2016

Undergraduate teaching assistant impact on student academic achievement.

Stephanie B. Philipp
University of Louisville

Thomas R. Tretter
University of Louisville

Christine V. Rich
University of Louisville

Original Publication Information

Philipp, Stephanie B., Thomas R. Tretter, and Christine V. Rich. "Undergraduate Teaching Assistant Impact on Student Academic Achievement." 2016. *Electronic Journal of Science Education* 20(2): 1-13.

ThinkIR Citation

Philipp, Stephanie B.; Tretter, Thomas R.; and Rich, Christine V., "Undergraduate teaching assistant impact on student academic achievement." (2016). *Faculty Scholarship.* Paper 136.
http://ir.library.louisville.edu/faculty/136

Electronic Journal of Science Education Vol. 20, No. 2 (2016)

Undergraduate teaching assistant impact on student academic achievement

Stephanie B. Philipp
University of Louisville, United States

Thomas R. Tretter
University of Louisville, United States

Christine V. Rich
University of Louisville, United States

Abstract

This study evaluated the impact that trained and supported undergraduate teaching assistants (UTAs) may have had on the academic achievement of students in the first semester of an introductory chemistry course for science and engineering majors. Framed by the concepts of Lave and Wenger's Community of Practice and Wheeler, Martin and Suls' Proxy Model of Social Comparison , the study used an untreated control group with dependent post-test only design. Covariates related to student academic achievement and contextual variables were also collected and used to build models for the final exam core outcome variable. Hierarchical linear models indicated that having a UTA gave students with above-average college GPA a statistically significant boost on final exam score. More importantly, having a UTA was associated with persistence into the next course in the two-semester introductory chemistry sequence, regardless of academic achievement.

Keywords: Undergraduate teaching assistant, hierarchical linear model; academic achievement, STEM; persistence

Please address all correspondence to: Stephanie B. Philipp, Department of Middle and Secondary Education, College of Education and Human Development, University of Louisville, Louisville, KY, stephanie.philipp@louisville.edu

Introduction

A recent National Center for Educational Statistics report on student attrition from undergraduate science, technology, engineering, and mathematics (STEM) programs showed that 48% of bachelor's degree-seeking students who entered STEM fields in U.S. colleges and universities between 2003 and 2009 left these fields by spring 2009 (Chen, 2013). This report also discussed the two paths students take when leaving STEM fields: switching to a non-STEM major or leaving college entirely: (1) Poor performance in STEM coursework relative to non-STEM coursework was strongly associated with students switching majors to non-STEM fields; (2) Poor performance in college overall was associated with dropping out of college completely.

Our university has a goal of increasing the number of STEM graduates across all science, mathematics, and engineering departments to meet growing demands for STEM professionals. Therefore, supporting improved student achievement and learning in STEM coursework is a major focus for the institution, and particularly for STEM and science education faculty.

This study reports on the results of an evaluation of an instructional program developed to engage students and support greater student learning in introductory science, engineering and mathematics courses across a large, urban, U.S. research-intensive university (Philipp, Tretter & Rich, 2016), with improved student retention of STEM majors as the major goal of the program. The program employed undergraduate teaching assistants (UTAs), who had been carefully selected, trained, and supported, to serve in a variety of instructional roles leading groups of no more than 25 students. Although the full program encompassed UTAs and accompanying faculty mentors working in nine departments across the University, this study examines the program effects in a large general chemistry course that serves as an essential gateway for most undergraduate science, engineering and mathematics majors. Other studies have reported on impacts of peer mentors on student achievement and persistence (Chapin, Wiggins, & Martin-Morris, 2014; Lewis & Lewis, 2008; Mitchell, Ippolito, & Lewis, 2012; Otero, Pollock & Finkelstein, 2010), but our study quantitatively examines the relationship between variables (academic and persistence) associated with higher final exam scores in a large-scale (n > 500 students) study using a multi-level model to explain variance in those scores.

