Collaborative Research: Longitudinal Impact of PLTL on Student and Peer Leader Retention of General Chemistry Concepts and Attitudes toward Chemistry

Studying the longitudinal impact of educational interventions is rarely conducted in science education.[1, 2] However, such investigations are necessary to examine the impact of interventions on long term student retention of concepts and utility of concepts in novel situations, in short on meaningful learning.[3] Without longitudinal investigations, there is the potential that educational interventions facilitate rote learning, where knowledge is used only for the immediate assessment and not subsequently retained. The proposed research explores the long-term impact of Peer-Led Team Learning (PLTL), a widely used reform pedagogy with a considerable evidence base for effectiveness within the implemented course. The proposed research relies on a parallel study design to explore the long-term impact of PLTL in both a naturalistic setting, an upper-level class in Analytical Chemistry, and in a controlled study. Combining the results in both studies will provide a robust picture on the long-term academic impacts of PLTL. Additionally, the proposed research examines the long-term impact of PLTL on students' affect, an area of recent attention given the clear importance of affect on students' persistence and academic achievement.

Peer-Led Team Learning

Peer-Led Team Learning is a nationally disseminated model for reform teaching (DUE-9972457 and DUE-0231349) that has been implemented at over 150 institutions.[4] In essence, PLTL relies on peer leaders, undergraduates who have successfully completed a targeted course, to return to the target course and lead students in small group learning activities. More specifically, PLTL has six critical components: 1) faculty involved with the learning activities and peer-leader training, 2) the learning activities are integral to the target course, 3) peer leaders are well trained and supervised, 4) the learning activities challenge students which promotes collaboration, 5) the course is set-up to promote group activity and 6) PLTL is supported by the institution.[5] One particular advantage of the approach is the ability to facilitate active learning with large classes by scaling the number of peer leaders.

Evaluations of PLTL have placed considerable attention on metrics of student learning within the target course. Past research has shown significant improvements with PLTL over traditional lecture-based instruction in student performance on common exams, final course grades and pass rates.[6-10] The strong majority of this research relied on quasi-experimental methodology resting on the assumption students from each pedagogy are comparable and represent independent observations. One exception is PI Bauer's recent work in considering students who were randomly assigned to PLTL that resulted in no evidence of academic benefit.[11] A possible explanation is that the random assignment provided a control for student motivation that was not achieved with prior quasi-experimental designs. PI Lewis found that the assignment of classes to PLTL (where all students were compelled to attend PLTL as part of their grade) led to a 13 to 15% improvement in the pass rate with PLTL while maintaining performance on a common exam.[12, 13]

In a follow-on study, PI Lewis investigated the longitudinal impact of students enrolling in PLTL compared to traditional lecture-based instruction in first-year General Chemistry, on enrollment and course grades in subsequent chemistry courses.[14] The results showed a significant difference in enrollment in the course immediately following General Chemistry favoring students who enrolled in PLTL. This difference is likely attributed to the higher pass rate associated with PLTL in General Chemistry. The difference is mitigated through student attrition in subsequent courses and is not discernibly significant in upper-level chemistry courses. No significant effects of PLTL in General Chemistry was found on course grades in subsequent chemistry courses. It may be the case that PLTL is effective in promoting students' retention of General Chemistry concepts but the metric of student grades in subsequent courses is too broad to measure this retention. Longitudinal research that examines assessments specific to the retention of General Chemistry content is needed to explore this possibility.

An additional area of research that has been considerably understudied is the impact of PLTL on peer leaders' content knowledge and skills. The largest portion of research has relied on peer leaders self-report of their learning gains from peer leading.[5] For example, Gafney and Varma-Nelson surveyed 119 former peer leaders from 11 different institutions.[15] Survey respondents rated thirteen different learning experiences on a Likert survey. Past experience as a peer leader had the highest average rating of the thirteen options exceeding their experience as a student in the PLTL model. In responses to an open-ended survey item nearly half of the respondents indicated that serving as a peer leader provided a more thorough knowledge of the discipline and helped in problem-solving abilities. The indication of learned knowledge and abilities indicated in the survey is promising, but reliance on peer leaders' self-report is problematic as it is uncertain that peer leaders have an appropriate context for rating the gains described or can utilize their learning gains in novel situations.

Measuring Student Learning

Creative Exercises (CEs) are an assessment technique that can measure students' use of content in novel situations.[16, 17] In brief, a CE provides students a prompt that describes a situation relevant to the course, such as "Dissolving 5.0 grams of $AgNO_3$ (*s*) into 500 mL of an aqueous solution of 0.1 M NaCl." Students are asked to describe as many statements as they can that are distinct, correct, and relevant to the prompt. Example statements may include creating a net ionic equation, solving the mass of the precipitate, identifying amount of excess reagent, describing physical characteristics of the precipitate based on ionic bonding or the charge or atomic structure of the ions. The prompt is designed to match content that is currently being assessed in the course (labeled the target concept), in this example, precipitation reactions. Credit is awarded for each statement that a student can list which satisfies the following criteria: each statement is correct, distinct from the prompt and other statements listed, and relevant to the target concept or prior concepts in the course. To score a CE, the instructor brainstorms a list of likely student answers and uses the aforementioned criteria to guide a decision on unanticipated student responses.

