponding author’s teaching responsibilities for more than a decade. The Chemistry Department staff has continuously provided assistance in identifying and scheduling rooms for group meetings, purchased materials for leaders, and provided hourly pay to leaders with more than a year’s experience. Several of these veterans are hired as coordinators for the program.

All together, the detail above demonstrates that PLTL as implemented at UNH incorporates the six critical elements for a strong program. Now, the other four explanations for the lack of a PLTL advantage in 2008 and 2009 are discussed.

Were Time-on-Task Controls Sufficient?

An important aspect of this study is the explicit attempt to control for time-on-task. In some studies, voluntary participation in PLTL may create unequal study effort relative to non-PLTL students. In this study, all non-PLTL participants were required to invest an amount of time equivalent to that of PLTL participants. We believe that this expectation was met by most students, thus removing one obvious potential source of imbalance in study time. We cannot claim that the total study time committed by PLTL students and those in the comparison group are equivalent since we did not attempt a full census, so it is possible that unknown differences in study time...
could mask an effect of PLTL alone. We can say that the documented study time in PLTL or alternate study activities could represent a significant portion of the outside-of-class study time. Students report an average of seven to nine “hours spent on this course per week” on their end-of-semester course evaluations. To the extent that this student report of their total time commitment is accurate, we have provided a control for a portion of that total time. Some studies have attempted to document study time and activity (Chan, 2014; Kamp, Dolmans, van Berkel, & Schmidt, 2012; Plant, Ericsson, Hill, & Asberg, 2005). The finding is that time by itself is not strongly related to performance, but quality of study is. One final comment is that the 2007 classes balanced time-on-task between PLTL participants and all other students as well. In that year, PLTL participants had stronger total exam scores (Figure 4 and Supplementary Table S5). This is the closest comparison to the 2008/2009 data, with the only difference being that 2007 involved student self-selection and 2008/2009 was random assignment. These results reaffirm the likely importance of initial student motivation in the interpretation of the achievement results of PLTL students.

**Do Other PLTL Implementations Include a Hidden Variable?**

Through professional conversations, we are aware that some PLTL programs have included quizzes during PLTL session time, although the published studies do not mention this use of quizzes. Nevertheless, in the interest of exploring mechanisms by which the PLTL advantage arises, we speculate about the potential effect of quizzing. There is a well-documented positive effect on student achievement called the “test effect,” which indicates that the process of preparing
for and working through a quiz may enhance learning and later test achievement (Pyburn et al., 2014). A study of PLTL vs. non-PLTL involvement may, therefore, indirectly involve comparing students who take quizzes (in PLTL session) with students who don’t (or who have some other expectation). Similarly, students not involved in PLTL may be given access to the PLTL materials but may not use them for study. Hidden distinctions like these could explain some of the apparent PLTL advantage found in other studies. At UNH, we did not include course quizzes during PLTL sessions.

Assessment and Regimen for Student Outcomes

Most other studies have relied on gross measures of student success, in particular, the overall grade distribution at the end of the semester vs the historical record for a course. Legitimate concerns may be raised with this kind of quasi-experimental historical comparison. These concerns are not addressed explicitly in those studies. In particular, one must be able to claim that course content, exam content, exam scoring and scaling, and student prior knowledge are the same, and that potential for subtle instructor/researcher bias, and the like are under control. Although it has been suggested that such controls may be unnecessary in higher education settings (Deslauriers and Wieman, 2011), we chose a design that avoids having to make this claim. In the current study, all of the aforementioned factors have been eliminated as potential confounding sources.
Controlling for Motivation

One critically important difference in this study is that a fully randomized design was used. Unlike many other PLTL studies (Hockings, DeAngelis, & Frey, 2008; Jayaratne, Thomas, & Trautmann, 2003; Lewis, 2011; Mitchell et al., 2012; Wamser, 2006), our study design eliminates self-selection bias. This is of concern because students who voluntarily participate in PLTL may have higher motivation and more positive attitude toward study and collaborative work than non-participants. In fact, in the years when PLTL was optional, PLTL participants as a whole did better than non-participants (Figure 4). This result reproduces what other PLTL studies have found in concurrent or historical comparisons. Furthermore, as partial confirmation of this motivational bias, we note that most students who dropped the course in the 2008/2009 randomized study never engaged with the invitation to participate in PLTL. Thus, our failure to find an achievement advantage suggests that a critical feature of the apparent success of PLTL is the initial motivation to elect to take advantage of this learning support structure. In the current study, motivation was controlled by the selection process—both PLTL participants and independent studiers had signed up for PLTL but only some were selected randomly for inclusion.

We initiated this discussion by enumerating several alternative hypotheses for the failure to find a performance advantage for PLTL students in the randomized study years 2008 and 2009, and the existence of advantage in other years at UNH and in studies at other institutions. We could find no consistent reason other than the motivation of students to elect to participate. Allowing for the possibility that particular subgroups of students might respond differently, we looked at disaggregated comparisons.