Study Focus

The focus for this study was the impact that trained and supported UTAs had on the academic achievement of less-experienced peers with whom they worked in weekly recitation sections connected to a large, first-semester general chemistry lecture for science majors (CHEM 201). The recitations had a small grade component to encourage attendance, but attendance was not mandatory. The UTAs were undergraduate chemistry majors who had previously been successful in the general chemistry course and were selected on the basis of grades in their major coursework, faculty recommendations as good communicators, and a desire to teach. Recitation sections were designed to meet at a time separate from the faculty-led, large lectures. About 20-25 students per section met once per week for 50 minutes to practice problems, build skill sets, and further discuss concepts that had been presented in the lecture. Recitation leaders had considerable autonomy with which they could lead the session. Framed by theory, this paper describes the UTA program, analyzes nested achievement data collected from general chemistry students in recitation sections using hierarchical linear modeling, and presents conclusions that may have implications for STEM program development and undergraduate teaching methods.

Theoretical Framework

The conceptual framework underlying this study is built on the relationships between specially trained and supported UTAs and the undergraduates they teach in terms of increased academic achievement in a STEM major. The outcome variable, academic achievement, is used here not to emphasize conventional academic outcomes (memorization of facts, development of basic skills, and employment of algorithms) as defined by Cohen (1994), but because students typically need to achieve a minimum grade in hierarchical STEM coursework in order to proceed to the next course in a program of study. Moreover, grades earned in STEM classes have been shown to impact student persistence in STEM coursework (Rask, 2010; Strenta, Elliott, Adair, Matier & Scott, 1994).

Three learning theories are interconnected to form the framework of this study: Vygotsky's conception of the Zone of Proximal Development (1978), Lave and Wenger's Communities of Practice (1991), and Wheeler, Martin, and Suls' Theory of Social Comparison for the Self-Assessment of Ability (1997). These three theories connect a) the assistance needed from the UTA to support the less experienced peers' cognitive development (Vygotsky, 1978), b) the relationship between UTA and student that can help acculturate the student into the community of practice (Lave & Wenger, 1991), and c) the students' requisite for an effective proxy to most accurately predict success in a course of study (Festinger, 1954; Wheeler, Martin, & Suls, 1997).

In Vygotsky's words, the zone of proximal development (ZPD) is defined as "*the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers*. (Vygotsky, 1986, p. 86, emphasis in the original). The more capable peer supports development by interactions with the less experienced peer, e.g., asking questions, modeling, discussing ideas, structuring tasks, and helping the learner to focus on what is important in the learning (Carter & Jones, 1994). Tharp & Gallimore (1991) described working in the ZPD as instructional conversation—an intentional, skillful dialogue between teacher and student. Moreover, both more- and less-experienced peers (or teachers and students) have reported developmental growth when working in the ZPD due to the give-and-take interactions that happen between the peers (Ash & Leavitt, 2003; Jones, Rua & Carter, 1998). In our study, UTAs were guided in holding instructional conversations with their students, eliciting student thinking, and working with the students using activities that supported developmental growth in learning chemistry. Because UTAs were to be working with students in their ZPD, we would expect to see developmental growth due to organized learning activities.

Related to Vygotsky's emphasis on social interaction between learner and more experienced teacher is Lave and Wenger's theory of communities of practice (Lave & Wenger, 1991). A community of practice is a group of people who are engaged in developing their skills as they interact with each other and practice their interest. The UTAs learned to develop communities of practice with disciplinary faculty, entry-level students, and each other, all in the interest of improving chemistry learning and teaching. The faculty acted as mentors to the UTAs and the UTAs took on the role of mentors to the entry-level students in their classes. The learning that happens among the members of the community is situated in the same context in which is it applied and is accomplished by socialization, imitation, and visualization. As the pivot between faculty and student, the UTA had the important role of sharing their recent experiences in general chemistry with their students as well as feeling empathy for the faculty, whose teaching responsibilities they now shared. Often the faculty imitated the UTAs' use of learning theory and strategies developed in conjunction with science education faculty and the UTA cohort. The entry level students were welcomed into and supported by the science learning community through the UTAs' mentoring, which was projected to attend to the psychosocial needs of the students and thereby improve student persistence.

