STRENGTHENING STEM PERFORMANCE AND PERSISTENCE:

INFLUENCE OF UNDERGRADUATE TEACHING ASSISTANTS

ON ENTRY-LEVEL STEM STUDENTS

By

Stephanie B. Philipp
B.S., University of Florida, 1988
M.S., University of North Carolina at Chapel Hill, 1990

A Dissertation
Submitted to the Faculty of the
College of Education and Human Development
of the University of Louisville
In Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

College of Education and Human Development
University of Louisville
Louisville, Kentucky

August 2013

STRENGTHENING STEM PERFORMANCE AND PERSISTENCE:

INFLUENCE OF UNDERGRADUATE TEACHING ASSISTANTS

ON ENTRY-LEVEL STEM STUDENTS

By

Stephanie B Philipp
B.S., University of Florida, 1988
M.S., University of North Carolina at Chapel Hill, 1990

A Dissertation Approved on

June 27, 2013

By the following Dissertation Committee:

____

Thomas R. Tretter, Dissertation Director

____

Robert Ronau

____

Sherri Brown

____

Christine Rich

____

Melissa Shirley

DEDICATION

This dissertation is dedicated to my husband, Craig,
for his unwavering belief in me and complete support of all my professional endeavors.

ACKNOWLEDGEMENTS

I would like to thank my dissertation advisor, Dr. Tom Tretter, for his patient guidance, enthusiastic encouragement, and wise counsel. I would also like to thank my dissertation committee; I have learned invaluable lessons from each one of you that I will do my best to pass on to my future students. I thank my children, Susannah, Kelly, and Alex, who have inspired me to be a life-long learner. Their enduring patience with Mom's incessant planning, writing, and studying over the last four years has not gone unnoticed. I would like to thank my family and friends: my parents, John and Marge Boggess for their reassurance and good-natured acceptance of a daughter who seems to be a professional student; my mother-in-law, Dixie Philipp, who celebrates each milestone with me; my sister, Kim, and sisters-in law Tina and Sandy and their husbands, special friends Peter, Shelley, Kimberly, and Stephanie and future colleagues Jill and Ingrid, and so many of the supportive faculty and staff of the College of Education and Human Development, who have all cheered me on each step of the dissertation journey. Lastly, I would like to thank the undergraduate teaching assistants and CHEM 201 instructors who willingly participated in this project and were so generous with their time and energy.

ABSTRACT

STRENGTHENING STEM PERFORMANCE AND PERSISTENCE:

INFLUENCE OF UNDERGRADUATE TEACHING ASSISTANTS

ON ENTRY-LEVEL STEM STUDENTS

Stephanie B. Philipp

June 27, 2013

Increasing retention of students in science, technology, engineering, or mathematics (STEM) programs of study is a priority for many colleges and universities. This study examines an undergraduate teaching assistant (UTA) program implemented in a general chemistry course for STEM majors to provide peer learning assistance to entry-level students. This study measured the content knowledge growth of UTAs compared to traditional graduate teaching assistants (GTAs) over the semester, and described the development of peer learning assistance skills of the UTAs as an outcome of semester-long training and support from both science education and STEM faculty. Impact of the UTA program on final exam grades, persistence of students to enroll in the next chemistry course required by their intended major, and STEM identity of students were estimated. The study sample comprised 284 students in 14 general chemistry recitation sections led by six UTAs and 310 students in 15 general chemistry recitation sections led by three traditional GTAs for comparison.

Results suggested that both UTAs and GTAs made significant learning gains in general chemistry content knowledge, and there was no significant difference in content

knowledge between UTA and GTA groups. Student evaluations, researcher observations, and chemistry faculty comments confirm UTAs were using the learning strategies discussed in the semester-long training program. UTA-led students rated their TAs significantly higher in teaching quality and student care and encouragement, which correlated with stronger STEM recognition by those students.

The results of hierarchical linear model (HLM) analysis showed little variance in final exam grades explained by section-level variables; most variance was explained by student-level variables: mathematics ACT score, college GPA, and intention to enroll in the next general chemistry course. Students having higher college GPAs were helped more by having a UTA. Results from logistic regression of persistence outcome variable showed that students are three times more likely to persist to CHEM 202 if they had a UTA in CHEM 201. Other positive predictors of retention included having strong college grades, and having strong ACT math scores. Coupled with HLM analysis result that UTAs were more effective at helping students with higher college GPAs achieve higher grades, the stronger persistence of UTA-led students showed that the UTA program is an effective program for retention of introductory-level students in STEM majors.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

BACKGROUND AND PROBLEM STATEMENT

A workforce educated in science, technology, engineering, or mathematics (STEM) is vitally important to creating new jobs, increasing competitiveness in the global economy, creating solutions to problems that plague our society, and educating our next generation of STEM professionals. Members of the 2005 "Rising Above a Gathering Storm" Report Committee recalled:

> The *Gathering Storm* committee concluded that a primary driver of the future economy and concomitant creation of jobs will be innovation, largely derived from advances in science and engineering. While only 4 percent of the nation's workforce is composed of scientists and engineers, this group disproportionately creates jobs for the other 96 percent. (National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, 2011, p. 4)

Government, education, and business groups are concerned that the quantity of college graduates with STEM majors is not sufficient to remain economically competitive with the rest of the world (Augustine, 2005; Business and Higher Education Forum, 2007, 2010; George, 1996; Kentucky Council on Postsecondary Education, 2007; National Academy of Sciences, 1999). Only 15.6% of U.S. college graduates earned degrees in STEM majors in 2007 (Business and Higher Education Forum, 2010). As reported by the Business and Higher Education Forum (2010), there were 3.8 million 9^{th} graders in the U.S. in 1997, according to the National Center for Educational Statistics (2009). By 2001, 2.7 million of those 9^{th} graders had graduated from high school and 1.7

million of those high school graduates chose to enter a 2 or 4 year college. Six years later, a total of 233,000 students earned bachelor's degrees in STEM majors (National Science Board, 2010). Moreover, the National Center for Educational Statistics (2009) estimates that half of the undergraduates who intend to major in STEM fields or who declare a STEM major switch to a non-STEM major between high school and college graduation.

Although some research concerning retention of undergraduate STEM majors was conducted before 1990 (Berryman, 1983; Hilton & Lee, 1988, Ware & Dill, 1986), many more studies since then have examined what kinds of students choose STEM majors, the characteristics of students who persist and who leave STEM majors, and reasons why undergraduates leave STEM majors and careers (e.g., Brainard & Carlin, 1998; Carlone & Johnson, 2007; Seymour & Hewitt, 1997; Swarat, Drane, Smith, Light, & Pinto, 2004; Tobias, 1990). Particularly notable findings that are common to many of the studies is that students, regardless of whether they leave or stay in a STEM program, find the teaching practices of faculty in their undergraduate STEM classes to be unhelpful for learning, the workload to be much greater than for their non-STEM peers, and non-STEM fields to be a more attractive career choice.

The increase in studies conducted on undergraduate STEM retention after 1990 may coincide with the creation of Project 2061 from the American Association for the Advancement of Science (AAAS), and the release of the report, *Science for All Americans* (Rutherford & Ahlgren, 1990). This report called for college and university mathematics and science departments, as well as education departments, to consider the guidelines proposed to increase science literacy in K-12 education as a basis for designing their own curricula for future math and science teachers. Scientific literacy is defined here as having the necessary knowledge about STEM concepts to participate in a democratic society in which decisions are made by its citizens involving scientific and technical issues. Although this level of scientific literacy does not require a college education, K-12 teachers, who take undergraduate level STEM courses, are instrumental in educating all K-12 students in fundamental STEM concepts and processes. Thus, the quality of undergraduate STEM education affects not only STEM majors, but also impacts the quality of K-12 teachers who are responsible for educating our country's youth.

In summary, the US is facing three problems in relation to undergraduate science education: (1) more STEM-proficient workers are needed for continuing global economic competition, (2) efforts to attract and retain students in college STEM programs are not as successful as desired, and (3) reform of undergraduate STEM education to attract and retain students and increase student achievement is a complicated process that involves all the stakeholders in post-secondary education: institutions and administrators, STEM faculty, business and government employers of STEM workers, and students interested in STEM programs of study.

Study Purpose

An NSF-funded project, Partnership for Retention Improvement in Mathematics, Engineering and Science (PRIMES), aimed at increasing retention in STEM programs through modified instructional methods in STEM courses, was implemented at the University of Louisville starting January 2012. This dissertation study was based on the PRIMES implementation, and focuses on qualitative and quantitative characteristics of the three core groups involved in the implementation project: undergraduates in introductory STEM courses, undergraduate teaching assistants (UTAs) assigned to small groups of these undergraduates, and STEM faculty who supervise the undergraduate teaching assistants and teach entry-level STEM students. The data collected from this implementation project were quantitative (e.g., test scores, and survey results) as well as qualitative (e.g., classroom observations, interviews, and artifacts, such as lesson plans, assignment materials, and assessments). The number of undergraduates of interest in this study, both first year students and their UTAs, provided enough power to estimate a two-level hierarchical linear model (HLM) for examining the factors that may impact undergraduate STEM academic achievement.

Previous research has shown that achievement and persistence in undergraduate science, technology, engineering, and mathematics (STEM) courses is influenced by the involvement or engagement of students in the first years of college (Tinto, 2001; Tobias, 1990) and that faculty action in the classroom make a difference in student achievement and retention (Braxton, Bray, & Burger, 2000). What is not well-known is how to engage different types of students in various settings, institutions, and academic disciplines. We know something about why students leave (Seymour & Hewitt, 1997; Strenta, Elliott, Adair, Matier, & Scott, 1994), but that does not automatically translate into what institutions, faculty, and students need to do to increase achievement and persistence. A model that describes not only effective programs for students but also institutional support for those who enact these programs is needed (Tinto, 2006).

The general purpose of this study is to investigate the impact that trained and supported UTAs as peer educators in a general chemistry course may have on student achievement, identity of students as STEM students, and the benefits incurred by the UTAs, who are at the heart of the PRIMES program. The UTAs received intensive pedagogy training and STEM faculty mentoring for their UTA role as a peer learning assistant for small groups of entry-level students enrolled in a general chemistry course. The specific purposes of this study are three-fold: 1) to describe the ways in which six UTAs deepened their content knowledge and used newly acquired pedagogical strategies for their own learning; 2) to explore how the UTAs' skills needed for effective peer learning assistance changed as a result of the UTA experience of pedagogical training, chemistry content support, and working with less-experienced peers, and 3) to determine the impact that UTAs had on the academic achievement and STEM identity of undergraduates.

Study Significance

This study will contribute new insights into the experiences of undergraduates considering STEM majors at a large research university and the impacts made by UTAs on the academic achievement of undergraduates in an introductory chemistry course. This study will also examine the development of the UTAs as peer learning assistants and the benefits acquired by the UTAs as a result of their teaching experience. The impact that the UTA program may have on the attitudes of the chemistry faculty involved as UTA mentors will also be described.

Study Limitations

Choosing and persisting in a college major is a highly personal decision made by a college student influenced by life experiences, relationships, aspirations, financial goals, and numerous other factors. Previous studies have examined students' personal choices of college major using ethnographic methods, with extensive surveys and interviews (Seymour & Hewitt, 1997; Strenta et al., 1994). This study is not designed to unpack that level of detail at the undergraduate level. Rather, this study will rely on group aggregates to make inferences about the collective set of individual experiences. The data collected in this study were from the second semester of the PRIMES project implementation; data from the first semester were used to pilot both implementation and data collection.

Research Questions

Grounded in the conceptual framework and theoretical foundation described in Chapter 2, the following research questions were asked to explore the relationships between peer learning assistance, student academic achievement and STEM identity:

1. How did the UTAs change as a scholar as a result of the UTA experience?
    a. In what ways did disciplinary content knowledge deepen?
    b. Which learning approaches did UTAs mention as aiding their own learning?
2. How did the UTAs' peer learning assistance skills develop over the semester(s) they were a UTA?
3. What impact did trained and supported UTAs have on the academic achievement of undergraduates in an introductory chemistry course?
4. How did the UTAs influence STEM identity development of undergraduates in the introductory chemistry course?

The order in which these questions were asked was intentional. The answers to the first two research questions were used to help explain or give context to the answers for the last two questions. If positive content knowledge and peer learning assistance growth were not observed for the UTAs, then the impact that UTAs would have on the achievement of their students would probably be very different from the impact that more knowledgeable and skillful UTAs would have on their students. Without answers to Research Questions 1 and 2, I would have little idea to what extent the UTAs were implementing the training they had received and if they perceived support for their peer learning assistance tasks.

Summary

This chapter outlined the background for this study including a problem statement, purpose of the study, study significance, and study limitations. Chapter 2 will discuss current literature concerning the state of STEM education, retention and attrition of STEM majors, and current learning undergraduate teaching assistant programs. A discussion of the conceptual framework for the study will also be presented. Chapter 3 will explain the methodology used in this study, describing the research design and the specific procedures used to collect the data. Chapter 4 will report the analysis of the data and results of this study. Chapter 5 will present the conclusions drawn from the study results, discuss implications for action, and will put forth recommendations for further research.

CHAPTER 2

REVIEW OF LITERATURE

This chapter comprises five main sections that summarize the theoretical and empirical knowledge base regarding the use of undergraduate teaching assistant programs that have been developed to support undergraduate STEM learning and increased retention. The first section describes the challenge for colleges and universities to increase the rate of retention for students in STEM majors and the motivation for increasing retention in STEM programs from initiatives proposed by both government and industry. The second section describes trends in STEM persistence and attrition in U.S. colleges and universities. The third section will review the research on the factors that may discourage many able and qualified students from persisting in their intended STEM programs and factors that may support student persistence in STEM courses. The fourth section will present the conceptual framework and identify and summarize research studies that have addressed the three interconnected concepts that frame this study: the use of undergraduate teaching assistants; academic achievement of undergraduates in STEM coursework; and development of STEM identity in undergraduate students. The fifth section will describe the UTA program implemented in this study that was modeled after programs that have shown evidence of success for increased academic achievement and strengthened STEM identity of undergraduate students. This section will conclude with a summary of the variables generated from this review of the literature that will be used in this study.

The Challenge to Increase Student Retention in STEM Majors

For the past several decades, national science organizations, industrial groups, and government committees have been warning of shortages of scientists and engineers who have the skills and knowledge to tackle our society's complex problems, such as energy supply and demand, disease control and eradication, materials development, environmental protection issues, and technology innovations (e.g., Grice, Peer, & Morris, 2011; National Academy of Sciences, 2007; U.S. Congress, 1990). In 1990, the U.S. Congressional Subcommittee on Science, Technology and Space met to discuss the projected shortfall of scientists and engineers needed to meet the scientific and technological challenges of the twenty-first century (U.S. Congress, 1990). This shortfall, according to the committee chairman and testimony from expert witnesses, would happen because fewer undergraduates were interested in or prepared for earning bachelor's degrees in STEM majors, coinciding with large numbers of expected retirements in research universities and federal agencies responsible for science and technology activity. The National Defense Research Institute also reported that, although the U. S. leads the world in many aspects of science and technology, persistent underperformance of U.S high school students on international math assessments and the observation that science and engineering careers have a limited attractiveness to U.S. students (Galama & Hosek, 2008) results in lower numbers of U.S.-born STEM graduates.

In response to the predicted shortfall of scientists and engineers, in 1990, the U.S. government promised to dramatically increase the National Science Foundation's budget for educational initiatives to fund STEM education programs that would help attract and retain undergraduate and graduate STEM students. Currently, the National Science Foundation directs the Science, Technology, Engineering, and Mathematics Talent Expansion Program (STEP), which seeks to increase the number of U.S. students receiving baccalaureate degrees in established or emerging fields within STEM. From the private sector, the Business-Higher Education Forum's Securing America's Leadership in STEM Initiative has a goal of doubling the number of STEM graduates from 2005 to 2015 by using a simulation model to find the "leaks" in the STEM pipeline through college (Business-Higher Education Forum, 2010).

Economic Pressures to Increase STEM Graduates

The current economic and political climates, as well as basic societal needs, promote the demand for STEM professionals. Many examples of this demand exist. Growing energy requirements, potable water obtainability, and reaction to recent natural disasters have increased the demand for geoscience professionals beyond the current supply (Gonzales & Keane, 2010). In power engineering, an aging workforce, where 45% of the professionals could retire in the next ten years, coupled with fewer college graduates in electrical engineering, could result in a shortage of qualified personnel to design, operate, and maintain electrical power systems (Grice et al., 2011).

Push for More Diversity in STEM Programs and Workplaces

Although non-white population groups make up almost 30% of the U.S. population as of 2006, they constitute only 9% of college-educated scientists and engineers in the U.S. (National Research Council, 2011). Moreover, minority groups that are the fastest growing segments of the population are the most underrepresented in science and engineering occupations (National Research Council, 2011). Participation rates in science and engineering occupations by gender are also differentiated: in some disciplines such as biology, women are over-represented, but in fields such as engineering and computer science, women participate at much lower rates than men. Scientific research and engineering project priorities are developed, funded, and supported by the people who are involved in the work. If the science and engineering workforce is more diverse, then the problems chosen to be tackled and the solutions recommended will also be more diverse. Increasing retention of all students interested in STEM majors could help increase the diversity of the STEM workforce. Although underrepresentation in STEM courses by well-prepared minority groups has been observed at many universities, research-based programs have been developed to increase equity in representation (Ong, Wright, Espinosa, & Orfield, 2011; Treisman, 1985, 1992).

Trends in STEM Persistence and Attrition

Discovering what is now known about graduation rates in colleges and universities is a logical place to start investigating the retention rate of STEM students. The National Center for Education Statistics (Horn & Berger, 2004) reported that between 1989 and 1995, total undergraduate fall enrollment in institutions of higher education increased from 11.7 million to 12.2 million. The completion rate for a bachelor’s degree in any program five years after starting postsecondary education was 53.3 percent for the 1989-1990 cohort. The study also showed that there was little change in bachelor’s degree completion rate (53.4 percent) from 1994 through 2000. The study did find a statistically significant increase in 5-year persistence rates for the cohort of students followed starting in 1995-1996. That is, a higher percentage of students who started post-secondary studies in the 1995-1996 cohort were still in college after 5 years

(17.2 percent), not having completed a bachelor's degree, as compared to the 1989/1990 cohort (13.3 percent). More recently, the National Center for Education Statistics (Aud et al., 2010) reported that undergraduate enrollment in postsecondary institutions increased to 16.4 million students from 2000 to 2008, and is expected to reach 19.0 million students by 2019. About 57 percent of first-time students seeking a bachelor's degree or its equivalent and attending a 4-year institution full time in 2001–02 completed a bachelor's degree or its equivalent at that institution within 6 years (Aud et al., 2010).

Based on the 1996 Beginning Postsecondary Students Longitudinal Study (Berkner, He, & Cataldi, 2002), 23 percent of beginning postsecondary students intended to major in a STEM field at some time during their postsecondary enrollment from 1995–96 to 2001. Of those students intending to major in STEM field in the 1995-1996 cohort, 34.8 percent actually earned a bachelor's degree and 18.6 percent were still enrolled as a STEM major by 2001. The number of mathematics, engineering and natural science bachelor's degrees awarded (except in biological science) decreased from 1985 to 2000 (National Academy of Sciences, 2007). The number of STEM graduates increased about 5% from 2003 to 2007, but the total number of bachelor's degrees awarded during that time period increased over 13 percent. In 2007, there were about 233,000 STEM graduates in the U.S., who earned 15.6% of all bachelor's degrees awarded that year (National Science Board, 2010).

Factors Impacting STEM Retention

Several studies have studied the characteristics of students who do not persist in STEM majors or have sought to discover the causes of attrition from undergraduate STEM courses (DeBoer, 1984; Hilton & Lee, 1988; Seymour & Hewitt, 1997; Strenta, Elliott, Adair, Matier, & Scott, 1994; Tobias, 1990). Knowledge of these factors may help to design an effective program for improving STEM major persistence.

Feelings of Science Competency

DeBoer (1984) surveyed freshman students after they received their first semester grades in science courses. He found that the intention to continue in STEM was positively associated with ratings of personal ability and negatively associated with the ease of the science courses for the "successful" students. In other words, students who believed they had strong ability in science, and that their science courses were challenging, were more likely to report plans for enrolling in advanced STEM coursework. Factors contributing to the students' sense of competence in science were not explored.

First-Year College Experiences

While Hilton and Lee (1988) were not able to infer the causes of attrition from STEM programs from their longitudinal study, they did find that the biggest net loss of students from interest in STEM majors occurred between the end of high school and the start of college. However, the pool of students who lost STEM interest included high school students who intended to major in a college STEM program but changed their mind by the start of college as well as those who intended a STEM major but who did not attend college at all. The second biggest net loss of students from STEM majors happened during the first year of college coursework, which includes introductory-level STEM courses required to continue progress toward a STEM major. Because most post-secondary institutions do not require nor encourage students to declare a major during the first year of study, researchers find it can be difficult to know if a student actually switched from a STEM major to a non-STEM major during this time period. However, students who are seriously intending on majoring in a STEM field must enroll in introductory-level STEM courses soon after entering university because of the hierarchical nature of STEM courses. Thus, a good proxy for a declared STEM major during the first year of college is enrollment in introductory science and math courses specifically required for STEM majors. Students enrolled in these types of courses who do not enroll in the subsequent STEM courses the next semester may be considered as departed from a STEM major. Moreover, students who graduated with a STEM degree had more likely persisted in STEM since the start of college; any movement of students into STEM programs from non-STEM programs during college was found to be negligible. This suggests that strategies to retain declared or intended college STEM majors during their first year of college and introductory-level courses that are gateways to STEM major programs are more likely to have an impact than strategies to recruit non-STEM college students into STEM programs.