During a CE assessment, students are generating their own statements related to the prompt and are showing evidence of cognitively linking two concepts. The emphasis on linking concepts matches Novak's model of human constructivism where meaningful learning would have more links made between new knowledge and previous conceptions, in contrast to rote learning or memorization.[3] Meaningful learning would be evidenced by greater long-term retention of the newly learned content, particularly as it relates to the utility of the content. The emphasis on the learner linking new knowledge with previous conceptions also matches numerous other models for learning including the constructivist theory paradigm that describes a learner building new knowledge based on the learner's previous conceptions and experiences, Marton and Saljo's Deep Learning versus Surface Learning, Linn and Eylon's Knowledge Integration Perspective and Stevens' *et al* model of learning progression.[18-21]

In this regard, CEs as an assessment technique is similar to having students generating concept maps given the shared emphasis on linking concepts. CEs have a simpler scoring method than concept maps, as no assumption on the structure of knowledge is needed. Instead, in CEs, the scoring system is based only on the number of statements the student provides that satisfies the above criteria. Scoring concept maps is problematic in that multiple structures for knowledge construction are plausible such as hierarchical, cyclical or associative (highly networked).[22, 23] The structure for knowledge can reasonably vary depending on the content, but the scoring of concept maps based on the structure must necessarily endorse one scheme (likely the instructor's scheme) over any others which may be equally plausible.[24, 25] In particular with large classes, concepts maps as assessments would need a scoring scheme that is flexible enough to consistently score a large variety of student maps created and the extensive time required to score a large number of student maps.

In contrast, a class set of seventy CEs can be scored in approximately one hour. Evidence for the validity of student responses to CEs as a measure of content knowledge in a General Chemistry classroom has been collected by PI Lewis through examining the content coverage, scoring structure, inter-rater reliability, and correlations with traditional chemistry assessments.[17] More recently, a qualitative investigation of student responses to CEs showed students successfully linking a wide variety of concepts throughout the course. Additionally, student responses showed several instances of students' misunderstanding the limits of models that are not shown with traditional assessments.[26] For example, students applied the ideal gas law to solutions, created electron configurations for molecules, labeled reactants as precipitates, and rated the Lattice Energies of molecular compounds.[26] Such misunderstandings indicate a lack of coherence in students' conceptualization of the content that is not traditionally addressed in instruction or measured by assessments.

Finally, CEs have been used once previously in upper-level chemistry courses to investigate students' long-term retention of earlier concepts. Warfa and Odowa developed a series of CEs for Biochemistry targeting thermodynamics, enzyme kinetics and the structure of biomolecules.[27] Student responses showed a strong reliance on current course content with thermodynamics and enzyme kinetics. In contrast, with the structure of biomolecules, students invoked a considerable range of concepts from foundational courses including pH, ionization, stereochemistry and organic functional groups. The difference in reliance on foundational knowledge is attributed to the context of the CE, in particular it is possible that providing details in the initial prompt relevant to foundational course material activates a greater reliance on foundational course material.

Attitudes toward Chemistry

Student understanding of chemistry ideas and skills is one important indicator of intellectual development, but two constructs in the affective domain provide complementary insight, in particular student attitude toward chemistry as a discipline and student self-concept as a learner of chemistry. An attitude is a learned tendency to respond positively or negatively to a given attitude object.[28] It forms through past experiences and significant people in one's life.[29] Attitude is often subdivided into components of affect (feelings and emotions), cognition (beliefs), and behavior (actions).[30, 31] Positive attitudes have been found to correlate with high chemistry achievement.[11, 30, 32-37]

The second affective construct is self-concept, an individual's persistent perception about themselves in terms of their understanding or ability to learn in a broad area of knowledge.[38-40] Students with stronger and more positive self-perceptions have been found to have higher academic achievement[32, 40, 41], including in college chemistry.[34, 42, 43]

PI Bauer has developed measures for both student attitude and self concept. The attitude measure has since been refined by Xu and Lewis and titled the Attitude toward the Subject of Chemistry Inventory version 2 (ASCIv2).[44] The ASCIv2 measures students' perceptions of the intellectual accessibility and emotional satisfaction associated with chemistry. The measure relies on a semantic differential where students identify a place on a seven point scale between two opposed adjectives such as confusing and clear. In exploring the validity of General Chemistry students' responses to the ASCIv2, a confirmatory factor structure showed reasonable fit and the scores showed predictive validity of students' academic achievement in the course. The second measure is the Self Concept Inventory (SCI) that is a Likert style instrument of 40 items.[38] The instrument measures students' self concept in: Chemistry, Math, General Academics, Academic Enjoyment and Creativity. Evidence for validity of student responses to the SCI include exploratory factor analysis and finding higher self-concept among upper-level chemistry students than with entering chemistry students.

Research Goals

The over-arching research goal is to provide a holistic picture of the long-term impact of PLTL on both peer leaders and students on content retention and attitude toward chemistry.