One factor that has not been investigated in the STEM student retention literature is the impact that a student's evaluation of their own capabilities may play in influencing the student's self-regulation in a STEM field of study. Because of the huge investment of time, effort and money in pursuing an undergraduate course of study in STEM, most introductory level students would want to realistically determine early in their program if they possessed the ability to earn a college diploma in the STEM field of their interest. Wheeler, Martin, and Suls (1997) describe the proxy model of social comparison for self-assessment of ability as a way that a person might compare himself or herself to another person who shares similar relevant attributes (a proxy). The proxy would be much more credible for the student (resulting in a more accurate evaluation) if the student were convinced that the proxy had succeeded on some relevant task (such as passing general chemistry) with maximal effort. For instance, UTAs would be credible STEM major proxies for students from a similar background and age if students also know that the UTAs had put forth their best effort to succeed in introductory STEM courses. If the UTA could be successful in coursework and persisting in the major, then students could reasonably estimate that they also could be successful. Students often perceive graduate students and professors as having a different background (age, country of origin, undergraduate program experiences) and a different level of effort from themselves, so that they are not perceived as very credible proxies.

Supported by these theories, a UTA program for a STEM-majors general chemistry course was devised, implemented, and evaluated for impact on the academic achievement of the entry-level students (Philipp, Tretter & Rich, 2016). The UTA program was based, in part, on what had been successful in previous programs that used more experienced undergraduates to assist less experienced undergraduates in STEM learning (e.g., Otero, Pollock, & Finkelstein, 2010).

Grounded in the theoretical framework, the following research questions were asked to explore the relationships between assistance offered by UTAs, student variables such as ACT or SAT scores and college GPA, and student academic achievement:

1. Which student variables were associated with academic achievement in CHEM 201?
2. What impact did trained and supported UTAs have on the academic achievement of undergraduates in CHEM 201?

Research Design and Methods

Quantitative data collection and analysis methods in a quasi-experimental study design were used to answer the research questions about measurable differential impact on student academic achievement and persistence related to the UTA program. A comparison group with dependent post-test only design was used. Because there was not a meaningful chemistry pre-test validated and used by the instructors for this entry-level course, the inclusion of select academic variables, such as ACT scores, number of mathematics and science high school AP courses taken, college GPA, and parental education level were used to control for any selection bias in initial student academic preparedness, experience, and ability between the treatment and comparison groups. The treatment group included approximately 300 students, each enrolled in one of fourteen weekly recitation sections led by one of the six trained and supported UTAs. The comparison group included a comparable number of students, each enrolled in one of fifteen weekly graduate teaching assistant (GTA)-led recitation sections. This resulted in recitation section sizes of approximately 20 students each; each recitation met weekly for 50 minutes, and all students also had office hours available for individual help from either UTAs and GTAs. Every UTA led at least two recitation sections for one of the four senior instructors and every GTA led five recitation sections for two of the four senior instructors. The UTA and GTA recitation sections were balanced over time of day, day of week, and the four senior chemistry instructors teaching four concurrent large lecture sections to minimize selection-history threat. The sample of students in this study did not include honors students, who were enrolled in a different course, nor did it include evening sections of general chemistry, often populated by non-traditional students having a very different college experience than the full-time students in the day sections of this course. Students did not know if they were registering for a UTA or GTA section at the time of enrollment into the course because the TA assignments were not identified to the students at registration. The demographics for both UTA (treatment) and GTA (comparison) groups are shown in Table 1. Recitation section attendance records were kept to minimize diffusion of treatment. A common cumulative multiple-choice final exam developed by the four course faculty and given to all students served as the dependent post-test. Although this exam is not a nationally validated exam, it aligned strongly with course objectives and instruction. Therefore, the final exam was not guaranteed to be a measure of what any general chemistry student should know, but was designed to measure the content knowledge of students in this particular course.

The nested nature of the data (students in recitation sections) was taken into account via two-level hierarchical linear modeling (HLM) of the student outcome (final exam score) with student level variables such as the seven variables listed in Table 1. Historical outcome data (scores from previous versions of a common final exam) from three prior years were used as an independent comparison to see if students performed statistically the same on the final exam as in previous years. The average final exam score for the cohort in this study was within the range of scores recorded from the past three years. Additionally, data on intended enrollment into the second semester of general chemistry and students' intended major were collected to evaluate student persistence in STEM majors.