Grading and Learning Environment

Strenta et al. (1994) surveyed thousands of men and women at four selective universities who were initially well-prepared for study in STEM fields, having higher than average high school math and science grades and high math SAT scores. The attrition rate for this capable group of students initially interested in STEM was 40 percent after four years of college. Strenta et al. found that the most significant cognitive factor for predicting students leaving STEM programs was low grades in STEM courses earned during the first two years of college. At the same time, students in science courses geared towards non-STEM majors earned significantly higher grades in those courses than did students in science courses required for STEM majors. This can be a disincentive to students, especially if the students are focused on college GPA as an indication of how well they are performing or as a condition to continue college enrollment. For example, many scholarship programs require students to keep a minimum GPA; if taking a STEM-required science or math course may result in a lower grade than taking a science course for non-majors, this serves as a disincentive for the student to remain a STEM major and risk losing the scholarship.

Students also expressed dissatisfaction with the highly competitive nature of STEM classes and the feeling that asking questions in STEM classes was not as welcome as it was in non-STEM classes. Students found STEM courses to be duller than courses in non-STEM fields, and introductory courses duller than advanced level courses in the same field. These findings seem to support those from Seymour and Hewitt (1991) and Tobias (1990). What Strenta and colleagues did not find was any indication of students experiencing overt or covert discrimination, sometimes called a "chilly climate" towards women or non-white students. As in Hilton and Lee (1988), positive net attrition from STEM majors to either non-STEM fields such as social sciences or humanities or a termination of college studies occurred during the first years of college. Rarely did students switch from a non-STEM program to a STEM program. Students in Strenta et al.'s study gave two other reasons (besides grades) for switching from STEM to non-STEM areas: (a) other fields of study were more interesting to them or (b) other fields of study made better use of their talents. These findings from Strenta et al. suggest that students favor STEM classes where questions and discussion of concepts are welcome, learning approaches are active and engaging and contribute to increased student competency. Furthermore, students need to know that hard work they are doing to prepare for their STEM classes is valued by their instructor and will be rewarded in the future with opportunities for meaningful careers.

Institutional Policy and Structure

Henderson, Beach and Finkelstein (2011) performed an analytic review of journal articles published between 1995 and 2008 that reported research on strategies used to reform instruction in undergraduate STEM courses. Successful programs to change instruction that were associated with higher student achievement included one or more of the following components supported by institutional structure: programs that were focused on one aspect of pedagogy or curriculum and lasted an extended period of time; use of performance evaluations and feedback to the participants in the program; programs that allowed for practice of new concepts and skills and reflection on that practice; and individualized solutions that aligned with cultural and organizational norms.

STEM Interest

Maltese and Tai (2011) completed a logistic regression analysis of factors that predict completion of a STEM bachelor's degree using data from the National Educational Longitudinal Study of 1988. The results indicated that the most of the students who graduate from college with STEM degrees made the choice to study a STEM field during high school, and that choice was related to a student's growing interest in mathematics and science rather than from taking advanced science course or earning higher grades in science courses. Significant predictors from high school experiences were the number of science classes taken during high school, but grades earned or academic level of those courses were not significant predictors. Students who completed more STEM credits in their first year of postsecondary study and those who earned higher grades in college than their peers were more likely to go on to earn a STEM degree, confirming previous studies (Hilton & Lee, 1988; Strenta et al.(1994). Another finding was that students involved in loan programs or work-study were no less likely to earn a STEM degree, meaning that access to a STEM degree is open to students regardless of financial situation.

Summary of Factors Impacting Retention

Improving retention of students in STEM majors has been an on-going interest for the past thirty years. Factors for improving STEM persistence and reducing attrition have been determined and confirmed from numerous studies. Factors that are worth consideration when planning a STEM retention improvement program include: encouraging feelings of science and math competency and foster growing interests in science and math; focusing on positive learning experiences in first-year or introductory-level STEM courses; combining introductory students with successful STEM students who can offer advice and encouragement about building a meaningful career using their STEM skills and knowledge; and obtaining institutional support that is necessary for effective long-term programs that use reflection on practice and regular evaluation for constant improvement. The next section describes STEM retention improvement programs that have used undergraduate teaching assistants (UTAs) to work with introductory level students in various capacities.

Undergraduate Teaching Assistant Programs

Undergraduate teaching assistants have been used in many roles for assisting undergraduate science teaching and learning by promoting active and collaborative learning, and to act as an intermediary between a course professor and the students enrolled in that course. Often, UTAs provide secondary instruction (discussion, recitation, or problem-solving sessions), while a faculty member or graduate teaching assistant provide the primary form of instruction (usually lecture format).

Treisman Model

One successful program for increased retention of students in science and math is the Treisman model that has been replicated in math and science classrooms at several universities (e.g., Conciatore, 1990; Swarat, Drane, Smith, Light & Pinto, 2004; Treisman, 1985). This model has the following components: (a) cooperative small group learning; (b) groups led by a more experienced undergraduate student; and (c) problem-based learning. Small cooperative learning groups allow students to learn from each other, to express their ideas in an environment that feels socially safer than a large lecture class, and more closely models the type of environment in which scientists work than does studying in isolation or passively listening to a lecture. Problem-based learning begins with an engaging, conceptually-based problem that is connected to the student's previous understandings. Solving the problem compels students to learn some new knowledge that is related to what they already know, so students take an active role in learning.

Using Treisman's model, Swarat et al. (2004) showed positive effects on retention of students in introductory biology, calculus-based physics, non-calculus based physics, and chemistry courses that were pre-requisites or gateway courses for declaring a science major. Retention was defined as successfully completing all three quarters of these gateway courses. Retention rates were significantly higher for participants who volunteered to be in the model program than for non-participants of equal academic ability. Differences between participants and non-participants in motivation and interest were not accounted for and could have possible provided insight into at least some of the increased retention rates. Factors found to be important in the implementation of the model were: recruitment of all students, not just recruitment of those who needed extra help and so stigmatizing the program; careful selection and preparation of peer mentors; faculty support and involvement, including responsibility for mentoring peer leaders; program materials made available to all participants; and a full time program coordinator and part time program evaluator who worked cooperatively in the same location.

Peer Led Team Learning

UTAs have also taught in programs called peer led team learning (PLTL), to lead small groups of students outside the primary classroom, using instructor-provided materials and working through problems with students in weekly meetings (Gafney & Varma-Nelson, 2007; Gosser et al, 1996; Gosser & Roth, 1998; Hug, Thiry, & Tedford, 2011; Lewis & Lewis, 2008; Tien, Roth, & Kampmeier, 2002). The UTAs in this program model had been successful students in the class which they were leading, were interested and skilled in communicating with less experienced undergraduates, and received a small stipend for their efforts. The role of the UTA in this model was to actively engage the less experienced students with the course material and with each other in a positive learning environment. In working with the students on small group problem solving, the UTAs modeled respectful discussion, constructive criticism, and an atmosphere of cooperative learning and equitable participation by all students. While most of the literature on PLTL programs provides insight into program implementation and results from participant surveys and interviews, Lewis and Lewis (2008) used a comparison group to study PLTL program effectiveness. They found that students who worked in a PLTL group showed statistically significant improvement in multiple academic measures over traditionally taught students.

Learning Assistant Model

Another UTA program is the Colorado Learning Assistant Model, in which successful and interested STEM undergraduate students are hired to be Learning Assistants (LAs) in various STEM courses across a university (Otero, Finkelstein, McCray, & Pollock, 2006; Otero, Pollock, & Finkelstein, 2010). These LAs meet weekly with the faculty instructor for the course to which they are assigned to plan learning activities and assess the learning progress of the students in the course. LAs also attend a weekly pedagogy seminar in the College of Education, where the LAs read relevant literature, reflect on their teaching practice, and share their teaching experiences with other LAs across the university. LAs work with less experienced undergraduates in recitation sections or as classroom assistants, facilitating interactive discussion and guiding problem solving sessions. The LAs are also encouraged to pursue K-12 science teaching. Although evaluations of this program have not reported retention rates nor followed persistence of students in STEM courses, findings of increased academic achievement have been reported. Students who had learning assistants in their introductory college physics courses made significant learning gains on the Force and Motion Conceptual Evaluation (Thornton & Sokoloff, 1998). In addition, the LAs themselves scored higher on this assessment on average than incoming graduate students who had not been LAs (Otero, Pollock, & Finkelstein, 2010).

Benefits to Students

In addition to the reported benefits for students who have team leaders, learning assistants or UTAs as reported above, other studies have also shown benefits for undergraduate students who have UTAs in the form of higher academic performance and a more positive class environment. Studying peer-led team leaders assisting in an organic chemistry course, Black and Deci (1999) found that undergraduate students performed better when they perceived their team leaders as more supportive of the students' own learning and cared about how they learned.

Using a combination of peer-led collaborative groups and guided inquiry activities, Lewis and Lewis (2008) reported overall increased student achievement on midterm exams and a common final exam compared to exam grades from courses taught in a traditional lecture format. The use of the peer-led guided inquiry program increased student achievement for all students equally; the program did not preferentially impact one type of student (for instance, students with lower SAT scores) more than another.

Benefits to UTAs

Benware and Deci (1984) showed that when people learned material for the purpose of teaching it to someone else, they were more intrinsically motivated, had higher conceptual learning scores, and perceived themselves to be more engaged in the material, even if they never actually performed the teaching. According to White (1959) and Deci (1975), intrinsic motivation is based on the need to make a meaningful impact on one's environment. Bargh and Schul (1980) also demonstrated that the cognitive processing used to study material in order to teach someone else is different from studying to take a test. Preparation for teaching material seemed to result in a more organized cognitive structure of the material for the teacher, helping the teacher to retain more information about the material.

Amaral and Vala (2009) reported that students who were initially unprepared for a rigorous chemistry course for majors took a semester-long remedial chemistry course with success. These students then went on to mentor other underprepared students in the remedial chemistry course, and earned higher grades in subsequent rigorous chemistry courses than students who were initially prepared for chemistry classes and had did not participated in the remedial course as a student or a UTA.

Gafney and Varma-Nelson (2007) found that 119 out of 570 former peer team leaders who had graduated from eleven different institutions ranked their experiences of leading a peer group to be the most significant learning experience (out of 13 suggested experiences) that they had as an undergraduate. These alumni peer leaders also overwhelmingly agreed that their experience helped them to appreciate the value of learning in small groups, appreciate the different learning styles people prefer, develop confidence to teach students and develop an appreciation for what it takes to be a teacher. However, many of these alumni peer leaders did not find that participating in a peer-led team as a student (not a leader) was a particularly important learning experience. Comments explaining the low ratings given to student participation in the peer groups mentioned poor execution of the workshop model by instructors and peer leaders; their peer leaders were not helpful in facilitating learning and the questions posed in the workshops were very tough. Some of the workshops had been used to introduce new material, instead of reinforcing what had been taught by the professor in lecture, thereby giving the peer leaders responsibility for teaching rather than assisting with practice. Based on these findings, one might think it is important to evaluate that what is actually happening in peer-led small groups is beneficial to student learning. Surveying students to confirm that the small groups are helpful and what might be done to improve the experience for learning would be one way to ensure the implementation of peer-led small groups is supporting improved undergraduate learning. One additional finding from this study showed that working as a peer team leader did not affect a student's career decisions, but it may have helped them develop confidence and skills in presenting and explaining complex ideas to others and to develop an appreciation for teaching as a vocation.

Weidert, Wendorf , Gurung, and Filz (2012) found by surveying UTAs and GTAs who had been UTAs that they considered the benefits of the program for themselves to be many: increasing use of effective teaching strategies, improving public speaking skills, experience in working with people, increase in self-confidence and personal insight, acquisition of knowledge about how students learn and behave, strengthen their content knowledge and connections with their major, and to prepare for teaching in graduate school or as future college faculty. UTAs reported that with increased teaching responsibilities came more satisfaction with the UTA experience.

Comparing the UTA experience to an undergraduate research experience, Schalk, McGinnis, Harring, Hendrickson, and Smith (2009) found that skills gained by the UTAs were very similar to those gained by undergraduates participating in a research internship: UTAs begin to identify as a scientist, develop more sophisticated communication skills for science, form more well-defined career goals, and increase their self-confidence. One noticeable finding from the reflective journals kept by the UTAs for this study was that the UTAs mostly concentrated on re-learning science concepts as preparation for teaching; the UTAs did not demonstrate much higher-order thinking, such as synthesizing ideas to form new opinions. Perhaps the UTA experience offered an opportunity to deepen content knowledge before trying to think scientifically in more sophisticated ways.

Benefits to Faculty

Fingerson and Culley (2001) found that UTAs played an intermediary role between professor and students, and gave feedback to instructors that helped the instructors to improve communication and instruction to their undergraduate students. Otero, Finkelstein, McCray and Pollock (2006) reported that interested professors, working with learning assistants to prepare for their course, have reconsidered what and how students learn. Romm, Gordon-Messer, and Kosinski-Collins (2010) believed their work as faculty was enriched by using UTAs in biology courses because the program was "harnessing the passion and innovation of the next generation of science teachers" (p.86).

Recommendations for Using UTAs

Based on extensive qualitative data collected from instructors, UTAs, and undergraduate students, Fingerson and Culley (2001) recommended that UTAs should be used in visible roles in the classroom to model active learning for their students. To do this, instructors need to plan ahead for how they will use UTAs in their classes, to maximize visibility and contact with undergraduate students and to enhance the educational benefits received by the UTA. The pedagogical relationship between professor and UTA is different than the relationship between the professor and a class of undergraduate students. The professor has a chance to mentor the UTA as a possible future faculty member, thus the UTA experiences for themselves the leadership and organizational effort it takes to be an instructor. Fingerson and Culley also recommended that UTAs be encouraged to reflect deeply on their experiences so the UTA and instructor can assess the overall experience had by the UTA. Last, evaluation of the UTAs should be performed by the undergraduates with whom they work, so that factors in choosing effective UTAs and ways to use UTAs can be improved for the benefit of undergraduate student learning and a positive course experience.

Theoretical and Conceptual 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 and stronger STEM identity (See Figure 2-1). 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 Situated Learning Theory (1991), and Wheeler, Martin, and Suls' Theory of Social Comparison for the Self-Assessment of Ability (1997).

Vygotsky's Zone of Proximal Development

What may help explain how trained and supported UTAs may assist their less-experienced peers develop an understanding of chemical concepts and improve problem solving skills in a general chemistry course is Vygotsky's theoretical construct of the zone of proximal development (Vygotsky, 1978). The more experienced UTA is trained to support the learning of the less experienced peer in an area of development that is just higher than the less-experienced peer can work on his/her own.

Lave and Wenger's Situated Learning Theory

The relationship between UTAs and their students can also be examined in light of Lave and Wenger's situated learning theory (1991). Learning is situated in a specific context and is a social process. This theory helps explain that learning is more than understanding content but is also understanding culture, norms, practices, and values within the discipline of study or a community of practice. To come to this understanding, newcomers, for example introductory STEM students, need the advice and support of the more experienced members of a community of practice. The community of practice needs to have three elements:

- a domain, or area of interest, like general chemistry,
- a community, where people share ideas, assist each other, and learn together, and
- a practice, in which participants are practitioners or are wanting to become practitioners. Developed over time, a practice is an agreed-upon way to solve problems, share resources, and pass on information.

For students to move toward full participation in the community of practice, they would need commit time and effort to the practice and take on more and broader responsibilities within the community. Over time, and with effective assistance of a more experienced member of the community (e.g., a UTA), they will develop an increasing sense of identity as a master practitioner (Lave & Wenger, 1991).

Wheeler, Martin, and Suls' Theory of Social Comparison for Ability

Of course, a UTA is not the only experienced practitioner who can help acculturate new students into the STEM community, but UTAs possess a unique position in the undergraduate STEM community as an intermediary between student and instructor. When an undergraduate student is trying to predict personal success in a course of study or in choosing a college major, he or she tries to accurately self-evaluate academic abilities and career interests. Not evaluating oneself accurately could cost a student time and money if the student must switch majors or leave college altogether due to realignment of ability with career aspirations or poor academic performance. The proxy model of social comparison for self-assessment of ability (Wheeler, Martin, & Suls, 1997) seeks to describe how people use social comparison to answer the question "Can I do X?" or" Do I have sufficient ability to perform a specific task effectively?" This theory builds on Festinger's theory of social comparison (1954). Students evaluate their ability by comparing themselves to a more experienced other in their community whose abilities are similar to their own and who has succeeded when putting forth a maximal effort. The UTA can be a credible proxy for the student if selected for similar age, experience, and culture.

These three theories connect a) the skills needed from the UTA to assist their less experienced peers in learning, b) the relationship between UTA and student that can help acculturate the student into the community of practice, and c) the students' need for an effective proxy to predict success in a course of study. Supported by these theories, a UTA program for a general chemistry course was devised and implemented, based on what had been successful in the previous programs that used more experienced undergraduates to mentor, teach, and otherwise assist less experienced undergraduates in STEM learning. The implementation of this UTA program in CHEM 201, the research questions chosen to study the UTA program, and the linkages to the theoretical foundation and conceptual framework are shown in Figure 2-1. The next section describes this program and the literature basis for the variables chosen for study.

Undergraduate Teaching Assistant Program: PRIMES

To address STEM retention improvement at the large, urban, research-intensive university where this dissertation study took place, a long-term university-wide program aimed at increasing retention of STEM majors by providing a more positive learning experience in introductory level STEM courses was implemented in January 2012. The program, Partnership for Retention Improvement in Mathematics, Engineering and Science (PRIMES) is a 5-year NSF-funded STEM Talent Expansion Program (STEP) project. PRIMES united faculty from the College of Arts and Sciences, Speed School of Engineering and the College of Education and Human Development at the University of Louisville to transform teaching and learning in STEM courses for three related groups: undergraduates in introductory STEM courses, more experienced STEM students serving as UTAs, and STEM faculty who are simultaneously teaching introductory level students and mentoring UTAs.

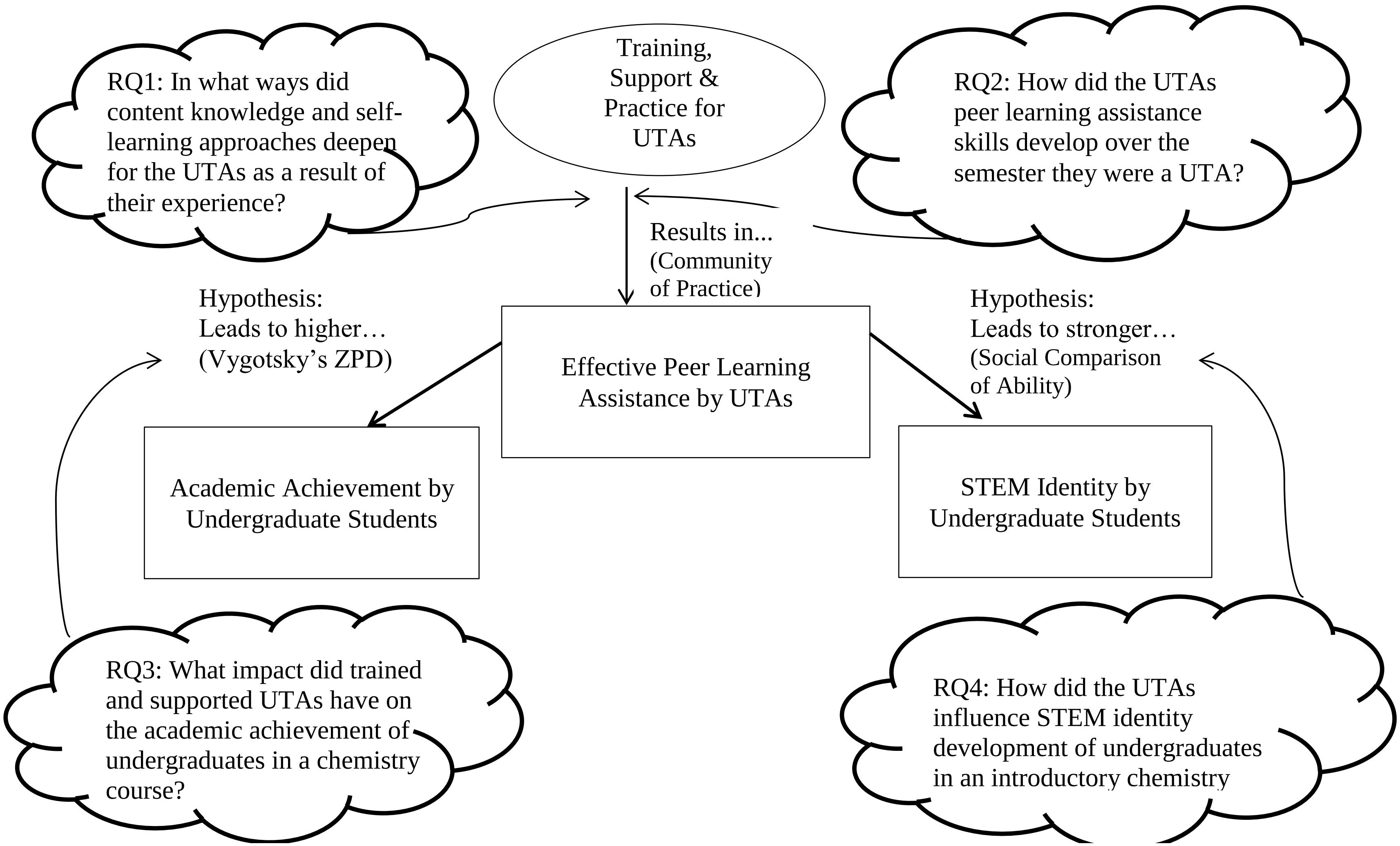

Figure 2-1 Illustration of how and where research questions link with the conceptual framework elements

In the semester that this study took place, PRIMES assigned approximately 60 trained and supported undergraduate teaching assistants (UTAs) competitively selected based on grades and professor recommendations from eight mathematics, science and engineering departments to teach small groups of undergraduate students enrolled in introductory courses across science, math and engineering departments. The UTAs were compensated for their work with a modest stipend and course credit. The UTAs functioned as peer educators for less experienced undergraduate students in laboratory, recitation, or supplemental instruction settings, as chosen by each department, using the Colorado Learning Assistant program (Otero, Finkelstein, McCray, & Pollock, 2006; Otero, Pollock, Finkelstein, 2010) as a model. As described previously in this chapter, other studies have also shown evidence of positive learning outcomes when peer educators are used to assist less experienced undergraduate students in introductory STEM courses (Amaral & Vala, 2009; Gosser & Roth, 1998; House & Wohlt, 1990; Hug, Thiry, & Tedford, 2011; Lewis, 2011).