Examining the long-term impact of education interventions, such as PLTL, is necessary to make the case that the benefits observed in past literature will last and that meaningful learning of content has been achieved. To achieve this goal, the proposed work seeks to develop and implement a series of assessments in an upper-level chemistry course, to measure student retention of General Chemistry (or introductory chemistry) content. In particular, the assessments will include a series of CEs that measure students' abilities to employ General Chemistry content in an upper-level chemistry course, representing a novel application for the foundational content. In addition, student understanding will be measured using a conventional longitudinal design to triangulate the results.

It is hypothesized that students with experience as peer leaders would exhibit particularly strong retention of General Chemistry content as a result of leading students on General Chemistry sessions on a weekly basis throughout a semester. It is also hypothesized that the learning gains observed in the literature for students who experience PLTL will result in long-term content retention as well.

Additionally it has been shown that students' affect toward chemistry impacts academic performance and persistence.[11, 34, 43, 45] The proposed work will determine the extent to which PLTL may affect student attitude towards the discipline and perceptions of themselves as learners of the discipline. It is hypothesized that the more intensive experience with content, through PLTL participation or even more with PLTL leadership, will generate a lasting positive attitude toward the discipline.

Methodology

The proposed research will adopt a parallel study design, employing two study designs concurrently to create robust results that address the research goal. The first study will occur in a naturalistic setting of students' first undergraduate class in Analytical Chemistry, often referred to as Quantitative Analysis. Quantitative Analysis is a typical upper-level required course for chemistry majors that relies on students successfully building upon knowledge learned in General Chemistry. The study will seek to measure students' applicable General Chemistry content knowledge within the course and in particular will strive to measure students' use of this content knowledge in applications particular to Quantitative Analysis. The strength of the design is the collection of data in the authentic environment, in particular by exploring the utility of prior knowledge. The weakness of the design is the lack of control over factors such as time between enrolling in General Chemistry versus Quantitative Analysis. This lack of control is inherent in studying naturalistic settings.

The second study will be designed as a conventional longitudinal study. Data will be collected for the same cohort across multiple time periods over a long period of time. This design allows for strong control over factors such as time and is ideal for demonstrating change over time. However, the study lacks the ability to apply knowledge in an authentic setting as described in the above naturalistic study.

The two-study design has complimentary strengths. By conducting both in parallel, the results may be combined to offer a more robust case regarding the long-term impact and identify the conditions under which long-term impact may be demonstrated.

Research Settings and Project Team

To broaden the applicability of the developed assessments and measure the impact of PLTL across institutions, the proposed research will take place across four universities. Of the four universities, two have active PLTL programs in place in General Chemistry and two do not, providing ready comparison. All four have American Chemical Society approved chemistry major programs. One comparative pair (PLTL and non-PLTL) is of institutions that are each large (greater than 40 thousand students) doctoral universities rated Highest Research Activity by Carnegie classification. The second comparative pair is of institutions that are each moderate-sized (5 to 15 thousand students) universities rated Higher Research Activity or Masters-granting by Carnegie classification. Each comparative pair of institutions is geographically proximal. Each university is described in more detail below.

The University of South Florida (USF) is a large research-intensive university that has a Ph.D. program in Chemistry. Each year USF enrolls approximately 2,950 students in General Chemistry I; 2,200 students in General Chemistry II; and 350 students in Quantitative Analysis. PLTL is in place in all classes of General Chemistry I except summer terms and in place in approximately half of the General Chemistry II classes. Peer leaders undergo weekly training that averages roughly two hours per week that focuses on pedagogical content knowledge and peer leaders work with students about two hours per week throughout a semester. Approximately 100 students per year serve as peer leaders supporting General Chemistry classes. In a recent survey 16% of students in Quantitative Analysis had either been or currently are a peer leader for General Chemistry.

The University of Central Florida (UCF) is a large research-intensive university that has a Ph.D. program in Chemistry. Last year UCF enrolls 2,631 students in General Chemistry I; 1,793 students in General Chemistry II and 154 students in Quantitative Analysis. UCF does not currently employ PLTL in chemistry.

The University of New Hampshire (UNH) is a smaller doctoral university with high research activity enrolling 600 students in General Chemistry I, 450 in General Chemistry II, and 25 students in Quantitative Analysis. PLTL serves about 40% of the General Chemistry population each semester, where students may elect to participate. About 40 student leaders support General Chemistry annually, and about 10% of them take Quantitative Analysis. Peer leaders participate in a weekly 1.5 hour course that provides an introduction to human cognition, group dynamics management, and pedagogical content knowledge. Each leader is then responsible for one 80-minute group study session weekly.

The University of Southern Maine (USM) is a larger Master's-granting institution enrolling 300 students in General Chemistry I, 200 in General Chemistry II, and 25 in Quantitative Analysis. USM does not currently employ PLTL in chemistry.