Table 1

Demographics of CHEM 201 Student Sample in Each TA Group

TA Group	**n**	**Male**	**Non-white**	**Parent having any college experience**	**ACT/SAT Math Z-Score (SD)**	**College GPA (SD)**	**Number of STEM AP Courses (SD)**	**Intention to enroll in CHEM 202**
GTA	310	64%	25%	75%	1.1 (0.80)	2.81 (0.85)	0.84 (1.21)	43%
UTA	284	64%	17%	79%	1.2 (0.80)	2.86 (0.80)	1.00 (1.23)	58%

[a] Z-score calculated using ACT or SAT national means and standard deviations (2012). National Mean Math ACT (SD) =21.1 (5.3) (ACT, Inc., 2014); National Mean Math SAT = 514 (117) (College Board, 2012)

Description of UTA Treatment

The treatment of the UTAs, which included pedagogical training and support, as well as content support, was implemented by education and STEM faculty during a three-day pre-semester workshop, bimonthly seminars, and weekly planning meetings with chemistry course faculty (Philipp, Tretter & Rich, 2016). The workshop introduced learning theories and active learning strategies such as questioning approaches, metacognition, mental models and formative assessment in STEM contexts. These topics were further unpacked and reflected upon in the semester-long seminar series, encouraging the UTAs to regularly practice the strategies with their students and reflect in writing and speaking on successes and improvements. UTA reflections, observations by researchers and faculty, and student evaluations of both UTAs and GTAs were analyzed previously (Philipp, Tretter, & Rich, 2014). Results showed that: UTAs were using the strategies discussed in seminars and reflecting thoughtfully on their practice; students perceived that UTA teaching skills (e.g., TA led effective discussions) and rapport-building skills (e.g., My TA encouraged questions) were higher than traditional GTA skills; and classroom observations by researchers confirmed that UTAs were employing active learning strategies and most students were reacting positively to them. The UTAs received course credit for the workshop and seminar series and a small monthly stipend.

The six UTAs in this study ranged from second-year to fourth year chemistry majors who had earned an A in the introductory chemistry course sequence they were now teaching. Moreover, the UTAs had been recommended by professors as being particularly effective in communications skills and all expressed a strong desire to work with their less experienced peers as an instructor. The UTAs met weekly as a cohort with the lecture section professors to prepare active learning activities and discuss obstacles that students might face with that week’s material that had been introduced in the lecture sections.

The content knowledge of the UTAs and GTAs were tested at the beginning of the semester by a comprehensive multiple choice exam similar to the one that their students would receive as their final exam. The scores for both UTAs and GTAs averaged 85% of the maximum points, and while there were differences in scores between individual teaching assistants, there was no significant difference between the UTAs and the GTAs average scores. The four professors leading the lecture sections and working with all the teaching assistants were satisfied that the teaching assistants had the minimum content knowledge needed to lead recitation sections and that there were no discernible differences in previous content preparation or knowledge between each group.

Description of GTA Group

The GTA comparison group consisted of three graduate students chosen for schedule availability and need for support through a teaching assistantship, which included a stipend and tuition remission. One of the GTAs had taught recitation sections of the course previously and the other two GTAs were teaching for the first time, typical of introductory chemistry GTAs at this institution. The GTAs neither received nor sought out graduate assistant pedagogical training available at the university, and met only occasionally with chemistry faculty. Based on pre and post-content knowledge tests, UTAs and GTAs were adequately and similarly prepared for the content knowledge taught in general chemistry. GTAs were not willing to participate in reflective writing or interviews about their experience. They did allow observation of their recitation classes, where they were observed twice during the semester, showing students how to work chemistry problems in whole class and small group sessions, waiting for questions from students while students completed pencil and paper chemistry problems in small groups or individually. The rapport that students had with the GTAs was observed to be very different from the positive rapport students shared with the UTAs. Students in GTA-led sections were observed to sit further away from where the GTA was teaching than students sat in UTA-led sections. Students shared little information with the GTAs, but were more likely to be passive recipients of example algorithmic questions worked by the GTA on the board. The comparison in this study was between the student outcomes from the “business as usual” recitation section instruction with typical GTA leaders and the student outcomes from a new program of trained and supported UTAs serving as recitation leaders.