Study Focus: UTA Program for CHEM 201

For this study, undergraduate students who had had been successful in an introductory level general chemistry course for STEM majors (CHEM 201) were selected to assist their less-experienced undergraduate peers in learning general chemistry concepts and associated calculations. The six CHEM 201 UTAs joined approximately 50 other UTAs from nine science, mathematics, and engineering departments for training and support from College of Education faculty and staff, College of Arts and Sciences faculty, and College of Engineering faculty.

Selection criteria for UTAs. The CHEM 201 UTAs were selected by chemistry faculty based on excellent grades, an application demonstrating interest in teaching, and recommendation from a professor vouching for the applicant's communication skills and ability to connect with peers. The UTA is in a unique position to support learning of course material in that they have recently been successful in this course and can recall their own struggles, misconceptions, and learning successes.

UTA training and support strands. In order to offer strong peer learning assistance to less-experienced peers, UTAs need both content knowledge support and pedagogical training. Chemistry faculty for CHEM 201 offered support for content unpacking and awareness of common misconceptions held by chemistry students, perhaps by the UTAs themselves. Science educators, including the author, from the College of Education trained and supported the UTAs in using active learning strategies, such as cooperative learning (Johnson, Johnson, & Smith, 1998) and problem-solving groups, questioning strategies, metacognitive activities, and formative assessment approaches that inform future instruction. One of the CHEM 201 professors also participated in the learning strategies instruction, presenting some topics and listening in on UTA small group conversations.

UTA training occurred in a three-day workshop prior to the beginning of the semester and in six hour-long seminars during the teaching semester. Attendance was mandatory for the pre-semester workshop and three out of the six seminars. Three other seminars could be completed by asynchronous online discussion participation in response to prompts, in addition to the usual submittal of a one-page written reflection on a chosen practice for that period of the semester. The workshop consisted of two parts: a two-day introduction to learning theory and pedagogical strategies for all UTAs involved in the PRIMES program and a one day session spent within the mathematics, science or engineering department in which the UTA was teaching. During the two-day session, UTAs listened to education and STEM faculty present short talks on pedagogical strategies, discussed how these strategies could be used in their own courses, and tried out the strategies for themselves with structured experiences organized by the faculty workshop presenters. Examples used to demonstrate pedagogical strategies used content knowledge from various science, mathematics, and engineering disciplines, including chemistry. Each workshop day ran from 9 am to 3 pm, with a half-hour break for a catered lunch. Many of the activities and approaches for this workshop were adapted from the University of Colorado Learning Assistant Model © and were used with permission. For the day spent within the disciplinary department, safety training, and disciplinary-specific and course-specific training and planning took place with the PRIMES STEM faculty and senior instructors involved with the courses using UTAs. This workshop day ran approximately the same amount of time and again offered a catered lunch for the UTAs and STEM faculty.

Seminars met six times during the semester for 50 minutes each, and were scheduled as a regular course so that UTAs were able to plan their academic schedule to avoid other academic commitments and attend the seminars just as they would any other course. The seminars incorporated two main subjects: the topic that UTAs were asked to focus on in the prior period (and for which they submitted a written reflection assignment), and the upcoming topic for the next focus period. The topics for the seminars included critical incidents from the first few weeks of classes, questioning strategies (e.g., convergent/divergent question types, wait time, equitable participation in questioning), metacognitive strategies to help students think about their own thinking during problem-solving tasks, and formative assessment strategies (e.g., eliciting student understanding, giving effective feedback, student self-assessment) (Moss & Brookhart, 2009), which connected questioning and metacognition strategies. The first goal of each seminar was to process the written reflection from the period prior to that seminar. The objectives of this process for the UTAs included sharing experiences to develop a sense of UTA community, imparting positive stories to learn from each other, and revealing negative issues to problem-solve together. The format of this first focus typically included some variety of small-group sharing (e.g. pairs, small groups, within disciplinary departments or in multi-disciplinary groups, etc.) so that all had an opportunity to share their experiences, followed by a brief whole-group reporting of main ideas so that all could benefit from cross-pollinating ideas across the larger group. The second goal of each seminar was to prepare for the upcoming focus. UTAs had read an article about the upcoming topic prior to seminar, so the second portion of the seminar processed that article. The main goal was to ensure that students left the seminar confident that they understood the intentions of the upcoming focus. A reflection on the practice of that topic was assigned to be submitted before the next seminar, giving the UTA time to choose a strategy to work on, use it with students, and write about how well the strategy worked with students and suggest what could be done differently next time.

The UTAs were supported in content knowledge growth by the chemistry faculty who taught CHEM 201. Weekly planning meetings were held in which faculty and UTAs discussed learning objectives and planned engaging activities to achieve those learning objectives with the students. During the meetings, content knowledge needed to assist students in meeting the learning objects was reviewed by the UTAs. Individual review for teaching preparation was the responsibility of the UTAs and they frequently mentioned the extensive time it took to prepare for class. This seemed to be a surprise for most UTAs, although they were not discouraged about the work. Many reported in their reflections that they came to a new realization of the responsibilities that faculty take on in order to teach effectively.

In particular, this study will focus on the six UTAs teaching 14 recitation sections of an introductory chemistry course for STEM majors, a comparison group of three GTAs teaching 15 recitations of the same course, and the approximately 700 undergraduate students enrolled in the 29 total recitation sections. Four senior instructors each taught a large lecture section (approximately 200 students per section) of the chemistry course and the students self-selected into one of the 29 recitations sections. The recitation sections were scheduled at various times from 8 am to 3 pm every day of the week. GTAs and UTAs were assigned to those scheduled sections after student registration, so there was little chance that a student signed up for a particular section in order to have a particular UTA or GTA.

Each UTA worked with students from one of the four senior instructors and the UTAs met weekly with that instructor to discuss learning objectives and plan active learning activities to help reach those objectives. Five of the six UTAs met weekly for 50 minutes with two sections of twenty- five students each, as well as offering at least two hours of one-on one help in their offices weekly. One of the UTAs worked with students from two senior instructors; this was her second semester as a UTA, and she had a double load of four twenty-five student sections. All UTAs did some minor grading of quizzes, but the focus of their assistantship was to actively teach and assist less experienced students. The UTAs, along with one or two volunteer faculty from their department, participated in pedagogical training guided by the College of Education and Human Development. This training consisted of two parts: (1) a three day workshop at the beginning of the teaching semester that highlighted learning theory, pedagogical strategies, and best teaching practices for STEM learning and (2) bimonthly seminars which focused on a different topic each seminar, such as formative assessment, convergent/divergent questioning, and development of metacognitive skills. UTAs were asked to write six reflections about their practice and use of pedagogical information presented in the workshop and seminars. UTAs were paid a small stipend from the PRIMES project and received one course credit in exchange for teaching and workshop/seminar participation.

Each GTA worked with students from two of the four senior instructors and the GTAs met occasionally with those instructors when problems needed to be addressed. These three GTAs met weekly for 50 minutes with five sections of twenty- five students each, as well as offering weekly at least two hours of one-on one help in their offices. Like the UTAs, GTAs' main focus was to teach and assist undergraduate students. Historically, the GTAs did not participate in any special on-going pedagogical training, met only occasionally with CHEM 201 faculty, and did not collaborate in planning recitation learning activities with the UTAs. This was still the case during the semester data were collected for this study. GTAs were paid a stipend and given tuition remission in exchange for teaching duties, supported by the Department of Chemistry.

In order to evaluate the effectiveness of this program on increasing STEM retention, two outcomes, student academic achievement and student STEM identity were chosen as reasonable measures for the one semester time-frame in which this study took place. The literature basis for choosing these two variables follows in the next section.

Academic achievement in STEM courses. Although not the only significant factor affecting student retention in STEM programs of study, student academic achievement in STEM courses does impact a student's ability to advance in a STEM program of study. Course grades have important motivation and entitlement effects that may be enhanced in science and engineering departments, which are well known for giving low grades with high variability (Goldman & Slaughter, 1976; Rask, 2010). STEM courses often demand great amounts of preparation in order to succeed in them and require high levels of mastery at each step in the hierarchy of course. The same factors that attract more high achieving students may result in their being graded more stringently (Goldman & Slaughter, 1976; Strenta & Elliott, 1987). Maltese and Tai (2011) found that students who completed more STEM credits in their first year and those who earned better grades than their peers were more likely to persist in a STEM program. Using a series of binary response models, Rask (2010) found that the grades received in a course are an important determinant of whether a student takes another course in the major.

Demographic variables impacting academic achievement. The following variables may be used in this study as covariates for academic achievement, in order to compare group characteristics.

ACT scores. ACT scores provide a criterion for college readiness based on thousands of student scores and success. The ACT has empirically developed Benchmarks for College Readiness in English composition, social sciences courses, college algebra, and biology (ACT, Inc., 2012). Students who meet a Benchmark on the ACT have approximately a 50 percent chance of earning a B or better and approximately a 75 percent chance of earning a C or better in the corresponding college course or courses.

Race/Ethnicity Identification. Black, Latino, and American Indian students in the United States graduate from high school at lower rates (about 60%of those who started high school in 2002 graduated in 2006) than White and Asian students (81% and 90% of whom graduated, respectively) (Education Trust, 2009). Black, Latino, and American Indian students who persist in high school frequently attend schools with fewer resources to develop the solid academic base necessary for succeeding in college science coursework. A lower percentage of Black men complete science majors than the percentage of white men that complete science majors; however about the same percentage of Black women complete science majors as White women (Johnson, Brown, Carlone, & Cuevas, 2011).

Parental level of education. Seymour and Hewitt (1997) found that parents are important in a student's decision to persist in science studies, but few studies have examined how family and personal factors combine to explain students' persistence in a science program. Family support can play a tremendous role in helping young adults to successfully adapt to college or university by buffering the negative effects of transition (Holahan & Moos, 1981). Parental involvement has been found to be an important predictor in a student's college science success (Ratelle, Larose, Guay, & Senécal, 2005), and so college educated parents should be able to help students transition more effectively in the first year of college. Additionally, level of education can be a proxy measure of family socioeconomic status, which has been found to affect students' academic success (Walpole, 2003).

STEM Identity. Science identity was conceptualized by Carlone and Johnson (2007) as a construct that may help explain factors that affect the experiences of students in science programs of study. Carlone and Johnson developed, tested, and refined a model of science identity that helped to understand women's experiences in science. To develop a strong science identity, one needs to develop the three interrelated components of the model: *competence* in science content, ability to *perform* relevant science practice, and *recognition*, by self and others, as a science person. They found that a person's racial, ethnic, and gender identities interacted strongly with recognition by meaningful others (such as science professors and internship supervisors, as well as family and friends) as a science person. Carlone and Johnson recommended further investigation into the performance aspect of the science identity model; how and why students learn to perform as a scientist and how their performance affects their science identity is not well-studied.

Hazari, Sonnert, Sadler, and Shanahan (2010) expanded Carlone and Johnson's science identity model (2007) by adding the construct of *interest* in their analysis of surveys on physics identity from a nationally representative sample of freshmen college students. Interest in science has been found to have an impact on student participation and persistence in science. The research surrounding Social Cognitive Career Theory

(SCCT) supports the impact of interest on career choice (Lent, Brown, & Hackett, 2002). SCCT claims that interest develops in an activity, like studying science, when a person feels competent doing the activity and thinks that doing the activity will produce a valued outcome. Hazari et al. found that for the 3800 college freshman they surveyed, recognizing oneself as a "physics person" correlated most strongly with interest in physics topics and recognition by others as a "physics person". Recognizing oneself as a "physics person" correlated significantly, but less strongly, with academic achievement in science and confidence in science ability. Based on the results of their study, Hazari et al. asserted that identity analyses will be useful in predicting longer-term persistence in science.

Research Variables

Based on the literature review and the implementation of the PRIMES project, several variables were selected to be explored in the present study (see Table 2.1). Variables were selected based on their importance in answering the research questions of interest and availability of the data from the implementation of the project. As a reminder, the research question guiding this study were:

1. How did the UTAs change as a scholar as a result of the UTA experience?
    a. In what ways did disciplinary content knowledge deepen?
    b. Which learning approaches did UTAs mention as aiding their own learning?
2. How did the UTAs' peer learning assistance skills develop over the semester(s) they were a UTA?

3. What impact did trained and supported UTAs have on the academic achievement of undergraduates in an introductory chemistry course?
4. How did the UTAs influence STEM identity development of undergraduates in the introductory chemistry course?

Table 2-1

Study Variables and Reasons for Selection

Variable Type	Reason
Student Variables	
Final Exam Grade	Measure in percentage points of STEM academic achievement from common CHEM 201 final exam
Persistence	Student enrollment in CHEM 202, the next course in the 2-course General Chemistry sequence. Required for STEM majors such as biology, physics, chemical engineering, biomedical engineering, etc.
STEM Recognition	Measure of student recognition of themselves and recognized by others as a ‘science person’ or a ‘math person’
STEM Interest	Measure of student interest in science and mathematics activities: experiments, talking about and learning more about science and mathematics topics
ACT/SAT scores (Mathematics, Reading)	ACT scores correlate with preparedness of student to achieve post-secondary course learning objectives.
Race/Ethnicity Identification	Studies report lower achievement and higher attrition from STEM programs for students in minority racial/ethnic groups
Parent education levels	First generation college students have little familiarity with expectations and structures of higher education. Also proxy for socioeconomic status, which positively correlates with academic achievement
Number of STEM AP courses taken in HS	The more AP courses a student has passed in high school, the more familiar they may be with post-secondary course expectations. Also speaks to student interest in STEM
College GPA	Grade point average earned at university before Fall 2012. Measures how well a student is doing academically in college.
TA Variables	
TA Impact on Academics	A measure of the degree to which students perceive TA to possess strong content knowledge, lead effective discussions, and be a valuable academic resource.
TA Rapport-Building Skills	A measure of the degree to which students perceive TA to encourage questions, be open in communicating, and care about students.
Content Knowledge	Measured by pre and post test scores on the TA pre/post Content Knowledge Test; similar to CHEM 201 Final Exam

CHAPTER 3

RESEARCH DESIGN AND METHODOLOGY

This study examined the impact of specially trained and supported undergraduate teaching assistants (UTAs) on the course experience of undergraduates in an introductory chemistry course required for many science and engineering majors. Concurrently, this study sought to describe the ways in which the UTAs developed teaching and communication skills and the benefits gained by the UTAs as a result of their teaching experience. By describing the UTA experience as well as the impact of the UTAs on their undergraduate students, this study strived for a rich understanding of a mentor-mentee relationship that could positively affect STEM undergraduates. To reach this understanding, both quantitative and qualitative data needed to be collected, analyzed and interpreted. Mixed methods, an integration of quantitative and qualitative methods, were chosen as the approach to reach this understanding.

The purpose of this chapter is to outline the study methodology including research design, study site and sample, instrumentation/measures, data collection procedures, and data analysis plan.

Research Design

An important role of research questions is to direct the selection of the research methods (Teddlie & Tashakkori, 2009). To answer the four research questions detailed in Chapter 2 requires both quantitative and qualitative data. Therefore, a mixed methods approach was used for this study.

A parallel multilevel mixed-methods design was chosen to answer the study's research questions and take advantage of the hierarchical context in which the study took place. The teaching assistant was a core unit of analysis; however, essential evidence about the impact of the UTA was gathered at the undergraduate student level. Quantitative data were collected at the student level in parallel with quantitative and qualitative data collected, also in parallel, at the teaching assistant level. Both types of data, quantitative and qualitative, were analyzed as described in the Data Analysis section of this chapter, and the results were used to make multiple inferences, which were then integrated to answer the four research questions. A table showing the phases of data collection for this study is shown in Table 3-1. The design approach for each research question is detailed below, along with how the design addresses threats to internal validity.

Research Question 1a – UTA content knowledge

To answer the first research question about UTA content knowledge growth, an untreated control group design with dependent pretest and posttest data was used. The treatment group comprised the six UTAs for CHEM 201 who participated in the PRIMES UTA program described in Chapter 2. The control group included the three GTAs who received no special pedagogical training, and met a few times individually with their senior instructor to discuss problems over the semester. The instructor did not give specific direction to the GTA about teaching strategies or activities, but also did not discourage the GTA from planning or discussing learning activities. The pre and posttest data were common final exams for CHEM 201 that had been previously administered to students in the past 5 years. The same multiple choice questions, representative of the concepts taught in the CHEM 201 course, were used on the pre- and posttests. The quantitative data were supplemented with reflections written by the UTAs over the course of the semester, concerning their experiences assisting students in their recitation sections.

Additionally, UTAs responded to an end-of semester guided reflection about perceived growth in content knowledge, and the responses from each UTA were compared to the content growth shown by the gain score from the content tests.

This design addressed internal validity threats in several ways. Because the TA groups differed by age and academic standing (undergraduate versus graduate level students), the groups were tested for content knowledge using a common pre-test/post-test. Although the tests questions were identical pre and post, the TAs did not review their tests after taking them, nor were they told what score they received so they had no evaluation from professors of what questions they had missed on the test. This may have prevented the pretest from “priming” TAs to re-learn specifically what they had missed on the pre-test in order to perform better on the post-test. That the test questions were identical pre and post meant that there should have been no change in how the test questions performed pre and post. There was no diffusion of treatment because the treatment (pedagogy workshop and seminars) was available only for the UTAs and professors confirmed that no GTAs participated in any formal pedagogical training during the semester or while in the graduate program. Because the motivation of the GTA was to focus on coursework and research in their graduate studies, there was little incentive for the GTAs to go above and beyond what was asked of them for the teaching assistantship: meet with five recitation groups each week and hold two announced office hours per week. UTAs, in contrast, earned a grade for participating in the pedagogy seminars and working with the senior instructors to plan learning activities.

Table 3-1

Phases of Data Collection

Research Question	Pre-Semester	Week 3	Week 5	Week 7	Week 9	Week 12	Week 15
1a. In what ways did disciplinary content knowledge for the UTAs as a result of their UTA experience?	Pretest for UTAs and GTAs (version of final exam for CHEM 201)						Post-test for UTAs and GTAs (similar version of CHEM 201 final exam)
1b. In what ways did self-learning approaches deepen for the UTAs as a result of their UTA experience? *(Quantitative & Qualitative)*		Reflection Critical incident Classroom Observation	Reflection Questions	Reflection Questions Part 2	Reflection Metacognition	Reflection Formative Assessment Classroom Observation	Final Reflection Undergraduate End of semester Survey Interview with CHEM 201 senior instructors
2. How did the UTAs' peer learning assistance skills develop over the semester(s) they were a UTA? *(Qualitative)*		Reflection Critical incident Classroom Observation	Reflection Questions	Reflection Questions Part 2	Reflection Metacognition	Reflection Formative Assessment Classroom Observation	Final Reflection Undergraduate End of semester Survey Interview with CHEM 201 senior instructors
3. What impact did trained and supported UTAs have on the academic achievement of undergraduates in an introductory chemistry course? *(Quantitative)*	Academic covariates collected for undergraduates						Final Exam for CHEM 201 Undergraduate End of Semester Survey
4. How did the UTAs influence STEM identity development of undergraduates in the introductory chemistry course? *(Quantitative)*	Academic covariates collected for undergraduates						Undergraduate End of semester survey

Research Question 1b – UTA Self-learning Approaches

The research design for describing how the UTAs used newly-learned pedagogical strategies for their own learning was a phenomenological description of how and/or when each UTA used these strategies for their own learning as mentioned in their reflections.