The project team will be composed of PI Lewis (USF), a graduate student at USF advised by PI Lewis, PI Bauer (UNH) and a graduate student at UNH advised by PI Bauer. PI Lewis is currently the General Chemistry coordinator at USF and has initiated PLTL in General Chemistry II at USF. His expertise is in Chemistry Education Research where he has evaluated the impact of PLTL on student success at multiple institutions.[12, 13, 46] PI Bauer similarly is an expert in Chemistry Education Research and has directed and evaluated PLTL at UNH since 2000. His original background was analytical chemistry teaching and research (plasma spectroscopy, environmental method development and assessment). He recently published analytical curriculum materials.[47] The project team will work with a group of faculty members who regularly teach Quantitative Analysis at each participating university (please see supporting documents) as discussed below.

Proposed Activities

Naturalistic Study

Year One; Design and Develop (July 2017 through June 2018)

The proposed activities begin with designing a set of assessments for Quantitative Analysis in Summer 2017. To start, the project team will request syllabi from each Quantitative Analysis professor and evaluate the syllabi for differences among topics or topic sequences. This information will be used to begin outlining common assessments that can be developed for common use at each university. The project team will then host a day and a half meeting, either in person or virtual attendance, for the panel of faculty members. The purpose of the meeting is to design a series of six Creative Exercises that target concepts which are common across all four settings. As informed by past research, three CEs will be designed as formative assessments for students and will be given as either homework or low-stakes in-class assignments. The primary purpose of these assignments is to familiarize students with the format of CEs. The other three CEs will be designed to implement as part of normal in-class exams. An example CE is shown in Figure 1, illustrating an analytical design issue requiring consideration of multiple concepts from general chemistry. Figure 1 also shows a partial list of potential correct answers, where the complete list will be used as a starting point for a scoring rubric. The instructors together will also design eight traditional assessment questions, either multiple-choice or closed-ended short answer, that will target General Chemistry concepts that the instructors identify as most relevant for Quantitative Analysis. These eight questions will be placed in the in-class exams.

Write down as many distinct, correct and relevant statements as you can about gravimetric or volumetric measurement of chloride by precipitation with silver ion. <u>Seven</u> distinct statements are needed for full credit on this problem.

Partial list of potential correct answers:

- Net ionic equation is Ag^+ (aq) + Cl^- (aq) → AgCl (s)
- To precipitate all the chloride, chloride must be the limiting reagent
- To ensure precipitation, excess silver drives equilibrium to the right
- Ksp for precipitation reaction must be large
- Chloride amount is calculated from mass of AgCl collected divided by molar mass AgCl
- If another ion is present that precipitates with silver, this may cause an error.
- Chloride forms a -1 ion by accepting 1 valence electron to complete its shell
- Silver forms a +1 ion by losing 1 valence electron
- Silver chloride is an ionic salt
- Silver chloride precipitates because it has a strong lattice energy
- Silver ion (chloride ion) is surrounded by water molecules
- Water molecules surrounding silver ion orient with O's facing the ion or water molecules surrounding chloride orient with H's facing the ion

Figure 1: Example Creative Exercise for Quantitative Analysis

In Fall 2017 an iterative development process will take place to ensure that these assessments measure student learning in line with the project team's intended processes. The designed assessments, both CEs and traditional, will be trial tested by incorporating them as assessments in Quantitative Analysis at USF and UNH (the PIs' home institutions). The instructor of the course and a member of the project team will score the CEs independently (and traditional assessments if automated scoring is not possible). Intra-class correlations, which measure the consistency in ranking of students across different graders, and Cohen's Kappa, which measure rater agreement while controlling for chance agreement, will be calculated as a measure of grading consistency. Past research suggests that intra-class correlations for CEs should exceed 0.8 and Cohen's Kappa should exceed 0.4.[17]

Additionally, instructor scores to the CEs and the traditional assessments (by question) will be collected along with scores on the assessments normally given by the instructor in the course. Correlations of scores between the set of common assessments and the instructors' normal assessments will be evaluated to examine the extent to which the assessments align. Past research suggests that correlations of approximately 0.5 between alternative assessment types are to be expected. This correlation value describes a moderate relationship, which is to be expected between alternative measures of chemistry knowledge. Correlation values much larger than 0.5 would indicate that the assessments developed were likely redundant with the normal assessments.

During the Fall 2017 semester, four students from each institution (eight total) will be recruited for interviews to investigate student response processes. Each student will be interviewed separately following the completion of each in-class exam that includes the new assessment items. The interview will first present the student with the same CE and traditional assessment items and ask the student to complete them as though they were taking them on an exam. The student will be encouraged to use a think-aloud approach describing their reasoning as they complete the task. An audio or video record will be obtained. Next the student will be presented with their answers from the actual exam questions and asked to describe the reason for any differences between the interview response and the exam response. The interviews will be reviewed by the project team for instances of students misinterpreting the intent of the assessment.

The resulting data will provide multiple lenses of validity to interpreting student responses to the developed assessments. First, the intra-class correlations between raters can provide evidence that the relative assessment scores within a class are not overly dependent on the rater. Second, Cohen's Kappa can provide evidence of any confusion associated with the scoring rubric. Third, correlations with normal assessments provide evidence of external validity using an additional measure of student learning. Finally, interviews on student response processes provide an indication of the extent to which students engage in the intended process for the assessments. During the Spring 2018 semester the project team will review the evidence for validity and generate recommended changes to the assessments developed.