Data Analysis and Results

Student Achievement

The impact of the trained and supported UTAs on students’ final exam scores was examined through hierarchical linear modeling (HLM) analysis. HLM takes into account that outcome data from individuals in groups may not be independent (individuals in the same course section may share similarities in outcomes based on some feature of the group), resulting in a more correct Type I error rate. Additionally, HLM allowed us to model both student-level and recitation section-level data at the same time in order to investigate relationships and interactions among the variables at both levels (Raudenbush & Bryk, 2002). We were interested in knowing if the relationships between the student variables varied by context.

As shown in Table 2, from the unconditional model (no variables in the model), the intra-class correlation coefficient (ICC) was calculated to measure the variance explained by clustering students in recitation sections. The ICC is the proportion of variability in the outcome (final exam grade) that is accounted for by clustering students in recitation sections; in other words, it is the between-section variance divided by the total variance. Because the within-section variance (397.58) was much larger than the between-section (intercept) variance (19.27), the ICC was very low (0.046). Therefore clustering students in recitation sections explained little variance in final exam scores (less than 5%). There was more variability in the final exam scores between students within each recitation section than variability between sections. However, because students were clustered in separate recitation sections, we decided to continue to use HLM to model relationships between student level variables as well as between student and recitation section variables.

The effects of seven student level variables (normalized ACT/SAT score, college GPA, enrollment in second semester of general chemistry course (coded as no = 0 and yes = 1), number of STEM AP courses taken, minority student status (coded as white = 0 and non-white = 1), parent educational level (coded as no college experience = 0 and any college experience = 1), and gender (coded as male = 0 and female = 1) were considered. In addition, the effects of three recitation section-level variables were considered: UTA treatment present (coded as no = 0, yes = 1), section-mean college GPA and section-mean normalized mathematics ACT/SAT score. A full two-level model was estimated in which section-level variables were explored as predictors of student level intercepts and slopes. Because independent variables were not manipulated, the variables (called predictors) in the model are correlational, not causal, in nature. The final model reported here contains only the statistically significant model variables or those trending toward significance ($p < .1$). The set of equations for this model using Raudenbush and Bryk (2002) notation is represented by:

$$Final\ Exam\ Score_{ij} = \beta_{0j} + \beta_{1j}*(College\_GPA_{ij}) + \beta_{2j}*(Persistence_{ij}) + \beta_{3j}*(Math\ Z\text{-}Score) + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(\text{Section-mean\_}College\_GPA) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}*(TA\_Type_{j}) + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}*(\text{Secton-m}ean\_Math\_Z\text{-}Score_{j}) + u_{2j}$$

$$\beta_{3j} = \gamma_{30} + u_{3j}$$

Final Exam score was predicted by three student level variables: student GPA in college, intention to enroll in the next section of general chemistry (Persistence), and normalized ACT or SAT mathematics test score (Math z-score). Section-mean college GPA was a section level predictor of the intercept (β_{0j}). TA Type (GTA=0; UTA=1) was a section-level predictor of the college GPA slope (β_{1j}). The section-mean math z-score was a section-level predictor of the persistence slope (β_{2j}).

The models are shown in Table 2 and include only variables that are considered statistically significant ($p < .05$) or trending toward significance ($p < .1$) for the exploratory nature of model building. Retention of three student variables (college GPA, enrollment in CHEM 202, and Math ACT score) in the Level 1 model explained 54% of within-section variance. Number of STEM AP courses taken in high school, parent education, minority status, and gender did not significantly explain any variance in final exam scores. Not surprisingly, higher college GPA, higher math ACT score, and enrollment in CHEM 202 were related to higher final exam scores. Addition of the three section-level variables (section-mean college GPA, TA type, and section-mean math ACT score) explained additional between-section variance in the intercept, college GPA slopes and persistence slopes. Because we did not want to remove possible section-level variables of interest from the model prematurely, we kept two section-level variables, section-mean math score and section-mean GPA, in the model to begin explaining between-section variance, even though there is a somewhat higher probability (10% instead of 5%) that we would get those values if there were no effect in the population. We have observed other researchers retaining theoretically sound variables that are trending toward significance ($p < .1$) for HLM exploration purposes (e.g., McCoach, O'Connell & Leavitt, 2006; von Secker & Lissitz, 1999). Moreover, it may be more valuable to look at effect size to assess importance of a variable in a model. In our case, the suspect section-level variables we chose to retain helped to explain about 15% of the between-section variance in the intercept and 20% of the variance in the persistence slopes between sections, so they may be valuable for other samples in future studies.