Research Question 2 – UTA peer learning assistance skills

The research design for Research Question 2 was a parallel mixed methods design to investigate the peer learning assistance skills of the UTAs from three perspectives: the UTAs themselves, the students in the UTA-led recitation sections, and the researcher and STEM faculty who worked with and observed the UTAs assisting their students. To describe the development of peer learning assistance skills perceived by each UTA, phenomenological descriptions of the UTA learning assistant skills developed during the UTA experience were identified from the reflections written by the UTAs. This type of research design is appropriate for the research question and the chosen perspective of the UTA because it focuses on the lived experiences of the UTAs as the UTAs were introduced to learning theories and strategies, practiced the strategies with their students, and reflected on their successes and challenges. Data were collected systematically across the duration of the semester in the form of first person reflections written by the UTAs in response to prompts about the pedagogical strategies they used with their students in the recitation sections.

To evaluate the peer learning assistant skills from the student perspective, an untreated control group with posttest only design was used. An end of semester course experience survey with questions about their TA's teaching skills, communication skills, and content knowledge was taken anonymously by the undergraduates in both the UTA-led recitation sections and comparison GTA-led recitation sections. This survey can be found in Appendix A. The reliability of the survey used with the undergraduates was estimated with computation of Cronbach alpha which is reported in the Chapter 4. The content validity of the survey was evaluated with principal components analysis and results from that analysis are also in Chapter 4.

To describe the peer learning assistance skills of the UTAs from the faculty point of view, an interview was conducted with CHEM 201 senior instructors who worked closely with the UTAs and field notes were gathered by me during two classroom observations over the semester. All data, quantitative and qualitative, were compared and contrasted to triangulate the results from the perspectives of UTAs, students, researcher, and faculty to strengthen the credibility of any one of the above data sources (reflections, survey, interview, and field notes). In the interest of disclosing any biases I might have projected on my observations, I have a masters' degree in analytical chemistry, am a licensed (accomplished practitioner) chemistry teacher in the state of Indiana, and have taught high school and college level general chemistry for several years.

Although how peer assistance learning skills developed for the six UTAs in this program may not be generalizable to UTAs everywhere, the descriptions from this method will characterize the actual skills used to provide treatment received by the undergraduate students in the recitation sections led by these six UTAs.

Research Question 3 - Impact of the UTA Program on Academic Achievement

To answer the third research question about the impact of the UTA program on academic achievement of the undergraduate students, an untreated control group design posttest only was used. The treatment group included the undergraduate students enrolled in one of the recitation sections led by one of the six trained and supported UTAs. The comparison group was a comparable number of students enrolled in GTA led recitation sections. Because there was not a meaningful chemistry pre-test validated and used by the senior instructors for this entry-level course, the inclusion of select academic variables, such as ACT scores, number of mathematics and science AP courses taken, and parental education level were used to control for selection bias in initial student academic preparedness and ability between the treatment and comparison groups. The nested nature of the data (students in recitation sections) was taken into account via multi-level modeling. Historical outcome data (scores from a common final exam) from three prior years were used as an independent comparison with the GTA comparison groups to see if students in GTA led sections performed statistically the same on the final exam as in previous years.

Although students were not randomly assigned to GTA or UTA recitation sections, registration into CHEM 201 took place through the recitations which were evenly scheduled over the day during all week days. UTAs and GTAs were assigned to recitation sections based on schedule availability and balanced across senior instructors. Every UTA worked for one of the four senior instructors (2 recitation sections) and every GTA worked for two of the senior instructors over 5 recitation sections. Students did not know the identity or type of TA until they came to their recitation section because the

TAs were not scheduled until the day before classes began. Academic covariates were used to compare the UTA-led student group to the GTA-led student group. Because the treated (UTA) and untreated (GTA) groups were so carefully scheduled over time of day, day of week and senior instructors, this minimized selection-history threat. With the large number of students involved in CHEM 201 (over 700 undergraduates), the groups were measured to be very similar academically and demographically, so selection-regression threats were minimized. The academic and demographic characteristics of the students who withdrew from the course or who did not choose to take the final exam were also examined to check for selection-mortality threats. Attendance records were kept in the recitation sections and short quizzes were often given in the recitation sections, so diffusion of treatment (students attending recitation sections in which they were not registered) was minimized.

Research Question 4 - Influence of the UTA Program on Student STEM Identity

To answer the fourth research question about the influence of the UTA program on the STEM identity of the undergraduate students, an untreated control group design posttest only was used. The treatment group included the undergraduate students enrolled in one of the recitation sections led by one of the six trained and supported UTAs. The comparison group was a comparable number of students enrolled in GTA led recitation sections. The number of mathematics and science AP courses taken in high school was used to contribute to initial STEM identity because enrollment in those elective courses which require specific mathematics and science prerequisites indicate high science or mathematics competence and performance levels, interest in mathematics and science study, and a recognition of the student as a “science person” or a “math person” by both the student and their teachers. The numbers of mathematics and science AP courses taken, as well as other possible academic covariates, were used to control for selection bias between the treatment and comparison groups. The nested nature of the data (students in recitation sections) was taken into account via multi-level modeling. The STEM identity outcome variable was measured by a set of eight questions concerning how the student identified herself as a scientist from the end of semester Course Experience Survey taken by nearly all the undergraduate students in CHEM 201. The validity of the survey for STEM identity was confirmed by principal components analysis and reliability within the student sample was estimated by computing Cronbach alpha. The end of semester Course Experience Survey can be found in Appendix A.

Study Site and Sample

The research site of this study was a medium-large, urban, Midwestern, research-intensive university. The study took place in the context of CHEM 201, an introductory level general chemistry course designed for students intending to major in a science or engineering discipline.

Undergraduate Sample

All undergraduate students enrolled in CHEM 201 during Fall were invited to participate in the study. Approximately 600 undergraduates were asked to complete a voluntary end of semester survey during the recitation section of CHEM 201 in which they were enrolled. The undergraduates agreed to take part in the study by completing and submitting the survey anonymously. Because the surveys were completed while in the recitation section and most students attended recitation sections regularly, the return rate for the surveys was very high (70%). Characteristics of the undergraduate treatment group and the undergraduate comparison group are found in Table 3-2.

The treatment group consisted of 342 undergraduates in fourteen UTA-led recitation sections across all four senior instructors. The comparison group consisted of 369 undergraduates enrolled in fifteen recitation sections led by the GTAs.

Table 3-2
Characteristics of All Undergraduates in Treatment (UTA-Led) and Comparison (GTA-Led) Groups

	UTA-led Undergraduates	**GTA-led Undergraduates**
N	342	369
Average ACT Mathematics score	27.0	26.7
# of AP STEM Courses	0.88	0.77
% Parent with College Experience	61	47
% Non-white	17	24
% Female	34	34

TA Sample

The combined TA sample included nine teaching assistants (6 UTAs, 3 GTAs) assigned to lead 25-person recitation sections 50 minutes per week.

The UTA sample included six trained and supported UTAs who took part in a pedagogy practicum course, described in Chapter 2. The UTAs were chosen from chemistry major applicants based on excellent chemistry grades and recommendations from chemistry faculty that attested to the UTA's desire to work with less-experienced peers, skills in communication, and a good work ethic. All six UTAs who taught sections of CHEM 201 were invited to participate in the qualitative strands of the study, designed as a multiple case study with data collection throughout the semester. All UTAs were of traditional college age (18-24) and were chemistry majors. Four of the six UTAs had also participated as a trained and supported UTA in another chemistry course during Spring 2012 semester and had returned to the program to repeat participation in the UTA practicum and teach CHEM 201 recitation sections. While five of the UTAs taught 2 recitation sections per week, one of the veteran UTAs taught a double load of 4 sections per week.

The GTA sample included three graduate students who had been awarded traditional departmental teaching assistantships that provided tuition remission and a stipend. Teaching assistantships generally expected 15-20 hours of work per week from the graduate student and required English-language competency measured by TOEFL score or successful completion of the Intensive English as a Second Language Program at the University of Louisville.

Instrumentation/Measures

UTA Content Knowledge

In order to measure whether any change in UTA content knowledge occurred during the semester, the UTAs and comparison group GTAs for CHEM 201 were administered an abbreviated version of the CHEM 201 final exam which had been used the previous year. This exam was developed by senior instructors and was similar in content and format to exams that have been given to CHEM 201 students in the chemistry department for several years, but with fewer questions so that the test could be easily completed in one hour. The administration of this exam took place the week before classes started and again during the last week of classes. The exam questions were the same each time. The 23-question multiple choice exams were scored by computer from scanned answer sheets and scores were reported as the percentage of points earned.

An abundance of qualitative data were collected from the UTAs regularly over the course of the semester in the form of written reflections which describe the UTAs' successes and struggles in assisting their less experienced peers to learn the course material. These narratives describe how learning to teach and working with students changed the UTAs' approaches to learning material in their own classes. Reflections from each UTA were collected in response to prompts about what the UTA considered to be foundational knowledge in chemistry, use of questioning strategies, planning and implementation of formative assessments, and introduction of metacognitive strategies. At the end of the semester, the UTAs were asked to reflect on their experiences and exemplify how their content knowledge, approaches to learning, attitudes towards their students, and opinions of the UTA program had changed over the course of the semester. The UTA end of semester reflection prompt can be found in Appendix A.

Undergraduate Academic Achievement in CHEM 201

This study measured the academic achievement of the approximately 700 undergraduates in CHEM 201 using the common final exam developed and given by the four senior instructors for CHEM 201. Versions of this exam have been given in common for at least 3 years that the class has been offered and the professors are satisfied that the content of the test assesses the major concepts covered in class. The 35-question multiple choice exams were scored by computer with oversight by each student's senior instructor, and the score reported as the percentage of answers which were correct. Historical data (average exam scores in the course each year) from last three years were available and were compared to the scores obtained in Fall 2012.

Undergraduate Perception of TA Academic Support

Survey questions were developed to evaluate the value of the UTA or GTA to the undergraduate students. Nineteen 5-point Likert scale items addressed issues related to effective peer learning assistance (Black & Deci, 2000; Gosser & Roth, 1998; Tien, Roth, & Kampmeier, 2001), such as the UTA trying to understand how a student is thinking before offering assistance and conveying confidence in the student's ability. The complete survey (Undergraduate Course Experience Survey), including questions about STEM identity discussed below, was pilot field tested with a group of ten undergraduate STEM majors who were not students in CHEM 201. Pilot students responded to the survey questions and gave feedback by answering the following questions:

1. Is the wording of the directions and the questions clear and unambiguous? (If not, please note which questions or directions are not clear)

2. Do some of the questions need to be rephrased or dropped from this survey? Please note which ones and why you suggest this.

3. Are there additional questions I should ask to find out about STEM identity of an undergraduate student?

As a result of the pilot field test, no substantive changes were made to the survey. Only minor modifications were made to clarify the questions. The surveys were submitted anonymously by the undergraduates to encourage students to candidly answer questions about the types of experiences they had with the UTA or GTA leading their recitation section. The undergraduate end of semester survey is found in Appendix A. Principal components analysis (PCA) was performed with the survey responses to address content validity for this sample of students.

STEM Identity

Because no instrument has been developed to measure STEM identity in undergraduates, eight survey questions were developed that aligned with the constructs of science identity detailed in the work of Carlone and Johnson (2007), including competence, performance, interest, and recognition. The survey was part of the same end of semester survey discussed above (Undergraduate Course Experience Survey) that was first pilot field tested and then administered to all the undergraduates in CHEM 201 during one of their last recitation sections. Undergraduate students were explicitly asked to consider the role their UTA or GTA played in their identity development and interest in STEM activities. As a result of the field test, no substantive changes were made to the survey. Only minor modifications were made to clarify the questions. A copy of the Undergraduate Course Experience Survey that includes the measures for TA academic support for undergraduates and STEM identity is found in Appendix A.

Data Collection Procedures

Data collection began in August 2012 and was completed by December 18, 2012, except for student academic covariate and demographic data obtained from the university's Institutional Research in March 2013. A chart showing when phases of data were collected is included here (see Table 3-1). Data from UTAs were collected by the senior instructor (pre and post content knowledge exam scores), the pedagogy practicum instructor (reflections), or the author (classroom observations) as part of the UTA program. Data from the undergraduates were collected by the university's institutional research database (student demographic data), the senior instructor (final exam scores) or the UTA (anonymous end of semester surveys collected in a manila envelope) at the end of the semester. Interviews with the senior instructors were carried out near the end of the semester by the author.

Data Analysis Plan

Research Question 1a – UTA Content Knowledge

Pre and post content test scores were reported for the six UTAs and three GTAs. Due to the small number of samples, and the dependent nature of the scores (pre and post semester test scores for each teaching assistant), a Wilcoxon Signed Rank test was performed at $\alpha = .05$ to test for significant differences in initial scores and in gain scores (using matched pairs approach for the growth question) between the UTA and GTA groups.

Research Question 1b – UTA Self-learning Approaches

The types of learning approaches practiced by the UTAs in their teaching and any change in using those learning approaches for their own learning were discussed by the UTAs in their reflections. Any comment made by the UTA in their reflections concerning how they were using what they had learned in the pedagogy workshop and seminars for their own learning approaches were reported in phenomenological descriptions for each UTA. Observations from the UTAs' classrooms and comments made by senior instructors concerning UTA self-learning approaches were added to the description. A summary of similarities and differences between the UTAs' descriptions of self-learning approaches was generated.

Research Question 2 – UTA Peer Learning Assistance Skills

Data from the six reflections written by each CHEM 201 UTA distributed throughout the course of the semester, field notes from classroom observations of the

CHEM 201 UTAs, and interview responses from the senior faculty working with the CHEM 201 UTAs were synthesized to create phenomenological descriptions of each UTA's skill growth. Using multiple methods and data sources (student reflection, researcher observations, and instructor interview) helped to strengthen the validity of the findings. For each UTA, the researcher reviewed the reflections, field notes and interview transcripts twice before creating a phenomenological description of peer learning assistance skills for each UTA. Similarities and differences between the UTA descriptions were noted. If conflicting evidence had been discovered, further probing by follow-up questioning and fact-checking would have taken place. The researcher's background as a former STEM student, a former undergraduate chemistry instructor, and an education researcher were made transparent in any inferences drawn from the data.

Data from the undergraduate Course Experience Survey, reduced to TA quality factors through PCA, were analyzed by comparing the means from the UTA-led students with the means from the comparison GTA-led students. Predictors for each of the TA quality factors were further explored using linear regression analysis.

One of the functions for the answers to this research question is to evaluate the reliability of the treatment implementation, both for the treatment of the UTAs in pedagogical training and the treatment of undergraduates in the UTA recitation sections, where UTAs should be using the skills they learned in the seminars and workshop to create a positive learning environment for their students. If the participants in this study do not report some difference between UTA and GTA recitation sections, then any difference found in student outcomes of academic achievement or STEM identity may be attributed to something else other than the UTA program. To evaluate UTA and GTA recitation section learning environments for any differences, both quantitative data (student survey) and qualitative data (reflections, observations, and interviews) were considered. The analysis of the quantitative data for this research question was used to explore outcome variables for the next two research questions while the qualitative data was used to contextualize the quantitative data to answer only this research question concerning UTA peer learning assistance skill development.

Research Question 3 – UTA Impact on Student Academic Achievement

Given the nested structure of the data, students within course sections, hierarchical linear modeling (HLM) (Raudenbush, Bryk and Congdon, 2004) was used to assess the relationships between student control variables such as ACT scores, number of STEM AP courses taken, and hours completed in college, and section-level variables such as the presence of a UTA and the amount of experience the UTA or GTA had in the program, with the outcomes of student course grades. 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 allows investigators to model both student-level and course section level data at the same time in order to investigate relationships and interactions among the variables at both levels (Raudenbush & Bryk, 2002).

Hierarchical linear model development. For this analysis, there were approximately 700 undergraduate students nested in 29 course sections of an introductory general chemistry course suitable for STEM majors. Course section was chosen as the cluster unit for two reasons: First, the cluster size of sections is much more uniform than the cluster size of students by individual UTAs and GTAs. The range of students per section ranges from 12 to 24 students while the number of students per TA ranges from 33 (Gary) to 103 (An Li) and is smaller for the UTAs who only teach 2 sections and much larger for the GTAs who teach 5 sections. It is possible that an individual TA may not perform in all sections in the exact same way, based on the qualities of the students in each section, time of day, day of the week, or senior instructor with whom students go to lecture. Second, students experience clustering in sections and section-level variables other than TA type may help explain between cluster variance. Moreover, for a given number of students, more information may be obtained from larger numbers of clusters with smaller numbers of students in each cluster than a smaller number of clusters having larger numbers of students per cluster (Snijders, 2005).

The dependent variable was the final exam grade measured as a percentage of answers that were correct. The effects of seven student level (also called Level 1) variables (ACT score, number of STEM AP courses taken, number of undergraduate hours completed by the end of the current semester, minority student status (white = 0, all other self-identified races/ethnicities = 1), parent educational level (0 = no college degree, 1 = college degree by either parent), and gender (female = 0, male = 1) were considered. In addition, the effects of section level (also called Level 2) variables were considered: UTA treatment present (no = 0, yes = 1), section-mean college GPA (current college success of students), and section-mean mathematics ACT/SAT z-score (how well students were prepared for college mathematical course work). Note that these effects are correlational, not causal, in nature because they were not experimentally manipulated in this study.

The HLM process began with estimating an unconditional model or random effects analysis of variance model (ANOVA) to determine the mean course grade (the intercept) and the between-section variance. Each student's course grade ($Y_{ij,}$ with i students and j sections) consists of the overall mean course grade (γ_{00}), the deviation of the section mean grade from the overall mean grade (μ_{0j}), and the deviation of the student's grade from the mean section grade (r_{ij}).

$$Y_{ij} = \gamma_{00} + \mu_{0j} + r_{ij}$$

Every student in the same section will have the same value for μ_{0j}, also called the random effect for the intercept, because we assume that μ_{0j} varies randomly across sections with a mean of 0 and a variance of τ_{00}. Using this model, between-section variance (τ_{00}) and total variance can be obtained and used to calculate the intra-class correlation (the ratio of between class variance and total variance). The intra-class correlation (ICC) is the proportion of the variability in the course grade accounted for by the section.

The model building process continued with addition of Level-1 or student-level variables to the model. A model that simultaneously included all 6 student-level control variables: Math z-score, AP STEM courses taken, minority status (white = 0; non-white = 1), college GPA, gender (female = 0, male =1), and parent education level (0 = no college degree; 1 = college degree), was estimated. For ease of interpretation, ACT scores and college GPA were grand-mean centered but number of AP courses variable was left in the raw metric. All other variables were dichotomous (0 or 1) as described above. If all variables were included in this Level-1 model, the set of equations would be

$$Y_{ij} = \beta_{0j} + \beta_{1j}(ACT) + \beta_{2j}(AP) + \beta_{3j}(GPA) +$$

$$\beta_{4j}(Gender) + \beta_{5j}(Minority) + \beta_{6j}(ParentEd) + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

In this model, γ_{00} (the main intercept) represents the predicted final exam grade for the model's reference student without taking into account the influence of her section – a white female student having an average math z-score score, no AP courses taken in high school, average college GPA, and parents who did not graduate from college,.

Adding student level variables can help explain some amount of both the within-section variance and between section variance. A proportional reduction in variance statistic can be computed (Raudenbush & Bryk, 2002) by subtracting the residual variance of the Level-1 model from the residual variance of the unconditional model. That difference is then divided by the variance of unconditional model to obtain the proportion of variance explained by adding the student level variables.

Finally, a full two-level model was estimated in which the presence of a UTA (0= no and 1 = yes) and the number of semesters of experience the UTA or GTA leading the section possesses, as predictors of the intercept. The main goal of the full model was to identify the differential presence (if any exists) of UTA vs. GTA, and also to identify the effect size of the UTA treatment on the undergraduate students' grades.

The proportion of variance explained after the addition of these section level variables compared to the Level 1 model can be calculated. It is not likely that the addition of the section level variables would help explain more within section variance, but could help explain more between section variance. Some interactions of interest are the effects of the UTA as a peer learning assistant on the course grades of females, or minority students. Researchers have related stories about the "chilly reception" these groups have experienced in STEM courses taught in traditional large class lecture and test formats (Seymour & Hewitt, 1997; Strenta et al., 1994). Measuring the relationship between having a UTA and final exam grades for these groups could indicate whether this program was successful in helping establish a supportive learning climate for a broader group of students than may have been the case.

Historical comparisons. The Fall 2012 final exam grades were compared with historical final exam grade means (Fall semesters, 2009-2011) in a one-way ANOVA. This comparison showed the similarity between this semester's grades and grades achieved in the last few years. Any significant differences between means were reported and faculty asked about possible reasons for shifts in the mean grade earned on the common final, which has been administered for the last several years.