Years Two and Three; Implement and Analyze (July 2018 through June 2020)

During Summer 2018 the faculty who designed the assessments will be reconvened for a virtual meeting. During the meeting the project team will present the validity evidence collected during the prior academic year and the recommended changes. The panel of faculty will discuss the recommended changes and reach a consensus on any modifications made to the assessments developed. The meeting will also discuss implementation decisions such as the testing conditions the assessments will be given in and the time allotted for each assessment. An effort will be made to make testing conditions comparable across settings but it is expected that logistical challenges will lead to some departures that will be noted for later analysis. The team will also discuss question security, student access to the research assessment items and the feasibility of using questions across multiple semesters. The final purpose of the meeting is to discuss Institutional Review Board approval at each research setting and to formulate a plan for the project team to apply for approval at each setting.

Implementation will be performed at each of the four research settings during Fall 2018 and Spring 2019 when Quantitative Analysis is offered. Across the four research settings, approximately 550 students take Quantitative Analysis each academic year. A brief survey will be given at the start of the term to the Quantitative Analysis class. The survey will be open ended and will be designed to provide information on the pedagogy students experienced in General Chemistry such as exclusively lecture, PLTL or other reform teaching styles. Additionally the survey will explore students' experiences helping others with General Chemistry either as a peer leader or informal teaching experiences such as tutoring General Chemistry students or leading study groups of General

Chemistry students. Each Quantitative Analysis class will also administer the affective measures ASCIv2 and SCI at the start of the term. Each class will administer three CEs as formative assessments and the other three CEs as summative assessments embedded in exams. The instructor for each class will create a copy of the student responses and then independently score the original student response.

The ungraded copy will be used to address the research questions. Both graduate students will begin scoring the CEs from in-class exams collected across all research settings. The decision to focus on in-class exams stems from earlier work, which showed that in-class exam CEs had stronger evidence of validity among student responses.[17] This decision also allows the instructors latitude to assign formative CEs as group-work or permit collaboration on take-home assignments if they choose. The graduate students will begin by independently scoring 25 CEs selected randomly and comparing their scores and discussing any discrepancies and any need to update the scoring rubric. Each graduate student will then grade the entire set of responses and discrepancies in scoring will be discussed and decided among the project team.

Starting in Spring 2019, the project team will also code student responses using the rubric as a starting code list and using an open coding scheme to add codes for incorrect student responses.[26] The coding will be performed where each response is coded independently by two members of the project team (each of the four members will code half of the total responses) using NVivo software. NVivo has the advantage of letting researchers in different locations compare coded lists remotely. The project team will discuss discrepancies between the two code lists until a consensus is reached. The scoring and coding of the student responses will provide the data to address the proposed research questions as described below.

The extent students utilize General Chemistry content will be determined through analysis of the codes pertaining to General Chemistry content. Each student's CE response will be classified as using a concept correctly, incorrectly or omitting the concept in their CE response. Additionally, success with General Chemistry concepts will be investigated by scores on the designed common traditional test items that target the same concepts. Thus, a measure of students' proficiency with each concept can be measured via traditional assessment along with their tendency to bring forth the content in a novel situation through CEs.

The coding of CE responses conducted in spring term will also allow a more nuanced qualitative description for how students utilize General Chemistry content. NVivo software facilitates an examination of codes that appear together and permit researchers to identify patterns. For example, codes that appear together may identify a sub-group of students that regularly use concepts related to solution chemistry and whether this topic is used correctly. Patterns among codes will be investigated as they have the potential to inform Quantitative Analysis regarding the correct and incorrect frameworks students bring forth from their General Chemistry experiences. As with past work with CEs, it is expected that student responses will also inform instructors regarding unexpected, incorrect links students make across content topics.

The impact of enrolling in PLTL as a student on retention of student concepts will be determined via the same set of student responses collected in Quantitative Analysis. The students will be delineated based on survey responses regarding their General Chemistry pedagogical experience. Students who participated in PLTL will be compared to students who did not experience PLTL at the same institution. Additionally, students who participated in PLTL will be compared to students at institutions that do not offer PLTL. The comparison groups across institutions will be stratified based on characteristics of incoming ability such as SAT sub-scores or General Chemistry final grades and on students' affect as measured by the ASCIv2 and SCI. Comparisons will likely be performed using Analysis of Variance and power analysis indicates that the planned sample will exceed the minimum sample size of 360 students needed to provide sufficient power (greater than 0.80) to identify small-medium (**f** = 0.175) effect sizes.[48] Other prevalent student survey responses, such as tutoring or joining study groups, will also be investigated for relationships to content retention.

Investigating the impact of experiences as a peer leader, will be explored in a similar fashion. Experiences as a peer leader will be determined via survey at the start of Quantitative Analysis. Similar to above, responses on CEs and scores on traditional assessments will be compared between peer leaders and non-peer leaders at the same institution and between peer leaders and a comparable group at the institutions without peer leading. The comparable group will be stratified to match incoming academic characteristics and student affect.