College GPA was centered on its grand mean (2.83), and section-mean college GPA was centered on its grand mean (2.70) for ease in interpretation. Therefore, in the full model, the overall intercept, γ_{00}, (48.85) now represents the predicted final exam score for a student with a college GPA of 2.83 who does not intend to enroll in CHEM 202 next semester, who has a math z-score of 0 (e.g. ACT math score of 27), and who is in a GTA-led recitation section that has a section-mean GPA of 2.70 and a section-mean math z-score of 0 (ACT math score of 26).

Table 2

Fixed Effects Estimates (Top) and Variance-Covariance Estimates (Bottom) for Models of the Predictors of Student Achievement

Parameter	**Unconditional Model (SE)**	**Level-1 Model (SE)**	**Full Level-2 Model (SE)**
Fixed Effects			
Intercept (γ_{00})	55.58* (1.14)	49.28* (1.79)	48.85* (1.72)
Section-mean college GPA (γ_{01})	—	—	7.57† (4.11)
College GPA (γ_{10})	—	13.57* (1.53)	9.31* (1.90)
TA Type (γ_{11})	—	—	8.60* (2.70)
Persistence (γ_{20})	—	3.48* (1.70)	10.07* (1.54)
Section-mean math z-score (γ_{21})	—	—	-6.19† (3.15)
Math z-score (γ_{30})	—	3.66* (1.06)	4.02* (1.10)
Variance Estimates			
Within-section variance (σ^2)	397.58	182.54	180.91
Intercept variance (τ_{00})	19.27	48.71*	41.45*
College GPA slope variance (τ_{11})	—	47.30*	37.63*
Persistence slope variance (τ_{22})	—	33.63*	26.68†
Math Z Score slope variance (τ_{33})	—	11.80*	13.92*
ICC	0.046		

*$p < .05$ $^{\dagger}p < .1$

The effect of the recitation-level variable, recitation section mean college GPA, on the intercept can be interpreted as the effect that for every unit that the recitation section-mean College GPA increased above 2.70, the final exam score would increase by 7.57 points. Being in a recitation section led by a UTA had a positive effect on the student's College GPA slope. This translated into 8.6 additional final exam percentage points for every unit above a college GPA of 2.83 for students who are in a UTA-led recitation section. Being in a recitation section with a higher mean z-score than 0 would have a negative effect on the student's persistence slope, with all other variables held constant. This meant that the higher the section-mean math score, the less points earned on the final exam score by a student who intended to enroll in CHEM 202. Because these relationships are correlational, not causational, one could interpret this part of the model another way: for students who are in a more well-prepared recitation section (higher mean math z-score), intent to enroll in CHEM 202 was not discouraged by a slightly lower final exam score.

Results showed no overall significant difference in final exam scores between students in the treatment group (having a trained UTA) and the comparison group (having a business as usual GTA), but after controlling for all other variables in the model, there was a statistically significant positive interaction effect on exam scores between students who had a higher than average college GPA and were also in a UTA recitation section. Using the full two-level model with parameters, predictions about final exam scores can be made for a given scenario. For example, a student with an above average college GPA of 3.33, an ACT math score of 27, having a GTA, enrolled in CHEM 202 next semester and in a section having a mean college GPA of 2.70 and a mean ACT score of 26 would score a 63% on the final exam. A student with a UTA and the rest of the variables same as above would score a 68% on the final exam. This highlights the statistically significant “UTA advantage” for students with stronger college GPAs.