Normality of grade sets and statistical test assumptions. According to University of Surrey Psychology Department (2007), a distribution with skew and kurtosis values in the range +2 to -2 are near enough to be considered normally distributed for most purposes, including hierarchical linear modeling. The normality of UTA and GTA group grades was established using SPSS. Non-normality (skew and kurtosis values much greater than +2 or much less than -2) was reported and the non-normal data transformed to categorical variables to establish normality of the data set. All data was examined to make sure it did not violate assumptions of statistical tests. When necessary, nonparametric tests were used.

Research Question 4 - UTA influence on student STEM identity

As in Research Question 3, a comparison between students in UTA-led sections and those in GTA-led sections was performed.

STEM identity was measured as a composite of the responses to the eight STEM identity questions on the survey. Principal components analysis was performed on the eight items to confirm that the number of factors measured. Items that did not load on one of the factors were dropped from the analysis because they did not explain any variance for the STEM Identity factors. The Kaiser Criterion (eigenvalues greater than one) and a scree plot were used to decide how many factors to retain. The proportion of variance in the items explained by the retained factors was calculated. The composite score from the retained items was used as the dependent variable for hierarchical linear modeling, with the same covariate factors and modeling steps as for Research Question 3. The higher the composite STEM Identity score, the more strongly the undergraduate student identifies as a STEM student.

CHAPTER 4

RESULTS

This chapter first presents the descriptive analysis of the groups of interest in this study: the four chemistry faculty instructors, the nine teaching assistants, and the hundreds of undergraduate general chemistry students enrolled in CHEM 201, including student attrition analysis. Then data preparation and instrumentation validation procedures are described. Finally, the statistical and qualitative analyses are performed and the results obtained for each research question addressed by this study are reported.

Descriptions of Study Groups

Chemistry Faculty

Four chemistry instructors (three tenured professors and a fifth year graduate student, identified as numbers 1 through 4 in Table 4-1 and termed 'senior instructors' by the Chemistry Department) taught the lecture portion of CHEM 201. Each senior instructor taught one large lecture section of CHEM 201 with approximately 200 students which met three days per week (Monday, Wednesday, Friday) for 50 minutes per day or twice per week (Tuesday, Thursday) for 75 minutes per day. The three tenured professors (1, 2, and 3) had taught sections of this course before this semester. The graduate student senior instructor was teaching this course for the first time, although she had instructional experience as a graduate teaching assistant (GTA) and National Science Foundation Graduate STEM Fellow in K-12 Education at the University of Louisville.

The three tenured professors' teaching experience ranged from 12 to 30 years and they had been designated undergraduate advisors in the chemistry department. Two of the instructors (1 and 3) had been chosen by students as Faculty Favorites in recent years. The four senior instructors used a common textbook, written by one of the senior instructors (3), and they created and used a common final exam. While each senior instructor taught his/her lecture section independently, they did meet before the semester to plan TA assignments and discuss course objectives, opportunities, and challenges, to minimize any substantial differences in course experiences between their lecture sections. The senior instructors worked collaboratively with undergraduates teaching assistants (UTAs) in weekly planning for recitation sections. Typical for CHEM 201 courses offered in the last five years, the senior instructors met with GTAs only occasionally during the semester, at the request of the GTA or when a problem was brought to the attention of the senior instructor.

Teaching Assistants

Nine teaching assistants (3 GTAs and 6 UTAs) were assigned to the 29 50-minute recitation sections of CHEM 201. Each GTA led recitation sections for 2 senior instructors so that the recitation sections were balanced across the four senior instructors to mitigate instructor effect. Most UTAs led two recitation sections for one of the senior instructors and the UTA-led recitation sections were also balanced across the four senior instructors. All TAs were young adults between the ages of 19 and 26. The UTAs were U.S. born and traditionally aged undergraduates including two sophomores and four seniors. According to the senior instructors, the GTAs chosen were typical for this doctoral-granting chemistry department: young adults having finished their undergraduate degree in the last 5 years and now working on their doctoral degree in the chemistry department. Two of the three GTAs had completed their undergraduate degree in another country and were not native speakers of English; however the graduate school requires international GTAs to demonstrate proficiency in spoken English before beginning to teach. The chemistry faculty and I also observed that the GTAs were acceptably proficient in spoken English.

Each graduate teaching assistant met with five recitation sections of 25 students each, one veteran UTA met with 4 sections of 25 students each, and the remaining five UTAs met with 2 sections of 25 students each. The CHEM 201 recitation sections were scheduled at various times during the day and multiple days throughout the week. The recitation sections had been scheduled in the university course catalog several months in advance of students registering for the course. The assignment of specific UTAs and GTAs to the recitation sections was made by the four senior instructors a few days before classes began and was mainly based on the TA’s availability along with balancing teaching assignments across senior instructors, week days, and day time hours. To address any selection bias, I examined the recitation section schedule and found that UTAs and GTAs were scheduled evenly over the course of the week and over time of day. Descriptions of the TAs involved in this study are shown in Table 4-1.

Student Attrition

A total of 711 students were enrolled across the four lecture sections and 29 recitation sections of CHEM 201 at the start of the Fall 2012 semester. There were 369 students in GTA-led sections and 342 students in UTA-led sections. A total of 117 students (16.5%) withdrew from the course or failed to finish the course by not taking the final exam. Fifty-nine students were from GTA recitation sections (16%) and 58 students from UTA recitation sections (17%). The students who withdrew or did not take the final exam had an average Math ACT score of 25.2, an average high school GPA of 3.01, an average college GPA of 2.01 and had earned an average of 40 hours college credit before Fall 2012. For analysis, data from 594 students, 310 in GTA-led sections and 284 students in UTA-led sections, was examined

During the last weeks of the semester, an end of course survey was administered to the 573 students who were still in the course and were given the survey in their recitation section. A GTA forgot to give the survey for one of her five recitation sections (21 students). Seventy-two percent of the students (414 students) responded to the end of course survey. To check for selection bias, the percentage of students responding to the survey from the 14 GTA recitation sections (70%) was similar to the percentage of students responding from the 14 UTA sections (75%).

Students

Characteristics of the students who enrolled in and finished CHEM 201 are shown in Table 4-2. Of the 594 students who finished the semester and took the final exam, 310 students were in 15 GTA recitation sections and 284 students were in 14 UTA recitation sections. There were more males than females in this course and the majority of students had at least one parent with college experience. The students in the GTA and UTA groups were very similar in terms of the academic covariates of parent college experience, math z-score, current college GPA and the number of advanced placement STEM courses taken in high school, so a potential selection bias was avoided.

Table 4-1

Teaching Assistants for CHEM 201 Recitation Sections

TA	Gender	Class Standing	Previous teaching experience (semesters)	Senior Instructor	Class Meeting Time/Day
			GTA Sections		
An Li	F	5th year	2	2	Tuesday 9-9:50 am
				4	Thursday 2: 2:50 pm
				4	Thursday 3-3:30 pm
				2	Wednesday 10-10:50 am
				2	Wednesday 12-12:50 pm
Erin	F	1st year	1	3	Friday 8-8:50 am
				3	Friday 10-10:50 am
				2	Friday 12-12:50 pm
				3	Thursday 1-1:50 pm
				3	Thursday 8-8:50 am
Rakesh	M	1st year	0	1	Friday 3-3:50 pm
				1	Friday 9-9:50 am
				1	Friday 10-10:50 am
				4	Monday 12-12:50 pm
				1	Thursday 8-8:50 am
			UTA Sections		
Brandy	F	Senior	1	1	Thursday 9 am
				1	Wednesday 12-12:50 pm
Gary	M	Soph.	0	3	Monday 8-8:50 am
				3	Thursday 1-1:50 pm
James	M	Senior	1	1	Friday 1-1:50 pm
				1	Wednesday 2-2:50 pm
Jason	M	Soph.	0	2	Tuesday 1-1:50
				2	Wednesday 1-1:50
Lisa	F	Senior	1	4	Friday 12-12:50
				3	Monday 9-9:50
				3	Monday 12-12:50
				4	Wednesday 12-12:50
Stacy	F	Senior	1	2	Tuesday 8-8:50
				2	Thursday 1-1:50

Table 4-2
Demographics of CHEM 201 Student Sample in Each TA Group

TA Group	n	Male	Non-white	Parent college experience	ACT/SAT Math Z-Score (SD)	College GPA (SD)	Number of STEM AP Courses (SD)
Combined	594	64%	21%	77%	1.15 (0.80)	2.83 (0.83)	0.92 (1.22)
GTA	310	64%	25%	75%	1.1 (0.80)	2.81 (0.85)	0.84 (1.21)
UTA	284	64%	17	79%	1.2 (0.80)	2.86 (0.80)	1.00 (1.23)

Data Preparation and Instrument Validation Procedures

Data Preparation

Quantitative data were collected from the senior instructors (final exam and course grades, and attendance at recitation sections), the university institutional research office (academic achievement and identity covariates: ACT/SAT scores, race, gender, college GPA, college hours earned, parents' education level, and number of STEM-related AP courses taken), and the CHEM 201 students (Course Experience Survey).

Data assumptions. The skew and kurtosis statistics were within commonly acceptable range (-2 to +2) for many of the outcome variables and covariates used in the study, indicating normal data distribution (see Table 4-3). One exception was recitation attendance, in which almost half the students attended all of the 14 recitation sessions and 7% of students attended less than half the sessions. The senior instructors communicated an expectation using participation points that students needed to be present at 80% of the recitation sections. Therefore, the students who attended at least 80% of the recitation sections were designated “high attenders” (n = 473), students who attended at least 50% but less than 80% of the sessions were called “medium attenders” (n = 78), and students who attended less than 50% of the session were “low attenders” (n = 43). These categories were dummy coded for use in statistical models or were used to split the data file into attendance groups.

Table 4-3
Psychometric Properties of Study Variables for the Student Sample

Variable	n	M	SD	Range		Skew	Kurtosis
				Potential	Actual		
Final Exam score (%)	594	55.6	20.4	0 - 100	5 - 100	.15	-.85
Recitation attendance (%)	594	86.6	20.1	0 - 100	0 - 100	-2.1	4.5
Math Z-Score	569	1.15	0.78	-	-1.15 - 2.81	-.09	-.21
College GPA	594	2.83	0.83	0-4.00	0-4.00	-.69	.29
TA Impact on Academics	411	28.5	6.5	0 - 39.1	7.8 – 39.1	-.47	-.24
TA Rapport Skills	411	10.3	1.9	0 -13.6	2.7 - 13.6	-.41	.51
Student STEM Recognition	411	10.2	1.9	0-12.5	2.5 -12.5	-.93	1.2
Student STEM Interest	411	10.0	2.0	0-12.2	2.5-12.2	-1.1	1.4

Missing data. There were two types of missing CHEM 201 student data on the student survey: (a) non-response for individual items which were aggregated into constructs as described below and (b) missing values on a single variable (e.g., ACT/SAT Math score or Parent’s education level). For missing data of the first type, less than 5 responses (1%) were missing from any one item on the Course Experience survey, so those missing responses were replaced with the sample-based mean for that item. For missing data of the second type, while the majority of the 594 students (89%) had ACT Math scores recorded with Institutional Research, 68 students lacked ACT scores. Of those 68 students, 43 students had SAT Math scores instead. Students’ ACT and SAT

Math scores were standardized by calculating a z-score using each test's published mean and standard deviation for the national testing population. The quantity *z* represents the distance between the raw score and the population mean in units of the standard deviation. For 2012, the ACT Mathematics subtest national average was 21.1 with a standard deviation of 5.3 and the ACT Reading subtest national average was 21.3 with standard deviation of 6.2 (ACT, Inc., 2012). The SAT Mathematics subtest national average was 514 with standard deviation of 117 and the SAT Verbal subtest national average was 496 with standard deviation of 114 (College Board, 2012). Because the percentage of missing ACT/SAT math scores was small (<5%), list-wise deletion of missing data points for HLM analysis was accepted to be a reasonable solution.

Eighty-three students (14%) were missing both father and mother education level data. For the remaining 511 students, all the students had mother's education level but 18 students were missing the father's education level. The parent's education level was evaluated by using the organizing the data into sets: no parent had attended college (even if the mother was the only parent known) (24%), one (either) parent had attended college (28%) or both parents had attended college (48%). These three sets were dummy coded for HLM analysis. Thus, depending on which variables were in the model for a specific HLM analytic step, the number of cases in each of those analyses would vary slightly depending on the variables included and those missing that variable.

There was one instance of missing TA data. One of the UTAs did not submit five of the written reflections from the semester to either the electronic database organized for that purpose or the faculty member in charge of evaluating them. Finally, the UTA admitted to losing the data. Rather than have him re-create reflections that were meant to represent his ability and disposition at certain points in time, I used the reflection he did submit for the last seminar, which was a retrospective account of his content knowledge and peer learning assistance skills development over the entire semester.

Instrument Validity and Reliability

Undergraduate course experience survey. Validation of the Undergraduate Course Experience Survey (Appendix A) was conducted by performing principal component analysis (PCA) on the survey items to extract orthogonal variables from the multiple items used on the survey measuring student perception of the TAs and the student STEM identity aspects. Some items on the survey were not designed to elicit answers to the research questions about TA impact on learning or student STEM identity. For example, items 31-39 were variables that could be used to characterize sections of students, so were not used in the PCA. Some items were not clearly understood by students; for Items 29 and 30, students were asked to respond to one item or the other, depending on their intended major, but many students responded to either both items or no items, so the questions were not used in further analyses. Items 20 and 21 were also unclear for students, with no established definitions for "interacting with the TA" and "mandatory or voluntary" attendance. Therefore, PCA was initially performed with Items 1-19 and Items 22-28. Items 22-28 focused on student STEM identity and were not expected to load with Items 1-19, which focused on TA effectiveness. Varimax rotation was chosen for the cleanest interpretation of components, which were retained if their eigenvalues were greater than unity (Kaiser-Guttman retention criteria).

Items 9, 12, 17, 18, 19, and 27 did not meet criteria of loading (correlation between the variable and the component) at least .40 on any one component with minimal overlap on (or correlation with) any other component (Stevens, 2009), so these items were dropped from the analysis. The remaining items were used in performing a final PCA. The items loaded onto one of four components with eigenvalues greater than 1 as shown by loading coefficients in Table 4-4. These four factors explained 69.5% of the variance in student responses from the course survey for undergraduates. These results indicate that the survey measured four factors:

1. Perceived TA Impact on Academic Success (10 items, $\alpha = .95$)
2. TA Rapport-Building Skills (4 items, $\alpha = .77$)
3. Student STEM Recognition (3 items, $\alpha = .84$)
4. Student STEM Interest (3 items, $\alpha = .82$)

The overall reliability coefficient (Cronbach α) for the 20 item survey was .91. The reliability coefficients were well within the norms of social science research.

TA Impact on Academics scores correlated strongly ($r = .699$) with TA Rapport-Building Skills, meaning that the higher students rated their TA's ability to impact their academic success in CHEM 201, the higher students rated the rapport-building skills of their TA. TA Impact and TA Rapport correlated significantly but much less strongly with both Student STEM Recognition and Student STEM Interest (all $r < .177$). Not surprisingly, student-reported ACT math subtest scores correlated significantly but not strongly with TA Impact ($r = .139$) and TA Rapport ($r = .145$). ACT math scores correlated a little more strongly with Student STEM Recognition ($r = .262$) and Student STEM Interest ($r=.223$). The higher the students' academic ability, the higher they rated their TA's ability to positively impact student academic success, their TA's skill in building rapport in the classroom, and the more the students recognized themselves and were recognized by others as a “science person” or a “math person” and were interested in science and mathematics. These relationships, although significant, were small in magnitude however.

Table 4-4

Factor Loadings for Principal Component Analysis with Varimax Rotation of Undergraduate Course Survey Items

Survey Item	Component 1	2	3	4
1 Course was enjoyable	**.723**	.357	-.002	.016
2 Course was valuable experience	**.850**	.240	.014	-.031
3 TA had strong content knowledge	**.709**	.033	.068	.117
4 TA gave clear explanations	**.824**	.171	.061	.027
5 TA led effective discussions	**.783**	.291	.037	-.023
6 Overall TA was excellent	**.846**	.312	.057	.008
7 TA gave choices for learning	**.697**	.434	-.018	.041
14 Success in future courses due to TA	**.781**	.316	.055	.024
15 Grade is higher due to TA	**.781**	.267	-.007	.052
16 Understand more content due to TA	**.820**	.223	.003	.024
8 Able to be open with TA	.294	**.732**	.113	-.033
10 TA encouraged questions	.355	**.670**	.102	.064
11 TA cares about me	.404	**.722**	-.014	.078
13 TA tries to understand me	.293	**.599**	.044	.109
22 I am a science or math person	.033	.078	**.884**	.195
23 Family/friends think I am science or math person	.043	.030	**.905**	.117
24 I want others to see me as science or math person	.044	.085	**.711**	.365
25 I am interested in experiments	.093	.004	.114	**.843**
26 I am interested in talking to others about STEM	-.004	.090	.191	**.865**
28 I want to know more about science or math	.005	.076	.399	**.738**

CHEM 201 final exam. One outcome measure used to measure student achievement in this study was the common final exam given to all 594 students enrolled in CHEM 201 at the end of the semester. The exam was given over a two hour block of time during exam week and consisted of 35 multiple choice questions. The first eighteen questions on the exam assessed student knowledge of material that had been presented in lecture and practiced in recitation sections the last 4 weeks of the semester and had not been previously assessed by a mid-term exam. These topics included electron configurations, chemical bonds, bonding theory, and molecular geometry and molecular orbital theory. The last seventeen questions assessed concepts and processes that had been previously assessed by at least three midterm exams, including atomic structure, thermochemistry, gas laws, chemical reactions, and stoichiometry. Twenty-nine questions assessed conceptual knowledge and 6 questions required students to calculate the correct answer. A periodic table and a chemical reference page were included for assistance in answering questions. This exam was created by the four senior instructors, and similar versions had been used for the past several years. Because almost 600 exams had to be graded quickly in order to submit final grades before the end of the semester, a multiple choice format was chosen. The multiple choice exam responses were collected on scan-able forms and grading was performed by a computer.

For the past three years, a similar final exam (both in format and content assessed) was given to CHEM 201 students and the average percentage correct ranged from 58.0% in 2009 to 62.4% in 2010. The average exam score for 2012 was 55.6% with a standard deviation of 20.4%. Standard deviations were not available for the past exams, an obstacle in performing a one-way ANOVA to check if these scores differed significantly.

However, if it was assumed that increasing numbers of students each year took the exams, as was related by senior instructors, (450 in 2009, 500 in 2010, and 550 in 2011) and with a similar standard deviation as the 2012 exam (SD = 20%), then a one way ANOVA would show that final exam scores differed significantly across the years, before the 2012 exam, $F(2, 1497) = 5.231$, $p = .005$), and with the inclusion of the 2012 exam ($F(3, 2090) = 11.256$, . $p < .001$).

CHEM 201 TA pre/post content knowledge test. A twenty-three question multiple choice test, similar to the content and format of the student final exam, was administered to the CHEM 201 TAs during the first and last weeks of the semester in which they were a TA. The content knowledge test included 4 questions requiring TAs to perform calculations and 19 conceptual questions, over topics similar to the student final exam. To minimize instrumentation threat to internal validity, an identical exam was given pre and post semester. To minimize testing threat to internal validity, the TAs were not told which questions they missed on the pre-test nor were they given access to their exams or answers after the pre-test, so brief exposure to the test minimized ‘priming’ the TAs for the content they would need to know. Neither ceiling nor floor effects were observed in the test data.

Data Analysis and Results

Research Question 1a – UTA Content Knowledge Growth

The purpose of Research Question 1a was to investigate how the content knowledge of the UTAs changed over the semester. In addressing this research question, two types of analyses were performed: a quantitative analysis of a pre- and post-CHEM

201 content knowledge test given to all TAs, and qualitative analysis of six teaching reflections that the UTAs wrote as part of their bimonthly pedagogy seminar.

Pre/Post content knowledge test. The UTAs and GTAs took the pre-test during the first week of classes, and then took an identical post-test fourteen weeks later during the last week of classes. The senior instructors gave the TAs their raw scores after the pre-test upon request, but did not give information to the TAs about which questions they missed nor did they give the TAs access to the exam after administration. Scores from the pre and post CHEM 201 content knowledge test are listed in Table 4-5. These scores denote the percent of questions correct out of 23 multiple choice questions. Seven of the nine TAs had positive gains in knowledge, while the other two TAs scored the same on pre and post-tests. The three TAs (one GTA and 2 UTAs) who scored the lowest at the beginning of the semester made the greatest score gains by the end of the semester. At $\alpha = .05$, the Related-Samples Wilcoxon Signed Ranks Test statistic was significant, indicating that on average, the combined TA groups significantly improved their knowledge content by the end of their recitation teaching assignment for CHEM 201.

Another way to analyze the change in scores was to calculate normalized gains (Hake, 1998). A normalized gain is the gain achieved as a proportion of the potential gain. This calculation takes into account widely varying pretest scores in a sample. For example, in our TA sample, An Li had a 13% gain score as did Lisa and Stacy. However, An Li's pretest score (16 out of 23 correct or 70%) was much lower than that for Lisa or Stacy, so a 13% increase in the post-test for An Li resulted in 83% of the answers correct (normalized gain of 0.43) while the same percent increase for Lisa and Stacy resulted in 91% of the answers correct (normalized gain of 0.60). Normalized gains are listed in

Table 4-5. Taking varying pre-test scores into account, there appeared to be no difference (e.g. similar averages) between the GTAs and the UTAs in terms of normalized gain scores.