Disaggregated Comparisons and Population Characteristics

We looked at differences in students’ performance trajectories over time (Exams 1, 2, 3 and final). Because exams varied in content and difficulty, scores were standardized to z-scores. Mixed-repeated measures ANOVA showed no significant improvement or decline in performance across the four exams within PLTL vs. non-PLTL groups.

Turning to affective characteristics, it is possible that students with different grades had different self-concept and attitude profiles, so we disaggregated by grade. Overall, high performers (A grades) reported higher attitudes and self-concept scores than low performers (C or below), as was reported in a recent study (Chan & Bauer, 2014). When looking at only the PLTL group or only the non-PLTL group, students with higher performance reported higher chemistry, math, and academic self-concept scores than lower performers. Furthermore, higher performers reported higher scores on: intellectual accessibility, emotional satisfaction, interest & utility, and lower scores on anxiety. Between group comparisons (PLTL vs. non-PLTL) showed a small differential effect with performance levels ($\eta^2 = 0.01$): students with A grades who are not in PLTL have a slightly stronger self-concept than those who are in PLTL, and the reverse is seen for students with C or lower grades. But this difference is not consistent across the components of self-concept nor between the study years.

Clear affective and performance differences were observed when disaggregating by other characteristics. In particular, first-year students show some evidence of outperforming non-first year students on tests. Furthermore, year in school and sex show some persistent differences in affective measures. Over the course of the semester, certain aspects of attitude and self-concept showed slight but significant decline. Our results complement the focus of other recent studies that have begun to track student affective development over time, for example differentiating by sex (Brandriet et al., 2011), race/ethnicity (Villafane, Garcia, & Lewis, 2014), college major and age (Adams, Perkins, Podolefsky, Dubson, Finkelstein, & Wieman, 2006), levels of performance.
(Zusho, Pintrich & Coppola, 2003), and course topics (Barbera, Adams, Wieman, & Perkins, 2008) using multiple surveys: Colorado Learning Attitudes about Science Survey (CLASS) (Adams et al., 2006), Chemistry Attitudes and Experiences Questionnaire (CAEQ) (Dalgety, et al., 2003), Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich & de Groot, 1990), and Attitudes to Subject of Chemistry Inventory (ASCI) (Bauer, 2008). The one clear message is that student affective trajectories do not have a uniform profile in large enrollment courses, and that disaggregation will be more productive for research.

Disaggregation of students by sex confirmed the existence of a difference in attitude toward chemistry between males and females. By the end of the semester, females reported a less positive attitude in terms of anxiety, fear, and intellectual accessibility (in 2008) and emotional satisfaction and intellectual accessibility (in 2009) than males overall. This is consistent with Brandriet et al.’s (2011) findings where females reported a less favorable attitude towards chemistry (emotional satisfaction, intellectual accessibility) than males. Salta and Tzougraki (2004) found that females from high-schools in Greece held a more negative attitude in terms of their belief of difficulty of chemistry. Likewise, Cheung (2009) found that Hong Kong males report a more favorable attitude toward chemistry lectures than females across different grades in high-school. In a meta-analysis covering literature from 1970 to 1991, Weinburgh (1995) found that males had a more positive attitude in all areas of science compared to females. These studies collectively show that females have an overall more negative attitude toward chemistry. Chemistry self-concept differed between males and females only at the beginning of the semester (in 2008), with males reporting a higher self-concept than females. However, no differences in self-concept were found for sex by the end of the semester for both years. Perhaps a single college course does not have sufficient impact on self-concept to cause substantial movement over a semester. In a pilot study, Lewis, Shaw, Webster, & Heitz (2009) observed a modest gain in chemistry self-concept within the semester when a student-centered pedagogy was used, however sex differences in self-concept were not studied.

Disaggregating by years of college experience confirmed the existence of more positive attitudes and chemistry self-concept among first-year students than among veteran students, a difference which was sustained over a semester. Why non first-year students have less positive attitudes and self-concept may be due to an existing fear, dislike, or lack of competence in chemistry which may have led them to delay taking chemistry. It is also possible they may be course repeaters who had a difficult experience the first time around.

One issue that could qualify the interpretation is the level of participation in survey completion. Students who elect to complete surveys may be qualitatively different from those who do not. We looked at the response rate for PLTL vs. non-PLTL students in completing both pre/post surveys. In 2008, 49% were in PLTL vs. 43% not; in 2009, 23% were in PLTL vs. 26% not. The response rates for both groups are comparable within each study year. Furthermore, we compared distribution by grade, sex, class year, and math SAT scores for survey completers (both PLTL and non-PLTL) vs. non-completers. We found both groups to be similar with regards to their grade distribution (A: 16% vs. 14%; B: 22% vs. 14%; C: 35% vs. 31%; D: 13% vs. 18%; F: 14% vs. 22% survey completers vs. non survey-completers, respectively), class year (FY: 56% vs. 50%; non FY: 44% vs. 50%), and math SAT scores (high: 19% vs. 21%; medium: 66% vs. 67%; low: 14% vs. 13%). The only difference found between completers and non-completers was regards to their sex (F: 69% vs. 44%; M: 31% vs. 56%). Hence students who decided to complete the surveys are not qualitatively different from students who decided not to complete the surveys (except for sex).