Conclusions and Implications

The two-level model of final exam score outcome revealed no significant difference in overall final exam scores between UTA-led and GTA-led students. A recent study where both UTAs and GTAs were given instructional preparation and support as laboratory teaching assistants reported similar results on course grades (Chapin, Wiggins & Martin-Morris, 2014). Student level variables such as amount of time spent studying, study strategies, and extra-curricular issues such as health, family, and adjustment to college life were beyond the scope of this study (and were not investigated in the Chapin et al. study either) and may have eclipsed the influences on grades due to a 50-minute once per week recitation. However, another similar study has indicated that using undergraduate peers in small-group guided inquiry-style instruction that replaced one of three lectures per week was associated with a small academic achievement increase, compared to traditional lecture-based instruction delivered by a professor (Lewis & Lewis, 2008). The context of this study differed from ours in that only one section of an introductory chemistry course was reformed and that each small-group size was much smaller (10 students per UTA) than our program requirements of 25 students per UTA. Concerned with sustainability of our program, we assigned a higher number of students per UTA in order to manage the cost of the reform and be able to justify that cost to the institution for continuation of the program after the grant period for this project ended. It is not certain that the student to UTA ratio is critical, but that could be explored in further study. Future studies may also need to address or take into account the variables outside the classroom that may affect student achievement in introductory level courses.

The model did demonstrate that having a UTA for students with above average college GPA is associated with a statistically significant boost on final exam score. This result is an indication that the UTAs are working in the zone of proximal development (Vygotsky, 1978) of students who are performing well in college. In other words, students who successfully had transitioned to college were able to take advantage of trained UTA support including higher order questioning, metacognitive strategies, and feedback from formative assessment in order to significantly increase their final exam score in chemistry. Future programs may need to be developed at the departmental level to address improved learning for students who are struggling academically. These programs may use additional pedagogical strategies known to increase student learning, but may also address extra-curricular or psychosocial issues that hinder student learning and achievement in the transition to college, such as time management, relationship counseling, and health guidance.

One student variable that was much lower for students in GTA recitations compared to those in UTA recitations was the intention to persist in a STEM major by enrolling in the second semester of general chemistry, CHEM 202. A chi square test confirmed that proportionally more students required to take CHEM 202 who had UTAs as recitation section leaders enrolled in CHEM 202 than did those who had GTA-led recitation sections: $\chi^2(1, N = 343) = 12.07$, $p = .001$). This second course is required for many STEM majors such as chemistry, biology, physics, and engineering disciplines such as chemical, industrial, mechanical, and bioengineering. Further examination of this disparity between the two groups is warranted and may elucidate factors about the UTA program that are helpful for improving persistence in STEM study.

There are several possible ways to investigate strengthening STEM student retention when conducting a semester-long examination of students in a retention improvement program. The focus of this study was to investigate grades, which are important for student progression in a STEM program. An equally or perhaps even more relevant consideration might be to examine enrollment in the next course as a measure of persistence (Barlow & Villarejo, 2004; Mitchell, Ippolito, & Lewis, 2012; Wamser, 2006), particularly if that enrollment could be increased among those students who achieved at least adequate grades in the first course. While our university does not require or even encourage students to declare a major during the first year of college, our students who intend to major in a STEM discipline usually enroll in required two-course sequences like CHEM 201 and CHEM 202 in their first or second year at university and take these courses in consecutive semesters. Given the hierarchical and structured nature of most STEM degrees, there is little room for deviating from the sequential path of prerequisite foundational courses (such as CHEM 201 and CHEM 202) if one is to stay on track to graduate with the degree in 4 years. Departure from the course sequence after the first semester usually indicates departure from a STEM major intention at our university. For increased retention of students in STEM programs, it is important for students to persist into that second course. The trained UTAs, with their recent experience in the course and their relationship as an intermediary between faculty and students, seemed to create an effective community of practice (Lave & Wenger, 1991) that welcomed the introductory level students into the STEM disciplines and encouraged a higher percentage of introductory students to persist into the second general chemistry course, regardless of grade performance compared to students with traditional GTA-led recitations. Further examination of how the UTAs created this community is a topic of an additional study.

In analyzing the existing data, another factor emerged that hinted at an influence by UTAs on the development or reinforcement of a sense of science identity in their students. This could be an indication of the UTA's successful function as a proxy model of social comparison (Wheeler, Martin, & Suls, 1997). Future work will include analyzing student data describing TA quality and student science identity specifically to investigate possible relationships between student persistence and the pertinent practices of the trained and supported UTAs.

Acknowledgments

The UTA training and support program and this study were funded by the National Science Foundation under grant #106830 from the Division of Undergraduate Education.