The five most problematic questions for the TAs on pre and post-test were:

1. Oxidation number and formal charge

3. Molecular geometry

5. Polarity

8. Molecular orbital theory

15. Balancing a redox reaction and calculating molar yield.

The bolded question numbers in Table 4-5 indicate where TAs answered the same questions incorrectly on both the pre and post-tests. The most problematic questions for the TAs, Question 8, requiring knowledge of molecular orbital theory, and Question 15, balancing a reduction-oxidation chemical equation, were missed by four out of the nine TAs on both the pre and post-tests.

Phenomenological descriptions of UTA content knowledge. To answer the research question concerning UTA content knowledge growth, the UTA was the unit of study or the case. Qualitative data in the form of first-person UTA reflections, faculty interview, and research observation field notes gathered for the six UTA cases were mined by the author for instances of content knowledge growth and are reported below. These reported instances were reviewed for accuracy by one of the senior instructors familiar with the CHEM 201 UTA reflections. Six guided reflections that asked the UTAs to relate what they were learning in pedagogy seminar to their recitation section practice were collected from the UTAs over the course of the semester. Because the reflections were written over regular intervals during the semester, they served as a chronological account of how the UTAs changed in content knowledge and application of learning theories for their own learning goals over the course of the semester. Similarities and differences between the UTAs' descriptions of content knowledge growth were summarized.

Table 4-5
Pre and Post Content Knowledge Test Scores for TAs

TA	Pre %	Post %	Gain score %	Normalized gain scores	Questions missed Pre-semester	Questions missed Post-semester
GTAs						
An Li	70	83	13	0.43	1, **3, 5**, 6, 9, **19**, 22	**3, 5**, 7, **19**
Erin	48	74	26	0.50	1, 3, 4, 5, 6, 7, **8, 9**, 14, **15**, 16, 21	**8, 9**, 12, **15**, 18, 19
Rakesh	74	78	4	0.17	**1, 14, 15**, 18, 22, 23	**1**, 3, 5, **14, 15**
UTAs						
Brandy	57	91	34	0.80	5, 7, 9, 10, 11, 14, 16, 18, 19, **21**	15, **21**
Gary	87	87	0	0	**1, 8, 22**	**1, 8, 22**
James	78	78	0	0	**3, 8**, 10, **15,** 19	**3**, 5, **8**, 12, **15**
Jason	57	74	17	0.40	**2, 6, 7, 8**, 10, 14, **15**, 16, 19, 21	1, **2, 6, 7, 8, 15**
Lisa	78	91	13	0.60	2, 8, 15, 19, 21	11, 23
Stacy	78	91	13	0.60	1, 7, 8, 15, 21	3, 19

Brandy. The semester in which this study took place was Brandy's second semester as a PRIMES UTA, but her first time as leader of a recitation section. Last year as a junior, she had taught in a laboratory section for general chemistry. Brandy had the second lowest score on the CHEM 201 pre-test, but she did not mention anything in her reflections about having any trouble with the content needed to assist her less-experienced peers in the recitation section. Her strategy was to prepare activities in advance: "By preparing lessons, I review the material several times and increase my own knowledge in the process" (Reflection 6). Her effective review of the material was evident as she gained 34 percentage points on the post-test, or a normalized gain of .80, the largest gain of any UTA or GTA. She shared that asking questions of students to help them think through a process helped refresh her own memory of the process, a metacognitive approach she used in her own learning (Reflection 2). The senior instructor with whom she worked related that Brandy overcame a few misconceptions at the start of the semester, and that in collaboratively preparing activities for the recitation section, the senior instructor was convinced that Brandy's content knowledge had increased over the course of the semester.

Gary. This study took place during Gary's first semester as a UTA. He was actually a sophomore chemical engineering major, but had been very successful in general chemistry courses the previous year. In his reflections, Gary was very confident in the depth of his own general chemistry content knowledge and did not express a desire to deepen that knowledge or acknowledge that the UTA experience changed his content knowledge in any way. His pre and post content knowledge test scores were identical and he missed the same three questions each time he took the test. The senior instructor who worked with Gary commented that he was a reliable UTA with a good foundation in the chemical concepts need for teaching CHM 201.

James. As a senior, this was James' second semester as a PRIMES UTA, and his second as a recitation section leader. His previous experience was as a PRIMES recitation leader for the second semester of organic chemistry and in other teaching capacities within the chemistry department before implementation of PRIMES. He was well-regarded as an experienced and capable TA by the senior instructors. James thought that the UTA experience "greatly improved my content knowledge obviously," (Reflection 6) although his content knowledge test score did not change from pre to post-semester. He missed five questions on the pre-test and five questions on the post-test, with three questions in common. James' philosophy for learning was captured by the statement "We are going to be teachers and students for our entire lives, so we might as well get good at teaching and learning" (Reflection 6). James' senior instructor was very comfortable with his depth of knowledge and experience in teaching.

Jason. Jason, a sophomore, was in his first semester as a PRIMES UTA for CHEM 201, having been a successful student in CHEM 201 just the year before Jason's self-described "passion for spreading the amazing possibilities of this universe and its properties" (Reflection 6) emerged from his qualitative data as a self-evaluated important characteristic for teaching assistants to possess, in addition to a solid content knowledge foundation. Although Jason's initial content knowledge score was relatively low compared to the other TAs, his post-test score improved considerably. Of the six questions Jason missed on the post-test, five of those questions he also missed on the pre-test. Observing Jason in two review sessions, Jason seemed sure of his content knowledge and he did not pass along misconceptions nor accept incorrect answers to his questions, although his questions to students were not as deep or divergent as some of the other more experienced UTAs.

Lisa. Similar to James, Lisa was a senior who had taught a chemistry recitation section as a PRIMES UTA in the previous semester, but her experience was in the second semester of general chemistry. Maybe her previous preparation for a general chemistry course helped her initial content knowledge because Lisa had one of the higher scores on the content knowledge pre-test. Moreover, she demonstrated a gain by the end of the semester. The two questions that she missed on the post-test were not ones she had missed on the pretest. She reflected, “I have found that my understanding has significantly improved through teaching others and I would like my students to experience that” (Reflection 4). Lisa did not take any science or math course during the semester she was the CHEM 201 UTA because she had completed all her science requirements and was finishing electives coursework while applying to graduate school in chemistry or chemical education. In terms of content knowledge change, she reflected, “I know that I have changed, but I feel like I haven’t been able to experience the changes, because I have not been exposed to any math or science courses of my own this semester. I think that my biggest changes have probably been in my understanding of and the way I look at new science concepts. I feel that I would be better able to adapt and wrap my mind around new topics after my exposure to so many different types of thinking” (Reflection 6). The senior instructor with whom Lisa worked had mentored Lisa before and continued to be impressed by her content growth.

Stacy. Stacy, a senior, had previous experience as a PRIMES UTA in a general chemistry laboratory section. Early in the semester, Stacy realized that "if you are able to explain the material to someone else, you confirm with yourself that you are able to work the problem and understand the concept" (Reflection 4). By the end of the semester, she believed that she had grown as a scholar by being a UTA and she was "spending more time making sure I know the concept of the questions rather than just knowing the answers" (Reflection 6). Preparing activities and questions for her students helped her deepen her own knowledge and forced her to think about her own thinking. Similar to Lisa, Stacy missed only two questions on the post-test after missing six on the pre-test (normalized gain = 0.60), with no repeated missed questions. Although Lisa and Stacy had each missed six questions on the pre-test (with three questions in common), they missed only two questions on the post-test, with no commonalities with the missed pre-test questions or each other. The senior instructor with whom Stacy worked mentioned the great amount of time and effort Stacy expended on preparation for and reflection on the recitation section activities.

Summary of similarities and differences. This summary is built upon a comparison and contrast of content knowledge gains and self-learning approaches in each of the UTA cases. Four of the UTAs, Brandy, James, Lisa and Stacy, had previous experience as PRIMES UTAs and all were seniors. Of those four, Lisa, Stacy and Brandy scored a 91 (the highest score) on the post-semester content knowledge test. James and Gary demonstrated no content knowledge growth according to the content knowledge test. Of the two UTAs with no previous experience and sophomore standing, Gary missed only 3 questions each testing event, and Jason scored the lowest of any of the UTAs on the post semester test.

All the UTAs except Gary wrote at some time during the semester that they felt they had deepened their content knowledge as a result of their UTA work. All of the UTAs except Gary and James did improve their content knowledge by the end of the semester as measured by the content knowledge test. The senior instructors were very comfortable with the UTAs content knowledge growth or with the starting content knowledge that the UTAs possessed. When observed in the recitation sections at the beginning and end of the semester, none of the UTAs had trouble answering students' questions, planning engaging learning activities, or connecting current concepts with familiar topics taught in CHEM 201.

All of the UTAs mentioned in their reflections that learning about how others learn impacted their approach to learning in their own classwork and in preparing for their peer learning assistance work. UTAs used the vocabulary and conceptual knowledge that had been presented and discussed in the accompanying pedagogy seminars to write about how their students learn. They had put some of the ideas they learned in the seminar into practice with their students and for themselves in their own science courses and were able to reflect on how that practice may have impacted content knowledge and their own ability to learn.

Research Question 1b – UTA Approaches to Self-Learning

The purpose of Research Question 1b was to investigate how the UTAs approached self-learning over the semester. In addressing this research question, phenomenological descriptions of self-learning approaches taken by the UTAs were crafted from six reflections that the UTAs wrote as part of their bimonthly pedagogy seminar.

Phenomenological descriptions of UTA approaches to self-learning.

Brandy. At the beginning of the semester, Brandy wrote that she had come to understand as a student and as an instructor that the best way to learn something was not to just listen to someone else's explanation, but to come to an understanding on your own, with help from a teacher's guiding questions (Reflection 2). For Brandy, the process of learning something new or coming to a deeper understanding about chemistry concepts was aided by her persistent review of material in preparation for crafting the guiding questions she used to help her students learn (Reflection 6).

Gary. Gary related a story about how a student's questions and his response to them helped to create a comfortable and engaging environment that he felt was necessary for learning. Therefore, one might assume that for Gary's own approach to learning, a comfortable and engaging environment is helpful. Gary did not relate how he might make that happen for himself or if he finds that kind of environment in his own coursework (Reflection 1). Like Brandy, Gary connected guiding questions with helping his own learning as well as helping his students learn to make sense of a concept (Reflection 4). Gary acknowledged that learning about how other people learn during his UTA experience helped him to learn more effectively in his own studies(Reflection 6).

James. From his previous experiences as a UTA and chemistry tutor, James found that connecting previously learned concepts with new concepts he wanted to learn was crucial for success (Reflection 1). James thought he should be determined in his approach to problem-solving, and if he didn't initially understand something, he "did not believe in ‘folding’ but using what knowledge you do have to answer a question that you may feel very uncomfortable about” (Reflection 4). He found it helpful for his own learning to see both sides of learning, as an instructor and as a student. He considered UTAs to be in a unique position as a bridge between instructor and student (Reflection 6).

Jason. Jason thought that the “review of chemistry principles helped me in other science courses I am currently taking” (Reflection 6). He believed that the UTA program helped him “learn to gather the most important information” (Reflection 6) in order to learn more effectively. He looked forward to improving his learning ability, “that way I can teach better and pass on more information” (Reflection 6).

Lisa. Preparation for teaching helped Lisa as a learner: “I have found that my understanding has significantly improved through teaching others” (Reflection 4). The use of questions as a learning tool was also a valuable skill for Lisa’s own learning: “I have found that as a student and as a teacher, I have a constant flow of questions that go through my mind. These questions help me to relate the material to older topics and to prepare myself and the students for topics that will come up in the future” (personal communication).

Stacy. Stacy reflected often on her own experiences as a student and how the UTA experience confirmed the usefulness of her learning approaches or introduced her to new and effective learning strategies for her own work. As a student, formative assessments helped her focus on the knowledge she did not yet know (Reflection 5). Additionally, “being a UTA and having to prepare and work with students weekly has allowed me to improve my ability to organize and express my own thoughts” (Reflection 6). Finally, as a UTA,

> “I have thought more about my own thinking and I have spent more time making sure that I know the concept of questions, rather than just the answers. I think that I have had to make sure that my understanding of the material is clear before I bring that knowledge to students and this has helped me to deepen my own understanding of the material. I was very interested and surprised at times by the many approaches that instructors use to help students learn material. (Reflection 6)

Summary of similarities and differences. All of the UTAs used at least one of the learning strategies discussed in the workshop and seminar for their own learning. Some of the UTAs wrote about using these strategies previously for effective learning, but as a result of the workshop and seminar training, now knew these strategies had names and were used intentionally by teachers to assist students in learning.

Summary of results for research questions 1a and 1b. Most TAs were able to answer more questions correctly on the post- test and no one scored lower on the post - test compared to the pre-test; however, four of the TAs (2 GTAs and 2 UTAs) still answered less than 80% of the 23 multiple choice questions correctly on the post-test. The fundamental chemical concepts of molecular orbital theory and redox reactions posed a particular problem for many TAs. All UTAs except one felt that they increased the depth of their content knowledge as a result of their UTA work. All UTAs stated that the ideas presented to them in pedagogy class about how people learn had, in turn, helped them to learn more deeply and more efficiently. Some of the UTAs were encouraging their students to study as if preparing to teach others, for that was the most effective way in their opinion to deeply learn new content and be able to connect new science concepts with familiar ones.

Research Question 2 – UTA Peer Learning Assistance Skills

The purpose of Research Question 2 was to evaluate growth in the peer learning assistance skills of the UTAs that may be impacting student learning, achievement, and course environment in recitation sections. This section triangulated results from the (a) the CHEM 201 students' responses on the TA Impact scale and TA Rapport-Building Skills scale from the Undergraduate Course Experience Survey, with (b) UTAs' six reflections on their peer learning assistance skills written over the course of the semester; (c) qualitative evaluations of UTA peer learning assistance skills from the senior instructors who worked closely with the UTAs, and (d) classroom observations made at the beginning and end of the semester by the author. The Undergraduate Course Survey results will be discussed first, and then a phenomenological description for each UTA will be constructed using the reflections, evaluations and observations collected for each UTA. Comparisons of the individual descriptions will be presented, and then the quantitative survey results will be reviewed in light of the qualitative findings in a summary section.

Undergraduate course experience survey. The results from the survey were reported as scores on the Perceived TA Impact on Academic Success Scale and scores on the TA Rapport-Building Skill Scale as described in Chapter 3. An independent-samples t-test was performed on both TA Impact score and TA Rapport score for GTA and UTA groups. Significant predictors of TA Impact and TA Rapport were discovered through linear regressions.

Perceived TA Impact on Academic Success Scale. The responses to the ten items loading onto the first component (termed Perceived TA Impact on Academic Success or TA Impact), listed in Table 4-4, were weighted by the loading factors and summed for each student to create a TA Impact score. The combined average TA Impact score was 28.51 out of a possible 39.07 (10 weighted items with a maximum rating of 5 for strongly agree).

TA Rapport-Building Skill Scale. The responses to items loading onto the second component (termed TA Rapport-Building Skills or TA Rapport), listed in Table 4-4, were weighted by loading factors and summed for each student to create a TA Rapport-Building Skills score. The combined average TA Rapport score was 10.29 out of a possible 13.62 (4 weighted items each with a maximum rating of 5 for strongly agree.)

Comparison of means. An independent-samples t-test was conducted on the TA Impact and TA Rapport scores to evaluate whether the GTA mean was significantly different from the UTA mean (Table 4-6). Levene's test for equality of variances was significantly nonequal at $p < .05$ for TA Impact; therefore the corrected degrees of freedom and t statistics were reported for TA Impact assuming the variances were not equal. The analyses indicated that students in UTA sections rated their TAs significantly higher on both TA Impact and TA Rapport than did students in GTA sections. Cohen's effect size value (d= .53) suggested being in a UTA section had a moderate practical significance for TA Impact score and a small to moderate effect (d = .38) for TA Rapport score. (Cohen, 1988)

Table 4-6
Comparison of Mean TA Impact and TA Rapport Scores

	GTA		UTA					
Variable	*M*	*SD*	*M*	*SD*	*df*	*t*	*p*	Cohen's d
TA Impact	26.79	6.64	30.15	6.02	399	5.355	<.001	0.53
TA Rapport	9.92	1.92	10.64	1.89	410	3.856	<.001	0.38

Linear regression analysis – TA Impact. Student-reported covariates such as ACT Math score, parent education level, number of STEM AP courses, and gender, as well as TA Rapport score, STEM Recognition, STEM Interest, and TA type were tested as predictors of TA Impact score using a linear regression model with backwards entry, where the software chose the sequence of variables to include based on those explaining the most variance coming first. TA Type, TA Rapport scale score and student gender significantly predicted TA Impact on Academics score (Table 4-7). Students having a UTA and rating their TA higher in TA Rapport score gave their TAs a higher TA Impact score; however females rated their TAs lower on the TA impact score than did males.

Table 4-7
Predictors of TA Impact on Academics

	Unstandardized Coefficients		Standardized Coefficients		
Predictors	**B**	**Std. Error**	**Beta**	**t**	**Sig.**
Constant	2.603	1.445		1.802	.073
TA Type (GTA = 0; UTA=1)	2.158	.535	.160	4.034	.000
TA Rapport	2.381	.139	.683	17.142	.000
Gender (0=male; 1=female)	-1.912	.542	-.137	-3.530	.001
Number of STEM AP courses	.469	.185	.099	2.538	.012
R^2	.561				
F	94.11				.000

Linear regression analysis – TA Rapport. Student-reported academic covariates such as ACT Math score, parent education level, number of STEM AP courses, and gender, identity variables STEM Recognition and STEM Interest and TA type were tested as predictors of TA Rapport score using a linear regression model with backwards entry, where the software chose the sequence of variables to include based on those explaining the most variance coming first. TA Rapport is a construct associated with peer learning assistance skills (“My TA encourages me to ask questions”) as well as a possible influence on STEM identity of students (“my TA cares about me as a person”). As shown in Table 4-8, TA Type and STEM Recognition score significantly predict TA Rapport Building skills.

Table 4-8
Predictors of TA Rapport Building Skills

	Unstandardized Coefficients		Standardized Coefficients		
Predictor	**B**	**Std. Error**	**Beta**	**t**	**Sig.**
Constant	7.576	.652		11.61	.000
TA Type (0=GTA; 1=UTA)	.687	.218	.178	3.15	.002
Student STEM Recognition	.230	.062	.209	3.70	.000
R^2	.520				
F	13.52				.000

Phenomenological descriptions of UTA peer learning assistance skills. This section will describe the peer learning assistance skills growth for each individual UTA, based on the six reflections written by the UTAs over the course of the semester, my observations of recitation sections at beginning and end of the semester, and interviews with the senior instructors at the end of the semester. Just two observations is a narrow evidence base over the semester, but the goal of the observations was to capture a description of the classroom environment and to record any examples of learning activities that the TAs were using on any given day. For the first observation, in the beginning of the semester, the day I visited many of the UTAs coincided with the session they chose to write about in their reflection for the second seminar. The descriptions of the UTA activities in my field notes closely matched the UTAs' descriptions of their recitation activities, so it was a confirmation of what had transpired in the classroom during one recitation section. The senior instructors comments were taken from an hour-long interview I had with them at the end of the semester about the UTAs. Although the senior instructors were not in the recitation section classroom with the TAs, they worked collaboratively with the UTAs on preparation of learning activities every week. Peer learning assistant skills included a focus on pedagogical topics presented and discussed in the pedagogy seminars as well as skills that have been shown in the literature to be effective for establishing a positive learning environment in undergraduate introductory science courses (Black and Deci, 2000; Lewis, 2011, Seymour & Hewitt, 1994; Tobias, 1990).

Brandy. A focus of Brandy's efforts to continually improve her teaching was learning to ask effective questions of her students to help them solve their own problems with the processes and concepts they were trying to master. Brandy began the semester by reading about different questioning strategies and developing her own questioning method. She wrote, "When I first started I wasn't sure what questions to ask, but as I've improved my method, I've managed to streamline the process" (Reflection 1). She found that when individual or small groups of students would ask her for help on a practice problem, she could ask them a series of leading questions, the kind of questions she would ask herself as she was working through the problem. The student's attention would then be focused on possible steps for solving their own problem. She realized early on that simply giving the students a quick answer when asked would not help them learn. Later in the semester, Brandy related the idea of guiding students thinking by asking a series of questions to a metacognitive strategy she used with her students called teacher-as -a-model, in which she talked out the series of questions she asked herself as an experienced problem-solver (Reflection 3). One skill that Brandy had trouble mastering was whole-class questioning. In the second reflection, she related that she tried whole-class questions with increasing wait time to ten seconds to allow her students time to think. "The same few students volunteered to answer questions while the rest of the class sat in silence and answered nothing even if they knew the answers (and based on their earlier quiz results, I know they knew the answers)" (Reflection 3). By the end of the semester, Brandy was still uncomfortable with whole-class questioning as she wrote, "I have yet to find a way to get them to respond to me in a lecture-type setting" (Reflection 6). Brandy listed patience, a positive attitude, and the ability to establish trust with students as the most important characteristics needed to help students learn science, and both her senior instructor and I observed these qualities in Brandy's interactions with her students throughout the semester (Observations, Senior instructor interview). Her students sat closely together in the front of the classroom (Observations), where she often used the chalkboard for talking her students through problems on which they had requested assistance. Although Brandy regularly prepared problems sets for them to solve and encouraged them to work in small groups, most students preferred to work by themselves, asking questions of her individually (Observations).