Additionally, the diverse range of data collected can allow subsequent nuanced investigations. One example would be in comparing the impact of serving as a peer leader versus participating as a student in a peer led class on content retention. This investigation would inform institutions regarding the essential features needed for content retention. If peer leaders exhibit much stronger concept retention than students in the course, the importance of incorporating peer leaders into structured group work would be emphasized over simply adopting instructor led group work.

The collection of data across multiple institutions also facilitates an investigation into the impact of PLTL on students' affect toward chemistry. PLTL leaders are hypothesized to show a stronger self-concept as a chemistry learner because of the additional intellectual focus on content as an instructional leader, and also a stronger positive attitude toward the subject matter. Further, affect measures in upper-level chemistry students has been under-studied and may inform future research project related to persistence or career choices of upper-level chemistry students. To explore the research question, Quantitative Analysis students' responses to the ASCIv2 and SCI will be scored for each sub-scale. The instruments will be checked for internal consistency using Cronbach's alpha for each sub-scale and comparing the values to those established in the literature.[38, 44] Confirmatory factor analysis will also be conducted to measure goodness of fit across sub-scales.

To determine the impact of enrolling in PLTL as a student or as a peer leader on student attitudes, comparisons will be conducted similar to the above analysis on content retention using the affect sub-scale scores related directly to Chemistry and Mathematics as outcome measures. These sub-scales are chosen as they've shown the most consistency in past literature and predictive performance of academic performance.[38, 43] The comparison across peer leading experiences will inform what experiences play a role in upper-level students' attitudes toward chemistry. Year three of the project will be dedicated to continuing analysis on the data collected throughout year two and the dissemination activities discussed below.

Control Study

The control study will run concurrently with the above naturalistic study and will be independent of the Quantitative Analysis setting. In Year One, students will be recruited from each research setting to participate in a project determined to measure General Chemistry knowledge at multiple points in time over a longitudinal design. Across the four research settings, 60 students (15 per setting) who completed General Chemistry without peer leading in Spring 2018 will be recruited. Additionally 30 students from USF and UNH (15 students per setting) will be recruited that were students in General Chemistry that used peer leading. Finally, 30 additional students who completed a peer leading assignment at USF and UNH will also be recruited. The resulting 120 students will be asked to complete a set of assessments related to General Chemistry knowledge.

The assessments include three CEs related to General Chemistry content. Additionally they will be asked to complete a set of three assessments adapted from the Measure of Linked Concepts format and related to General Chemistry content.[49] Measure of Linked Concepts are closed-ended versions of CEs. In a Measure of Linked Concepts, students are presented a prompt and a series of typical student responses from prior CEs. Students are asked to determine whether the statements are true or false and to rate their confidence on each one. An example Measure of Linked Concept is shown in Figure 2. Like CEs the Measure of Linked Concepts can measure students' ideas throughout the course. The example in Figure 2 focuses on chemical kinetics but also investigates student understanding of intermolecular forces, ideal gas law and Lewis structures. The confidence scale can be used to provide an indication of students' chance guessing.

Consider a reaction chamber initially with **0.10 M of C_4H_8** that undergoes the reaction: **$C_4H_8(g) \rightarrow 2\ C_2H_4(g)$** with **Rate = $1.12 * 10^{-2}\ s^{-1}\ [C_4H_8]$**
For statements 1-6 indicate whether each statement is true or false. For each statement rate your confidence in your assessment on a scale of 1 [not at all confident] to 5 [very confident]

1. After 50 seconds, the concentration of C_4H_8 will be 0.057 M. (T)
2. The initial rate of formation of C_2H_4 will be 0.00224 M/s. (T)
3. The strongest intermolecular force between the chemicals in the reaction chamber will be hydrogen bonding. (F)
4. To change the rate law to: **Rate = $1.5 * 10^{-3}\ s^{-1}\ [C_4H_8]$** will require increasing the temperature. (F)
5. If the reaction chamber volume is 25.0 L, there are 140 grams of C_4H_8 initially. (T)
6. A Lewis structure for C_2H_4 would have 12 valence electrons present. (T)

Figure 2: Example Measure of Linked Concept

Five participants will be recruited among the original set to take the assessments and follow up with a think aloud interview. In the think aloud interview, each participant will describe their reasoning for each response given. The purpose of the interview is to determine students' response processes to the assessments and to identify any areas where the assessments need to be modified. The remaining identified participants will be given the set of assessments in May of 2018 (year one) thereby occurring immediately upon the participants' completion of General Chemistry or serving as a peer leader. A member of the project team will coordinate and administer the assessments in a controlled test-like environment at each research setting. Efforts will be made to schedule participants simultaneously and participants will be instructed to work independently on the assessments. The same assessments will be given annually in May in year two and year three of the project. Participant attrition over the course of the study is expected and discussed below and will be partially mitigated through the use of compensation to participants for each annual session they attend.