References

ACT, Inc. (2014). *2012 ACT National and State scores.* Retrieved from: http://www.act.org/newsroom/data/2012/states.html

Ash, D., & Levitt, K. (2003). Working within the zone of proximal development: Formative assessment as professional development. *Journal of Science Teacher Education, 14*, 23-48.

Barlow, A. E., & Villarejo, M. (2004). Making a difference for minorities: Evaluation of an educational enrichment program. *Journal of Research in Science Teaching, 41*(9), 861-881.

Carter, G., & Jones, M. G. (1994). Relationship between ability-paired interactions and the development of fifth graders' concepts of balance. *Journal of Research in Science Teaching, 31*, 847-856.

Chapin, H. C., Wiggins, B. L., & Martin-Morris, L. E. (2014). Undergraduate science learners show comparable outcomes whether taught by undergraduate or graduate teaching assistants. *Journal of College Science Teaching, 44*, 90-99.

Chen, X. (2013). STEM Attrition: College Students’ Paths Into and Out of STEM Fields (NCES 2014-001). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*, 1-35.

College Board. (2012, September 24). 2012 College-bound seniors total group profile report. Retrieved from http://media.collegeboard.com/digitalServices/pdf/research/TotalGroup-2012.pdf

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*, 117-140.

Jones, M. G., Rua, M. J., & Carter, G. (1998). Science teachers' conceptual growth within Vygotsky's zone of proximal development. *Journal of Research in Science Teaching, 35*, 967-985.

Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Lewis, S. E., & Lewis, J. E. (2008). Seeking effectiveness and equity in a large college chemistry course: An HLM investigation of peer-led guided inquiry. *Journal of Research in Science Teaching*, *45*, 794-811.

McCoach, D. B., O'Connell, A. A., & Levitt, H. (2006). Ability grouping across kindergarten using an early childhood longitudinal study. *Journal of Educational Research*, *99*(6), 339-346.

Mitchell, Y. D., Ippolito, J., & Lewis, S. E. (2012). Evaluating peer-led team learning across the two semester general chemistry sequence. *Chemistry Education Research and Practice*, *13*(3), 378-383.

Otero, V., Pollock, S., & Finkelstein, N. (2010). A physics department’s role in preparing physics teachers: The Colorado learning assistant model. *American Journal of Physics*, *78*, 1218.

Philipp, S.B., Tretter, T.R., & Rich, C.V. (2014). *Strengthening STEM performance and persistence: Influence of undergraduate teaching assistants on entry-level STEM students.* Proceedings from the annual National Association of Research in Science

Teaching conference, Pittsburgh, PA. https://www.narst.org/annualconference/2014 conference.cfm

Philipp, S.B., Tretter, T.R., & Rich, C.V. (2016). Development of undergraduate teaching assistants as effective peer mentors in STEM courses. *Journal of College Science Teaching, 45*(3), 74-82.

Rask, K. (2010). *Attrition in STEM fields at a liberal arts college: The importance of grades and pre-collegiate preferen*ces [Electronic version]. Retrieved November 1, 2012, from Cornell University, School of Industrial and Labor Relations site: http://digitalcommons.ilr.cornell.edu/workingpapers

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods, (2nd Ed.)*. Newbury Park, CA: Sage.

Strenta, A. C., Elliott, R., Adair, R., Matier, M., & Scott, J. (1994). Choosing and leaving science in highly selective institutions. *Research in Higher Education 35*, 513-547.

Tharp, R. G., & Gallimore, R. (1991). *The instructional conversation: Teaching and learning in social activity*. Center for Research on Education, Diversity & Excellence. Retrieved from: escholarship.org/uc/item/5th0939d

Von Secker, C. E., & Lissitz, R. W. (1999). Estimating the impact of instructional practices on student achievement in science. *Journal of research in science teaching*, *36*(10), 1110-1126.

Vygotsky, L. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (1986). *Thought and language* (rev. ed.). Cambridge, MA: MIT Press.

Wamser, C.C. (2006). Peer-led team learning in organic chemistry: Effects on student performance, success, and persistence in the course. *Journal of Chemical Education, 83*, 1562-1566.

Wheeler, L., Martin, R., & Suls, J. (1997). The proxy model of social comparison for self-assessment of ability. *Personality and Social Psychology Review*, *1*, 54-64.