Gary. At the beginning of the semester, Gary reflected on how the students in his recitation sections were silent during the class and he "needed them to become engaged in the classroom in order to maximize their learning of the material" (Reflection 1). Fortunately, one student finally spoke up during a recitation, in frustration about learning what seemed to be an overwhelming amount of disconnected facts. Gary was able to show the students a pattern formed by the facts and give the students a level of comfort with both learning strategies and asking questions out loud. As the semester progressed, Gary was able to use the ideas about questions presented in the pedagogy seminar with his whole class and related the concepts of questioning to metacognitive strategies and formative assessment. Gary mentioned that formative assessments helped him "assess where the class is with learning and understanding the concepts," (Reflection 5), but did not give a specific examples of how he uses the information from formative assessments in his own teaching. Gary's tone of voice, when speaking to the whole class, was less conversational and more formal than the other UTAs. However, with individual students, he was relaxed and friendly (Observations). Gary's senior instructor thought Gary's peer learning assistance skills had improved over the semester, especially given that Gary was just a sophomore (Senior instructor interview).

James. According to his senior instructor, because of James' experience as a recitation section leader, his skills for assisting student learning were more advanced than some others. From the beginning of the semester, James prepared open-ended questions within his learning activities that "allowed for students to get their questions answered while not having to directly ask them" (Reflection 1). He took responsibility for "not letting students' misunderstandings persist while at the same time not excluding other students in the discussion" (Reflection 1). He was adamant about incorporating students' input in the class and being prepared to follow students' lead about their interests in the topic of discussion. His recitation sections were observed to include lively discussions between the students and him and within small groups of students. James used questions to connect familiar concepts with newly learned concepts for his students and to evaluate students' learning. He used student answers to questions to adapt his teaching to the students' learning needs. When his students had trouble with a set of true and false questions, he reasoned aloud through each question, "treating the questions not like we were taking a test question with only one right answer but studying for it" (Reflection 4). James thought of incorrect answers from students "not as failures, but as do-able fixes" (Reflection 5). He was motivated to continue to improve his peer assistance skills to "engage the whole class and not just those who want to learn" (Reflection 6). James was aware of his unique position between instructor and student due to recent experience in CHEM 201, "The biggest part of being UTA is helping people how only you know they need help…we are just as much course guides as academic resources" (Reflection 6). James' rapport with the students was genuinely warm yet professional. In a recitation section, several students complained loudly about how something had been taught in their lecture. James was able to calm the students' emotions, defend the senior instructors' methods, and help the students understand the concept in question in a very calm, professional way that seemed to make the students feel at ease (Observations). James is considered to be one of the strongest UTAs in the chemistry department, and the senior instructors rely on him to lead the other UTAs (Senior instructor interview).

Jason. Jason was concerned about maintaining a welcoming environment in his recitation sections. "The incoming students were intimidated or overwhelmed with emotions in their first college class and I was the face of helping them through that" (Reflection 6). I observed Jason twice during the semester, and both times he enthusiastically greeted his students as they entered and kept up a positive attitude throughout the class time (Observations). Jason had plenty of low-cognitive level review questions ready to use with the whole class; although most students did not take notes, all were attentive and many different students participated in answering the questions. Jason did not give the students much wait time to answer the questions, but the questions were mostly recall and he had no trouble getting most students to participate in answering the questions. Occasionally Jason asked the students to explain their reasoning for an answer, but the students did not ask for reasons or clarification. (Observations) Jason's senior instructor observed that Jason offered extra review time to help students and that his enthusiasm was infectious. He often worked with Stacy, a veteran UTA, to plan learning activities and Jason found those activities effective in helping students to establish learning cohorts that functioned well for the students to review with each other outside of class time (Reflection 6). As the semester progressed, Jason continually sought to improve his peer learning assistance skills and spent much time preparing for his recitation sections (Senior instructor interview). Although I did not observe Jason challenging his students with higher-level questioning as other UTAs did with their students, he did create a positive learning atmosphere that attracted students to attend recitation sections (Observations).

Lisa. Even as a veteran recitation section UTA, Lisa learned valuable lessons from interactions with her students. "I considered that these questions [using conversion factors] would arise, but had not laid out a plan of action….being able to anticipate your students' need is only helpful if you are also able to address that need" (Reflection 1). Lisa took responsibility for planning after that incident and preparation was always evident during subsequent classroom observations and meetings with her senior instructors. Lisa and her senior instructors planned engaging learning activities that included divergent questioning, student interaction, and formative assessment. "My hope for an activity was to show that there are so many 'correct' possibilities when working with ions—it isn't always right or wrong" (Reflection 2). Her goal of asking divergent questions to help her students learn was one she struggled with: "I have a hard time taking the very straightforward topics in CHEM 201 and converting them to questions that may have several correct answers" (Reflection 3). To help her students become more aware of metacognitive strategies, she asked the students to explain their answers to their peers because "I have found that my understanding has significantly improved through teaching others and would like my students to experience that" (Reflection 4). Encouraging students to work in small collaborative groups was not effective for Lisa. "I thought that allowing students to do 'group work' would allow them the most time to grow and learn in their own way. But I actually think that when students were in groups, they were more counterproductive. I also noticed that students were much more hesitant to ask questions when in groups" (Reflection 6). Toward the end of the semester, Lisa used more whole-class discussion, in which she called on all students to participate using her knowledge of her students and created a "safe" atmosphere where it was comfortable to participate in discussion (Observations, Reflection 6). The senior instructors who mentored Lisa believed that she was the best UTA in the program due to her dedication to her students' learning and continual improvements in her peer learning assistance skills. Lisa said to me in seminar, early in the semester, "Without a TA in recitation sections, many CHEM 201 students would be drowning. And it's not ok to just let people drown."

Stacy. From the beginning of the semester, Stacy tried to foster effective communication between the students in her recitation section by encouraging multiple methods of solving problems and making those methods visible to all students. She used questions to answer student questions, "to help students understand a concept better as they worked it out for themselves" (Reflection 2). She also asked her students to "talk through a problem" (Reflection 3) to find out where in the process the student was having the most difficulty. This required that her students be willing to communicate with her and the class, a willingness that was supported as they gained trust in her and the learning environment. Although she believed that getting her students to volunteer was difficult (Reflection 5), I observed her doing it with ease and student comfort (Observations). She used different types of formative assessments often in her recitation section to give feedback on her students' learning and to know where to change her instruction to meet student needs. Stacy's senior instructor met with Stacy weekly to reflect on her UTA practice (Senior instructor interview). The goal of her work was to help her students learn and she was serious about doing the best job she could to help them. "I strived to be a better TA than the TAs I had in the past" (Reflection 6). One of the senior instructors commented that by working with Stacy, she was inspired to rethink some of her own practices (Senior instructor interview).

Comparison of descriptions. The analysis reported in this section is built upon a comparison and contrast of peer learning assistance skill development descriptions of the individual UTA cases.

UTAs put into practice the pedagogical strategies they learned about in seminar with varying self-reported skill. Some UTAs, such as Jason, James, and Lisa found whole-class questioning to be effective for student learning and they felt their students enjoyed and appreciated group learning. Brandy found that her students did not respond to whole-class instruction and preferred to work individually or in small groups, interacting with her in a more personal way. No matter which strategies the UTAs used for student learning, according to their reflections, their main commitment was to increase student learning and to engage their students in the concepts and processes required for success in CHEM 201. All the UTAs communicated a desire to help their less-experienced peers because they themselves had been in introductory classes not long before this semester and remembered their struggles to succeed in the course. They perceived that their unique position as a more experienced peer would help the students in a way that the senior instructors could not.

These findings were confirmed by classroom observations during the semester. Both times I paid an unannounced visited to each of the six UTAs, they were interacting with the students and encouraging the students to interact with each other. There were planned activities involving input from every student, sometimes in small group or whole class discussion or activity, sometimes in written format. Often the students were required to physically move to join a group, make a choice, or create a product. An example UTA recitation section started with the UTA entering the classroom shortly before the start of class with graded quizzes to pass back that had been taken the week before. The UTA knew most of the students, especially by the end of the semester, and so passing back the quizzes was a quick job and also served as a way to take attendance and greet each student by name. The UTA then introduced the objective for the day, such as practice working with ionic compounds, and launched into an explanation of the activity that had been planned. The UTA then proceeded to facilitate the activity by passing out materials, such as cards with ion names printed on them, and encouraging the students to begin sorting themselves into positive and negative ions. Students who were confused about where they belonged as an ion were allowed to stay at their seats and observe the other students' choices. In a few seconds the UTA posed questions to the students who had decided to join either the positive or negative group. As the students justified their choices, incorrect thinking was subtly corrected through guiding questions, and those students who had hesitated to join a group had more clues about where they belonged and were able to make a more informed choice. After the ions were correctly segregated into positive and negative groups, the UTA directed the students to move to join other ions to form electronically balanced ionic compounds. Some students did not have a "match" to make a compound, so the UTA pointed out to the whole group other ionic compound possibilities for these students. The learning environment was one of exploration and justification of choices, not right or wrong answers with punitive consequences for students with wrong answers. By the time the activity was over and all students had found possible matches with other ions, students were talking with each other about ionic compounds and asking the UTA questions to extend their learning. There was a comfortable atmosphere to the classrooms run by UTAs—students sat closer together and nearer to the front of the room or wherever the UTA stood to illustrate examples for students questions. Students actively sought UTA interaction and received it, and students were more vocal in questions, complaints and compliments. Sometimes, UTAs and students had conversations about future course choices and inquiries into the UTAs career plans. This amount of interactivity and active student engagement was typical across all of the UTA recitations sections observed.

Except for TA-led whole class discussion or working in small groups on assigned problem sets, I did not see such interactivity in a GTA classroom. Students did not actively pursue questions in GTA classrooms in the way I saw them do in UTA classrooms. During many GTA-led recitation sections, I did not see the GTA leave the front desk or table to assist students individually or in small groups and few students asked questions. Often the GTA showed examples of problems on the chalkboard with little student input, or assigned students to work on problem sets and waited at a front desk for students to ask for assistance. I saw one GTA work individually with one student for 40 minutes, while the rest of the class worked silently and independently on a problem set. At the end of the semester, I did see one GTA direct students to the chalkboard to demonstrate problems for the whole class and students complied that day, although the GTA had warned students she was going to “make them participate.” From this comment by the GTA, I assumed that students had not been willing to participate in group activities previously.

Summary of results for research question 2. UTAs scored significantly higher than GTAs on TA Impact and TA Rapport scores reported by CHEM 201 students. Students did perceive both UTAs and GTAs as having strong content knowledge; when asked to respond on the end of course survey to the statement “My TA had a strong knowledge of the course content,” 84 percent of students in GTA-led recitation sections replied agree or strongly agree and 85 percent of students in UTA-led sections responded agree or strongly agree. This means that students in UTA recitation sections perceived their TAs to be stronger in other peer learning assistance skills: to give clearer explanations, lead more effective discussions, and give more choices for student learning than students in GTA recitation sections. Students in UTA recitation sections also perceived their TA to be more open in communication, to more caring and understanding, and to encourage more questions than students in GTA recitation sections. Being a student in a UTA recitation section had a moderate effect size on TA Impact score and a small to moderate effect on TA Rapport score. Besides TA Type, gender was a significant predictor of TA Impact, with males rating their TAs higher on TA Impact, controlling for TA Type. The more advanced placement STEM courses a student took in high school, the stronger the student rated their TA. TA Impact was also positively influenced by TA Rapport. TA Rapport, in turn, was predicted by TA Type, with UTA-led students rating their TAs higher on TA Rapport. TA Rapport was also predicted by STEM Recognition, which relates how a student identities him or herself (“I think I am a science person” with TA qualities such as “My TA cares about me as a person”.

All UTAs were practicing with their students many of the skills they were learning in the pedagogy seminar. Veteran UTAs continued to improve throughout the semester and new UTAs had the opportunity to practice and reflect on the skills they had discussed in the pedagogy seminar. The students in UTA recitation sections recognized the UTAs’ aims to continually improve their peer learning assistance skills by rating these skills higher than students in GTA recitation sections had rated their TAs.   All UTAs seemed genuinely concerned about their students' learning, and wanted their students to be comfortable and confident in the class. Almost all UTAs believed that any of their students could learn some science in CHEM 201. Almost all UTAs commented about how surprised they were at the extra time it took to prepare for their recitation sections.

Research Question 3 - Impact of the UTA Program on Academic Achievement

The purpose of Research Question 3 was to compare academic achievement by students who had UTA recitation sections with those who had GTA recitation sections, after controlling for academic covariates such as college math readiness, college GPA, number of STEM courses taken in high school, and parents' education level. This comparison was initiated using a 2-level hierarchical model with final exam score as the outcome variable.

Another outcome variable for student academic achievement, persistence, was operationalized in this study by a student enrolling in the second semester of general chemistry (CHEM 202) for the next semester. Of the 594 CHEM 201students participating in this study, 342 (58%) had declared or intended to declare a major in a STEM program which required CHEM 202. 128 students had declared majors or intended to declare majors in programs not requiring the second semester of CHEM 202, and 124 students had not recorded any intended major at all. To further examine student achievement, chi square tests were conducted to check for significant differences between students in GTA-led recitation sections and UTA-led recitation sections for persistence (enrollment in next semester of general chemistry). To reveal significant predictors for persistence, a logistic regression was performed.

Hierarchical linear modeling of student achievement data. The analytical process began by estimating the degree of relationship among students in the same section, which is captured by calculation of the intraclass correlation coefficient (ICC). To determine the ICC, an unconditional model was estimated in which each student's score on the final exam consisted of three elements: the overall mean (γ_{00}), the deviation of the section mean from the overall mean (μ_{0j}), and the deviation of the student's score from his or her section mean (r_{ij}).

$$\textit{Final Exam Score}_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

Every student in the same section has the same value for μ_0, allowing the dependence of scores from the same section to be modeled. The variance of the μ_0 is the between-section variance, called τ_{00}. The variance of r_{ij} is the within section variance, σ^2. The ICC is the ratio of the between section variance τ_{00} and the total variance ($\tau_{00} + \sigma^2$). With Final Exam Score as the outcome, within groups variance σ^2 was = 397.58 and between groups variance, τ_{00}, was 19.27, resulting in an ICC of 0.047, interpreted as 4.7% of the variance in final exam scores explained by clustering in sections. This is a very small amount of variance to be explained at the section level. The remaining variance in scores could be explained by within section (Level 1) variables.

Level 1 Model. The Level 1 (student variables) model was built using plausible student level variables (listed in Table 4-9) associated with final exam scores: Statistically significant predictors of Final Exam Score were kept in the model: College GPA (the most recent indicator of student's success with college course load);

Persistence (enrollment in the next sequential chemistry course); and Math Z-Score (indicator of college preparedness especially in math-intensive courses such as general chemistry). College GPA was centered on the grand mean GPA (2.83). This Level 1 model explained 54% of the within groups variance (σ^2 was reduced from 397.58 in the baseline unconditional model to 182.54 in this Level 1 model). Adding the student level variables did nothing to explain the between sections variance; it actually increased to 48.71 from the baseline 19.27. The variability in the relationships between the final exam score, college GPA, and math z-score were statistically significant, confirming there is more between section variance left to explain. From the tau correlation matrix, college GPA, persistence, and math z-score variables had a negative relationship with the final exam score. This meant that the slope for these predictors is less steep as final exam score increases. In other words, they may not have as much effect on the final exam score at higher scores.

Having tried all student predictors that were available and were theoretically plausible predictors of final exam score, the model that explained the most variance possible was represented by the following set of equations:

$$\begin{gathered}
\textit{Final Exam Score}_{ij} = \beta_{0j} + \beta_{1j}*(\textit{College_GPA}_{ij}) + \beta_{2j}*(\textit{Persistence}_{ij}) \\
+ \beta_{3j}*(\textit{Math_Z-Score}) + r_{ij} \\
\beta_{0j} = \gamma_{00} + u_{0j} \\
\beta_{1j} = \gamma_{10} + u_{1j} \\
\beta_{2j} = \gamma_{20} + u_{2j} \\
\beta_{3j} = \gamma_{30} + u_{3j}
\end{gathered}$$

The parameter estimates for all variables in this model are found in Table 4-10. In this model, γ_{00} (49.28) represents the predicted final exam score for a student with a college GPA of 2.83 (the grand mean), who does not intend to enroll in CHEM 202 for the next semester, and who has a math z-score of 0 (holding ACT-Math score constant at

21.1 or SAT score constant at 514). The college GPA slope, γ_{10}, (13.57), represents the expected change in final exam score per unit change in GPA above or below the grand mean of 2.83, after controlling for persistence and math z-score. The persistence slope, γ_{20} (3.48), represents the differential between students who did not intend to enroll in the next chemistry course in the sequence (CHEM 202) in the next semester and those who do intend to enroll, after controlling for all other variables in the model. The math z-score slope, γ_{30} (3.66), represents the expected change in final exam score per unit change in student math z-score.

Table 4-9
HLM Analysis Variables

Variable	Description
	Dependent Variables
Final Exam Score	Percent correct responses on common 35 question multiple choice exam
	Student Level Variables
Math Z-Score	Student ACT or SAT math subtest score standardized with test population mean and standard deviation
College GPA	Grade point average for all college credit as of December 2012
Recitation Attendance	Dummy coded as "high" > 80%, "medium" <80% but >50% and "low" <50%
Persistence	Student registered for next chemistry course in sequence (CHEM 201) counted as 1, not registered 0
Parents' Education Level	Any parent college experience = 1; No college experience listed for either parent = 0
Race	Non-white = 0; White = 1
Gender	Male = 0; Female = 1
Number of STEM Advance Placement Courses	Number of advanced placement or international baccalaureate courses with STEM focus taken in high school
Reading Z-Score	Student ACT or SAT reading/verbal subtest score standardized with national test population mean and standard deviation
	Section Level Variables
TA Type	GTA = 0; UTA = 1
Mean Math Z-Score	Section mean for Math Z-Score
Mean College GPA	Section mean for College GPA
Mean TA Impact Score	Section mean for TA Impact Score
Mean TA Rapport Score	Section mean for TA Rapport Score

Full Level 2 Model. A full Level-2 model was estimated in which Level 2 variables were explored as predictors of intercepts and slopes. Section-mean college GPA was a predictor of the intercept. TA Type (GTA=0; UTA=1) was a predictor of the college GPA slope. The section-mean math z-score was a predictor of the persistence slope. The set of equations for this model:

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

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(\text{Mean}\_College\_GPA) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}*(TA\_Type_j) + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21}*(Mean\_Math\_Z\text{-}Score_j) + u_{2j}$$
$$\beta_{3j} = \gamma_{30} + u_{3j}$$

Section-mean college GPA was centered on its grand mean (2.70). Therefore, in this 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 score of 21.1), and who is in a GTA-led recitation section that has a mean GPA of 2.70 and a section mean math z-score of 0.

The effect of the Level 2 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 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. Thinking about 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.

Table 4-10
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*

*$p < .05$
**$p < .1$

Using the above full Level 2 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 score of 26, 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 61% on the final exam. A student with a UTA and the rest of the variables same as above would score a 65% on the final exam.

UTA impact on student persistence. A chi-square test of independence was performed to examine the relationship between TA Type (GTA or UTA) and enrollment in the next semester of general chemistry for all 594 students participating in this study. The relationship between these variables was significant, χ^2 (1, N = 594) = 13.64, p <.001. Students having UTAs as recitation section leaders were more likely to enroll in the next semester of general chemistry.

Of the 343 students declared or intending to declare majors requiring CHEM 202, 189 students were in UTA-led recitation sections and154 students were in GTA-led recitation sections. 135 out of 189 UTA-led students (71%) enrolled in CHEM 202 while 82 out of 153 GTA-led students (53%) enrolled in CHEM 202. A chi square test confirmed that proportionally more students required to take CHEM 202 who have UTAs as recitation section leaders enrolled in CHEM 202 than did those who have GTA-led recitation sections: χ^2 (1, N = 343) = 12.07, *p* =.001).

Additionally, of the students who had no declared or intended major (n = 124), 50% (24 out of 48) of UTA-led students enrolled in CHEM 202 while 41 % of GTA-led students (31 out of 76) enrolled in CHEM 202. According to a chi-square test, the relation between TA Type and intention to enroll in CHEM 202 for undeclared students was not significant, χ^2 (1, N = 124) = 0.106, *p* =.745). A third group of students, those who declared majors that did not require CHEM 202 (n = 128) demonstrated similar results. Of the 47 students in UTA-led recitation sections, 15% enrolled in CHEM 202.