The set of assessments will be scored based on each student's performance on the set of CE's and on the set of Measure of Linked Concepts. For CEs the total number of statements that satisfy the distinct, correct and relevant criteria will be tabulated. As with the naturalistic study, the project team will independently score 25 CEs and discuss discrepancies to reach consensus. The discussion will be used to clarify the scoring rubrics and two members of the project team will score the remainder of the CEs independently with discrepancies discussed by the project team. Student performance on the Measure of Linked Concepts will be measured by percent correct on each statement.

The analysis will use repeated measures analysis of variance to compare student performance across the three time points of administration and examine the interaction effects of participation in peer leading versus time. Separate analyses will be conducted on participating in peer leading as a peer leader or as a student in a peer-led class. A power analysis indicates that complete data for 60 students is needed to for sufficient power (greater than 0.80) to identify a small-medium interaction effect.[50] Estimating participant attrition is problematic given the dearth of longitudinal work in post-secondary science education. The proposed work will use a conservative estimate of 25% attrition per year which is in excess of a previous multi-institution longitudinal study.[51] Thus, the initial planned sample of 120 students would provide a resulting 60 cases of complete data for sufficient power.

This analysis will inform students' retention of concepts while controlling for the knowledge each participant demonstrated immediately following their experience as a General Chemistry student or peer leader. Content retention can then be compared directly between those students and peer leaders and between students who experienced General Chemistry with peer leading and those who had traditional instruction.

External Formative Feedback via Advisory Board

This project will benefit by formative feedback from an advisory board. The advisory board has a diverse range of expertise including experience with developing Analytical Chemistry curricula, developing chemistry assessments and evaluating PLTL. The following members have confirmed their willingness to participate as an advisory board (please see Supplemental Documents):

Patricia Mabrouk is a Professor at Northeastern University and has pertinent expertise in bioanalytical chemistry through her research on heme proteins and conducting polymers. She was editor and author of an ACS Symposium series book on problem-based learning in the analytical sciences entitled "Active Learning: Models from the Analytical Sciences" published in 2007.[52] She introduced problem-based learning in Quantitative Analysis at Northeastern University and developed a course entitled Bioanalytical Chemistry.

Regina (Gina) Frey is the Florence E. Moog Professor of STEM Education, with her primary appointment in Chemistry. She is Executive Director of The Teaching Center and co-Director of a university education-research center (CIRCLE). Since 1994, Frey has focused on the development, implementation, and evaluation of peer-led pedagogies, including PLTL, Process-Oriented Guided Inquiry Learning (POGIL), and a transition program for underprepared students in General Chemistry.[7, 53] Frey consults with universities across the country on incorporating active-learning strategies (particularly PLTL) into STEM curricula. As executive director, Frey works with faculty from all disciplines to develop, evaluate, refine, and improve their teaching; discussing strategies for incorporating new pedagogical methods and tools into their courses.

Jack Barbera is an Associate Professor of chemistry education at Portland State University where his research group focuses on the use of both qualitative and quantitative methodologies for assessment instrument development, modification, and evaluation. The Barbera research group has worked on several instruments for the assessment of student motivation, epistemological beliefs, and several content areas within chemistry.[54-60] Dr. Barbera's work using qualitative methods to provide evidence in support of the response process validity of instrument items makes him an ideal candidate to serve in an advisory role for this project.[56, 58, 59, 61]

The project team will host a virtual meeting with the advisory board every July and November throughout the duration of the project. In preparation for each meeting, the project team will provide a report detailing the activities and outcomes associated with the project since the preceding meeting. The report will also detail the future plans regarding of the project. During the virtual meeting the project team will make a brief presentation on any unresolved issues and seek the advisory board's feedback on how to proceed. Toward the end of the project, the project team will send portions of developed manuscripts intended for publication to select members of the advisory board, based on relevant expertise, for feedback.

Dissemination

The creation of student assessments for Analytical Chemistry, particularly Creative Exercises, will be of use to both instructors and education researchers who are involved in Analytical Chemistry. First, the student assessments will be disseminated to an established network of Analytical Chemistry instructors (please see supplementary letter by Juliette Lantz). Second, the assessment templates will be presented in manuscripts for journals that target audiences of post-secondary chemistry faculty such as the *Journal of Chemical Education* or *Chemistry Education Research and Practice*. Finally, a workshop for faculty will be conducted at the Biennial Conference in Chemical Education (BCCE), the largest education conference for chemistry faculty. The workshop will target measuring long-term retention of concepts with the project methodology used as an exemplar and concluding with faculty, working in groups, designing a plan for measuring long-term retention. The project team will also organize a symposium regarding PLTL at the BCCE conference, which will emphasize both the results on the impact of PLTL and a faculty discussion on how to initiate a PLTL program at audience members' home institutions

The project will also generate results regarding students' use of General Chemistry content in Analytical Chemistry and the impact of PLTL on students' long-term content retention. The former results will be of interest to Analytical Chemistry instructors and will be disseminated through both publications to journals that target faculty as mentioned above along with consideration for publishing in the topic specific journal *Analytical*

Chemistry. Additionally we would prepare a report on the strengths and weaknesses in applying content knowledge that we observed for dissemination through the aforementioned network of Analytical Chemistry instructors. Results pertaining to the impact of PLTL on long-term content retention will be of interest to education researchers, program evaluators and higher education administrators. Dissemination of these results will target chemistry education research specific journals listed above and more general journals such as *Higher Education*.