In the context of this study and student population at a single institution over two subsequent years, PLTL was not shown to provide a learning advantage or a change in student attitudes or self-
concept relative to students engaged in an equivalent amount of study time in independent groups or tutored settings. In this respect, the failure to find a PLTL advantage can be cast in a different way—one can say that students who spend time working in groups with an experienced peer succeed just as well as students who take serious advantage of traditional professor-led review or tutoring sessions and self-organized groups.

Implications

The results of this study lead us to make recommendations for both instructional practice and research. First, we are not arguing that PLTL is not valuable as an academic program. Our results, confirming those of other studies, show that PLTL helps students learn if they take advantage of it. However, it seems that any student making a sincere commitment to study can be just as successful via other mechanisms. As reported by other studies, the quality of the engagement may matter most (Chan, 2014; Kamp, Dolmans, van Berkel, & Schmidt, 2012; Plant et al., 2005). Since our results strongly suggest that the students who reap the benefits of PLTL may be the ones predisposed to seize the opportunity, at-risk students may be opting out. Understanding why students opt-out to their own disadvantage would be useful to know. From an instructor’s point of view, talking with students about what it means to engage in quality study would serve all students, and may encourage PLTL participation for those who need it most. Furthermore, as most traditional support mechanisms (tutoring, office hours, etc.) are person and schedule limited, one big advantage of PLTL is that it expands the opportunities for quality study time. This might be particularly important for larger classes and those with a history of substantial student attrition. A second big advantage is that PLTL offers a professional development opportunity for the leaders, a benefit that has been reported on elsewhere (Gafney & Varma-Nelson, 2008). On any campus, students are our biggest resource and evidence suggests that their intellectual development is strongly mediated by their interactions with peers (Astin, 1993). Why not leverage their intrinsic motivation to help? A professor’s time and influence can only go so far. PLTL expands the opportunities for meaningful intellectual engagement with the subject matter.

In terms of research, some important issues are raised by this study regarding the design and interpretation of future studies.

- We should investigate mechanisms not just final outcomes. In terms of the inherent value of PLTL, we suggest that future studies adjust the focus away from whether PLTL provides a learning benefit to enhancing the value of the PLTL learning environment and to finding out which students might benefit most. This means attending to how leaders create and sustain a positive supportive learning environment (Black & Deci, 2000), to the nature of the discourse occurring within groups (Christian & Talanquer, 2012;) and to the characteristics of individuals within groups (Jensen & Lawson, 2011; Sandi-Urena, Cooper, & Stevens, 2011).
- A curriculum model is not a monolithic entity. It is a collection of interrelated parts. As such, it is risky to call any model “a” treatment because those parts are implemented and adapted by human beings. Thinking of a model as “a” thing tends to direct the argument to “yes it does” and “no it doesn’t,” when we’re really interested in knowing “how it does or doesn’t” and for whom. In defense of our study results, we provided substantial fine-grained detail about our implementation of PLTL. This level of detail is not always included in published reports, perhaps because authors and editors are always looking to streamline manuscripts. This could mean that important hidden variables might be overlooked, such as quizzing during PLTL sessions or uniformity of PLTL content coverage among groups or degree of use of PLTL materials by non-PLTL participants. Again, this is thinking mechanistically. Thus, we would argue that more detail is always
better than less in study reports. Should a meta-analysis of PLTL be conducted in the future, having those details would be very valuable.

- Knowing what happens does not necessarily provide insight into how it happens. The discussion section points to numerous factors (anticipated and unanticipated) that could play a role in how effective “PLTL” may be for students. Providing rigorous control of those factors in order to demonstrate causality is challenging but sharpens inferences. Having data across multiple settings and conditions helps to eliminate confounding explanations.

- One clear limitation of this study is that it was performed at a single institution, whose student population is different from that of many other institutions. To provide a few points of comparison, UNH students have a mean combined SAT score of 1,090, 81% received some kind of financial assistance, and the minority student population is about 10%.

One could ask, given the outcomes of the current study, whether the commitment to PLTL will be sustained at UNH. Yes it will; in fact, we have expanded into other courses. The driving force is student enthusiasm of both leaders and group participants, and the commitment of several instructors to support the program for their courses. But the awareness that the benefits of PLTL may follow from the initial motivation to participate raises concerns regarding at-risk students who may have opted out. This has caused us to be careful in how we talk about PLTL at the start of the semester. Instead of emphasizing the potential for enhancing one’s grade through participation (an extrinsic motivator), we make explicit how PLTL could build self-efficacy and its supportive environment (to support intrinsic motivation). We have also been talking with student academic support and student life organizations on campus for advice on how best to recognize, advise, and refer students for assistance before they get into academic difficulties.

References


*Journal of Research in Science Teaching*


*Journal of Research in Science Teaching*


Supporting Information

Additional supporting information may be found in the online version of this article at the publisher’s web-site.