Broader Impacts

The proposed work will generate results that will inform undergraduate education in the sciences. In particular, it will explore the impact of a reform pedagogy on long-term concept retention relative to traditional lecture-based pedagogy. The results have the potential to offer a unique lens of measuring instructional effectiveness, which is beyond performance on the most proximal student assessments. The measure of instructional effectiveness, with an emphasis on long-term concept retention, can inform both post-secondary science instructors and discipline-based educational researchers. The results will be disseminated in academic journals that target both instructors and educational researchers. In addition, the project will generate a series of assessment that can be used in an upper-level Analytical Chemistry course, such as Quantitative Analysis, to explore students' use of General Chemistry content. The results and assessments generated will be of use to instructors of the course as they can serve to describe how students build on their General Chemistry content knowledge in learning Analytical Chemistry. The results and assessments will be disseminated via a network of instructors who regularly teach the course and via academic journals. Finally, the proposed work will be presented via a workshop at the Biennial Conference on Chemical Education to present a means for investigating long-term concept retention in post-secondary chemistry.

Results from Prior NSF Support

Windows on the Inquiry Classroom: Full Course, Instructor-and-Apprentice Annotated Video for Professional Development in STEM Inquiry Teaching (Christopher Bauer, DUE-1245730, 1/1/14 – 12/31/16, $199,999). **Broader Impact**: A twenty-eight-session inquiry-based interdisciplinary STEM course "Fire and Ice" addressing the perception, movement, creation, and application of heat and temperature, and historical development of these ideas, has been captured on video. It is being edited, organized, and tagged into forms which are available and searchable freely through the internet and can be downloaded (UNH Library Scholarly Repository) or streamed (YouTube) as ten-minute snapshots. One can eavesdrop on students, instructor, or graduate interns in real-time action or reflective interviews. Course materials are also available in the Repository. This course is a resource for professional development, research, high-school teacher or graduate student training, and the like. **Intellectual Merit**: Two research studies have been reported. (1) On the reflective development of graduate student interns embedded in this course based on the framework of "teacher noticing". Data included critical-time interviews, daily class debriefings, and written journals. (2) On student understanding of thermal concepts assessed using a modified and tested version of an existing instrument. Self-concept for learning and attitude toward the subject were also assessed. We found evidence of improved student understanding but no change in attitude or self-concept.

These results have been presented at two national conferences. No publications have been produced under this award currently, however the project is still ongoing.

Improving Large Lecture Gateway Chemistry Courses through Flipped Classes with Peer-Led Team Learning (Scott Lewis, DUE-1432085, 10/1/14 – 9/30/17, $329,994). **Intellectual Merit:** This project is ongoing and has resulted in the development and evaluation of a reform pedagogy designed for General Chemistry with large class sizes. The reform pedagogy combines trained peer leaders to implement active learning with a large class and the flipped class approach using videos posted online to maintain content coverage. The evaluation of the pedagogy has examined student performance on common exams and student pass rates, using class-level as the unit of analysis (each class represents one data point). Results to date show that the pedagogy is effective in promoting student achievement, with statistically significant improvements on the metric. **Broader Impacts:** To date the pedagogy has been conducted in six classes of General Chemistry II impacting approximately 1,500 students and has involved five faculty members in designing and implementing reform pedagogy. An article describing the pedagogy, practical aspects of implementation and evaluation of the impact of the pedagogy has been published.[46] This work has also been presented at a national and regional conference. The data collection associated with the project has also led to a publication on the academic performance of at-risk students in General Chemistry, in particular examining means by which they succeed.[49]

Developing a Student Centered Assessment for General Chemistry (Scott Lewis, DUE-0941976, DUE-1416006, 3/10/10 – 07/31/15, $159,450). **Intellectual Merit:** The Student Centered Assessment project has recently completed and has resulted in the development of a series of open-ended Creative Exercises assessments for General Chemistry. Work related to this project has led to presenting of evidence for validity of Creative Exercises and validity of an ACS Examinations Institute General Chemistry exam. A qualitative analysis of student responses to the assessments have shown evidence of students misinterpreting the limits of models in their attempts to link concepts throughout the course and the development of closed-ended analogs. **Broader Impacts:** Work related to this project has been presented at national and regional conferences, in a series of workshops and has been described on the Royal Society of Chemistry's *Education in Chemistry* blog. Ultimately the workshops led to 19 different science and math instructors employing the assessments in their classes. The blog discussion has led to 16 additional instructors requesting a copy of the assessments. This work has resulted in five peer-reviewed publications, two on the validity of chemistry assessments, one detailing students' efforts to link concepts in General Chemistry and one on the longitudinal impact of reform in General Chemistry on student progression throughout the chemistry curriculum; and one introducing the closed-ended form of student assessment.[14, 17, 26, 49, 62] The resulting work has also been presented six times at national conferences and numerous regional conferences and invited seminars.