Of the 81 students in GTA-led sections, 17% enrolled in CHEM 202. The intended majors for this third group of students were a diverse mix, ranging from STEM fields, such as computer engineering to fields not considered in STEM such as English or political science.

Predictors of persistence. Because persistence is a categorical dependent variable, logistic regression was used to explore the predictors of persistence. Variables tested to predict persistence were TA type, CHEM 201 final exam score, college GPA, math z-score (ACT/SAT), parent education level, and section mean scores for TA Impact, TA Rapport, STEM recognition, and STEM Interest. A five-predictor logistic model (obtained from backwards entry) was fitted to the data to test the research hypotheses regarding the relationships between the likelihood that a student would enroll in CHEM 202 and TA Type, final exam score, college GPA, math z-score, and parent education. Examining the odds ratios, having a UTA gives a student three times the chance of enrolling in CHEM 202 than having a GTA. The higher the final exam score, college GPA and math score, the more likely the student will enroll in CHEM 202. However, a student with a parent having college experience would be less likely to enroll in CHEM 202.

Table 4-11
Logistic Regression for Persistence

Predictor	B	S.E.	Wald	df	Sig.	Exp(B)
TA Type Code (GTA = 0; UTA =1)	1.160	.319	13.191	1	.000	3.188
Final Exam(%)	.025	.009	7.008	1	.008	1.025
College GPA	1.064	.251	17.943	1	.000	2.899
Math z-score	1.313	.252	27.188	1	.000	3.718
Parent Ed (no college=0; college = 1)	-.920	.387	5.652	1	.017	.398
Constant	-4.989	.813	37.695	1	.000	.007

Summary of results for research question 3. There was little variance in final exam scores explained by clustering students in sections. Addition of the three student variables explained 54% of within section variance. Addition of the three section variables explained 15% of the between section variance in the intercept. In this model, a student's final exam score was significantly related to the ACT or SAT math subtest score and college GPA. Holding all other variables constant, the intention of the student to take the next sequential chemistry course, CHEM 202, was related to a 3.48 point increase in final exam score. This was especially important for students who were in sections that averaged lower on the math z-score. Being in a UTA led section did more strongly influence the college GPA effect on the final exam score.

Another outcome of interest, persistence in STEM that leads to taking the second semester of general chemistry, was explored. Students who were required to take CHEM 202 for their declared or intended major were more likely to enroll in CHEM 202 if they were in a UTA-led recitation section. Significant predictors of persistence, besides TA type were final exam score, college GPA, math z-score and parent education level.

Research Question 4- Influence of the UTA Program on Student STEM Identity

The purpose of Research Question 4 was to explore the relationships between aspects of the UTA program and the reported STEM identity of the CHEM 201 students, operationalized as recognition of themselves as “science persons” or “math persons” and interest in science and mathematics activities.

Undergraduate course experience survey. The results from the survey were reported as scores on the Student STEM Recognition Scale and scores on the Student STEM Interest Scale. These scores were planned to be used as outcome variables in an HLM analysis of student STEM identity.

Student STEM Recognition Scale. The responses to items loading onto the third component (termed Student STEM Recognition), listed in Table 4-4, were weighted by loading factors and summed for each student to create a Student STEM Recognition scale. The combined mean Student STEM Recognition score was 10.19 out of a possible 12.50 (3 weighted items each with a maximum rating of 5 for strongly agree). The mean UTA Student STEM Recognition score was 10.44 and the mean GTA Student STEM Recognition score was 9.94. Mean Student STEM Recognition scores for each TA group means are shown in Table 4-12.

Student STEM Interest Scale. The responses to items loading onto the fourth component (termed Student STEM Interest), listed in Table 4-4, were weighted by loading factors and summed for each student to create a Student STEM Interest scale. The combined mean Student STEM Interest score was 9.98 out of a possible 12.23 (3 weighted items each with a maximum rating of 5 for strongly agree). The mean UTA Student STEM Interest score was 10.12 and the mean GTA Student STEM Interest score was 9.83. Student STEM Interest scores for each section along with TA group means are shown in Table 4-12

An independent -samples t-test was conducted on the Student STEM Recognition and Student STEM Interest scores to evaluate whether the GTA mean was significantly different from the UTA mean. Levene's test for equality of variances was significantly nonequal at $p < .05$ for both scores; therefore the corrected degrees of freedom and t statistics were reported assuming the variances were not equal.

Table 4-12
Comparison of Mean Student STEM Recognition and Student STEM Interest Scores

	GTA		UTA					
Variable	*M*	*SD*	*M*	*SD*	*df*	*t*	*p*	Cohen's d
Student STEM Recognition	9.94	2.14	10.44	1.63	374	-2.643	.04	0.54
Student STEM Interest	9.83	2.14	10.12	1.82	391	-1.485	.353	-

The independent samples t-test indicated that the mean Student STEM Recognition score for students in UTA sections was significantly higher than the score for students in GTA sections. No significant difference was found between the reported STEM interest of students in UTA groups and students in GTA groups.

Inter-correlations among factor scores. Scores on both the Student STEM Recognition Factor and Student STEM Interest Factor correlated significantly with each other ($r = .503$, $p < .001$). The more students recognized themselves or were recognized by others as "science persons" or "math persons" , the stronger their interest in science and mathematics activities such as experiments, science discussions, or learning more about science or mathematics.

Hierarchical linear modeling of STEM identity variables. Given the statistical difference in STEM recognition between students in UTA-led sections compared to GTA-led sections, an HLM model was used to investigate in more detail possible relationships of other variables with STEM recognition. However, because the STEM interest scores were not different across UTA and GTA groups, no parallel HLM model was computed for that outcome variable. The analytical process began by estimating the degree of relationship among students in the same section, which is captured by calculation of the intraclass correlation coefficient (ICC). To determine the ICC, an unconditional model was estimated in which each student's score on the Student STEM Recognition scale consisted of three elements: the overall mean (γ_{00}), the deviation of the section mean from the overall mean (μ_{0j}), and the deviation of the student's score from his or her section mean (r_{ij}).

$$\textit{Student STEM Recognition Scale Score}_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

Every student in the same section has the same value for μ_0, allowing the dependence of scores from the same section to be modeled. The variance of the μ_0 is the between-section variance, called τ_{00}. The variance of r_{ij} is the within section variance, σ^2. The ICC is the ratio of the between section variance τ_{00} and the total variance ($\tau_{00} + \sigma^2$). With Student STEM Recognition score as the outcome, within groups variance σ^2 was = 3.538 and between groups variance, τ_{00}, was 0.1257, resulting in an ICC of 0.0343, interpreted as 3.43% of the variance in final exam scores explained by clustering in sections. This is a very small amount of variance to be explained at the section level.

Level 1 model. The Level 1 (student variables only) model was built using student level variables captured by the Undergraduate Course Survey that were associated with student STEM recognition. Statistically significant predictors of Student STEM Recognition were kept in the model: student reported Math ACT Score (indicator of college preparedness especially in math-intensive courses such as general chemistry) and TA Rapport Building Scale score (evaluating student's perception of the TA's encouragement and understanding of student's needs). Math ACT scores and TA Rapport were centered on their respective means: mean Math ACT score was 28.63 and mean TA Rapport score was 10.29. This Level 1 model explained 15% of the within groups variance (σ^2 was reduced from 3.54 in the baseline unconditional model to 2.99 in this Level 1 model).

Table 4-13
Fixed Effects Estimates (Top) and Variance-Covariance Estimates (Bottom) for Models of the Predictors of Student STEM Recognition

Parameter	**Unconditional Model**	**Level-1 Model**
Fixed Effects		
Intercept (γ_{00})	10.18* (1.14)	10.36* (0.10)
Math ACT Score (γ_{10})	—	0.11* (0.03)
TA Rapport Score(γ_{20})	—	0.15* (0.05)
Variance Estimates		
Within-section variance (σ^2)	3.54	2.99
Intercept variance (τ_{00})	0.126**	0.025
Math ACT-Score slope variance (τ_{11})	—	0.002
TA Rapport Score slope variance (τ_{22})	—	0.0003

*p < .05
**p < .1

The small amount of variance explained by clustering in sections suggests that any impact of being in a UTA section would be minimal. The relatively small coefficients in Table 4-13 for the significant predictors confirms that any statistically significant predictor would nevertheless still offer only a small effect size measure. Additionally, standard deviations and means in Table 4-12 suggest that there is a ceiling effect on this particular measure since the overall mean of each group is within approximately one standard deviation of the top of the scale. Given that all students responding to this survey were in a STEM majors course, there tended to be a preponderance of responses at the high end of the recognition scale.

Summary of results for research question 4. Students in UTA-led recitation sections rated themselves significantly higher on STEM Recognition than did students in GTA-led recitation sections. Additionally, exploration of predictors of strong STEM recognition included the TA rapport variable, suggesting that those who felt a stronger rapport with their TA tended to more positively rate their STEM recognition. Given that the students in UTA sections reported stronger TA rapport as reported in results for Research Question 2, this suggests a possible positive impact of UTAs on STEM recognition. By contrast, there were no differences between UTA and GTA students on STEM interest.

Summary of Results

Most UTAs believed that they increased the depth of their content knowledge as a result of their UTA work and the post content knowledge test results showed that both UTAs and GTAs had an equivalent knowledge base of general chemistry concepts. All UTAs affirmed that the pedagogical strategies presented to them in seminar and workshop had not only helped them assist their students' learning but had also impacted their own learning in more advanced coursework.

Students in UTA recitation sections perceived their TAs to be stronger in peer learning assistance skills: to give clearer explanations, lead more effective discussions, and give more choices for student learning than students in GTA recitation sections. Students in UTA recitation sections also perceived their TA to have better rapport-building skills: to be more open in communication, to be more caring and understanding, and to encourage more questions than students in GTA recitation sections. The more a student recognized themselves and were recognized by others as a "science person" or a "math person", the higher they rated their TA as caring and encouraging. UTA reflections, researcher observations, and senior instructor comments on agreed that UTAs were practicing with their student the skills they were learning in the pedagogy workshop and seminars.

The better students were doing in college and the more prepared they were for the mathematical aspect of college coursework, the better they did on the final exam, regardless of TA type. Also students who went on to enroll in the second semester of general chemistry did better on the final exam. Students who were in sections led by a UTA and had higher than average college GPAs scored better on the final exam. In other words, the better students were doing in college, the more the UTA was able to help them score well on the final exam, even after controlling for a suite of variables that had been included in the model. College GPA was not predicted by ethnicity or gender, so UTAs were helping all students.

Results from the logistic regression of the persistence outcome variable showed that students are three times more likely to persist in CHEM 202 if they had a UTA in CHEM 201. Other positive predictors of retention included having strong college grades, and being well-prepared (e.g. strong ACT math scores) to take on STEM coursework.

Having a UTA and having a TA with good rapport-building skills were positively related to students recognizing themselves and being recognized by others as a ‘science person’. There seemed to be no difference between students’ STEM interest in UTA-led recitations and GTA-led recitations.

CHAPTER 5

DISCUSSION

The purpose of this study was to examine a UTA program for retention improvement by investigating whether training, support and practice helped UTAs develop the content knowledge and skills needed to effectively assist introductory chemistry students in learning and to measure any impact that the UTAs may have had on student achievement and the identity of students as STEM students. Four research questions addressed this purpose. This chapter discusses the findings and conclusions for each research question, and then summarizes the implications of the findings across all four research questions.

Research Questions 1a and 1b - UTA Content Knowledge and Learning Approaches Growth

The purpose of Research Question 1a and 1b was to examine the ways in which UTAs deepened their content knowledge during their UTA experience and used newly acquired learning strategies in their own scholarship.

Deepening Content Knowledge

Most GTAs and UTAs increased content knowledge of general chemistry concepts after teaching a semester of CHEM 201, which supports the findings of studies reporting cognitive benefits for those who prepare to teach others (Bargh & Schul, 1980; Schalk, McGinnis, Harring, Hendrickson, & Smith, 2009; Weidert, Wendorf , Gurung, &

Filz (2012). That there was no substantial difference between UTAs and GTAs in learning gains on the pre/post content knowledge test suggests that in-depth review of the material in preparation for teaching, which was performed by all TAs, supported their own learning gains. The seminar in which the UTAs participated, but the GTAs did not, did not focus on content preparation, so it was not expected to make a difference in content knowledge growth for the UTAs. Perhaps by working closely with their senior instructors and each other, the UTAs reconceived naïve notions about some chemical concepts, while GTAs had the option of reviewing content with senior instructors and had more academic and research experience to create a fairly strong content foundation. GTAs' pre-test scores were lower than many of the UTAs' pre-test scores and GTA post-test scores were not any higher than some of the UTA post-test scores. The pre and post content test scores alleviated any possible concerns, if they existed for any of the faculty, that the UTAs were not as prepared, content-wise, as the more experienced GTAs.

Of the five most missed questions on the content test, four were conceptual in nature and from the material learned at the end of the semester. Hence, there was little time left in the semester to continue practicing this challenging material or applying it in different contexts with the students. Moreover, the most problematic concepts were related to electron assignment, chemical bonding, Lewis structures and molecular geometry. How to teach these complex concepts has been specifically discussed in the chemical education literature (Cooper, Grove, Underwood & Klymkowsky, 2010; Cooper, Underwood, Hiley, & Klymkowsky, 2012; Packer & Woodgate, 1991; Suidan, Badenhoop, Glendening, & Weiunhold, 1995). Previously, suggestions for teaching how to write and use Lewis structures and identify molecular geometries merely involved sets of rules or heuristics for students to follow in order to get the right answer (e.g., Packer & Woodgate, 1991). Only recently have chemical education researchers examined how molecular structure and properties have been traditionally taught, and they have discovered those teaching strategies are in conflict with how most people actually learn (Cooper et al, 2010). It is likely that most UTAs, GTAs, and now their students, are still conceptualizing molecular structure using heuristics that have little meaning for the TAs or students, rather than on a sure comprehension of how molecular structure predicts chemical behavior. By not fully understanding these difficult abstract concepts, novice and more experienced students alike will continue to have trouble using molecular structure to predict chemical behavior.

Six out of nine TAs missed questions on the post- test that they had also missed on the pre-test. Some of these questions were the problematic molecular structure questions just described, while others were concerned with other topics such as electron configurations and thermochemistry. This suggests that UTAs and GTAs, like their students, possess misconceptions that are resistant to change (Strike & Posner, 1992), even after reviewing these concepts in depth in preparation for teaching. Because the missed questions tended to cluster in a relatively few domains across all TAs and they were still frequently missed even after teaching those topics to others for a semester, this is a concern because it suggests that these mistakes were authentic rather than a careless error of some sort. If the TAs hold misconceptions about a chemistry topic, there is a good chance that they will pass those misconceptions on to their students.

Five out of six UTAs reported that they had deepened their content knowledge related to material in this freshmen-level chemistry course. The UTAs did not report learning new material or material with which they were previously unfamiliar, but that by having to review the material in order to clearly explain or develop questions to ask students or to plan active learning activities, they felt more confident about their foundational chemistry knowledge. They often described the effect as “solidifying their knowledge.” For four out of the six UTAs, this was shown by learning gains on the content knowledge test. For the two UTAs who had the same scores on both pre and post-tests, their test scores were high enough (78 and 87 percent correct) that the senior instructors were not uncomfortable about their content knowledge foundation.

Self-Learning Approaches

By focusing on how others learn, during seminar discussions, practice of skills in the classroom, and regular reflection on their teaching practice, the UTAs were able to transfer that focus to their own learning without difficulty. In fact, many of the UTAs had already been using the strategies discussed in seminars but didn’t realize the strategies had a name or that others found them useful too. Although it was not the sole intent of the UTA program, self-improvement of the UTA’s learning skills may increase the quality of the chemistry practitioner graduating from the University, as a scholar and possible future teacher in graduate school, K-12 education, or in workplace training. The PRIMES UTA program employs on average 15 different students each year who are chemistry majors and many who are seniors. The chemistry department graduates on average 24 students each year. Therefore, a good portion of the chemistry majors in the department will have UTA experience before they graduate.

As a member of a community of practice with their students and senior instructors, the UTAs were motivational to others with their application of newly acquired learning skills. In talking with their students about what learning strategies worked for them, UTAs were credible models of a STEM student who had succeeded by putting forth best effort: using metacognition, actively practicing problem-solving, and using formative assessments to focus on what they needed for improvement. For the senior instructors, the UTAs were examples for the depth of learning that could happen when students applied research based learning strategies to their own studies.

Research Question 2 - UTA Peer Learning Assistance Skills

The purpose of Research Question 2 was to describe the development of UTAs' skills for assisting their less-experienced peers in learning general chemistry content. From the student perspective, comparison was made between the UTAs' and GTAs' peer learning assistance skills. The UTAs also reflected on their practice to report their perceptions of their own skill growth.

The Student Perspective

Using the Course Experience Survey, peer learning assistance skills for both UTAs and GTAs were evaluated by the students, resulting in TA Impact on Academic Achievement and TA Rapport-Building Skills scores. The higher TA Impact and TA Rapport scores earned by the UTAs may have been related to one of the main differences between the TA groups—the pedagogical training program required for the UTAs. The seminar and workshop experiences in the program were directed toward supporting the UTAs in helping their students learn in a positive environment and using research based learning strategies. Additionally, the amount of time given during the seminars for UTAs to share their perspectives on their practice with each other and chemistry faculty was effective for building a UTA culture of learning assistance. Not only did the UTAs participate in a community of practice (Lave & Wenger, 1991; Thiry, & Laursen, 2011) with their students in the classroom, the UTAs were also vital members of a community of practice for teaching and learning with chemistry faculty and education faculty.

In analyzing the relationships between student variables and TA Impact score, Math ACT score and number of STEM AP courses taken in high school did not significantly predict TA Impact score. This suggests that no matter how well prepared they were for college chemistry study, UTA-led students perceived their TA as having a larger impact on their academic success than students who had a GTA.

Another difference between the two groups was that the UTAs were current majors in the same program in which they were teaching and the GTAs had attended another university (perhaps even in another country) for their undergraduate degree. The UTAs had recent knowledge of the expectations and customs of the CHEM 201 program and had probably been students in one or more of the current instructors' courses. This may have allowed the TAs to seem more effective to the students in that they had recent knowledge of course expectations and requirements.

The UTA Perspective

It was evident from UTA reflections that the UTAs were practicing skills learned in seminars. Additionally, the UTAs were connecting the topics (questioning, metacognition, formative assessment) in practice instead of thinking of them as discrete constructs. This suggests that the UTAs were accepting the treatment offered in the pedagogical training program and delivering the intended treatment to CHEM 201 students.

Because of the training and support for the UTAs from the pedagogical training program, the UTAs were equipped with techniques for uncovering their students' understandings and areas of confusion. And because of the UTAs' familiarity with CHEM 201's requirements and expectations, they were situated to best anticipate likely difficulties their students may experience with the content and proactively assist them in understanding challenging content. Evidence from the UTAs suggested that they recognized the complications in helping others master challenging chemistry ideas, identifying areas where they would want to do better if they were to repeat the UTA experience. In spite of the challenges, they still believed that they had positive impacts for some of their students.

Research Question 3 - Impact of the UTA Program on Academic Achievement

The purpose of Research Question 3 was to measure how achievement of students in the UTA-led group may have been impacted by the UTA program, after accounting for differences in the treatment and comparison groups and variables outside the recitation section experience. Achievement was operationalized by final exam grades for one set of analyses, and a second approach to documenting achievement was persistence in STEM by enrolling in the second semester of general chemistry.

Impact on Final Exam Grades

The HLM model of final exam score outcome revealed little variance associated with the clustering of students in recitation sections. Student level factors that were not captured in this study, such as amount of time students spent studying, how students spent their time studying, previous chemistry-specific preparation, and extracurricular variables such as living conditions, emotional conditions, family issues, financial issues, health issues, and student adjustment to college life, may influence grade performance far more than 50 minutes per week in a recitation section. Thus, any variance that might be attributable to section-level variables (such as type of TA) tended to be overshadowed by the much larger variance associated with individuals.

Student level academic variables that were collected in this study explained some of the within section variance. The overall effects of college GPA and math z-score, although statistically significant, were small. Another relationship was found between final exam score and whether or not the student enrolled in the second semester course for general chemistry, CHEM 202. This suggests that students who enrolled in CHEM 202, either because their major required the course or they possessed an internal motivation to take the course, were more serious about preparing for the final exam, or possessed a higher ability on the final exam. Alternatively, because the student did well on the final exam (and in the course as well), the student was encouraged or motivated to take the second semester of general chemistry. Because student persistence in the general chemistry course sequence indicates an intention to persist in many STEM majors, this is an interesting relationship to study in more depth, so student persistence will be examined further below.

The model also showed that students who were in sections led by a UTA and had higher than average college GPAs scored better on the final exam. In other words, the better students were doing in college, the more the UTA was able to help them score well on the final exam, even after controlling for a suite of variables that had been included in the model. This relationship could be desirable for retention of high quality STEM students because the UTAs are effectively helping those students who are most ready to learn at the college level and are making a successful adjustment to college. College GPA was not predicted by ethnicity or gender, so an alternative explanation of UTAs working more effectively with those of a certain gender or ethnic group could be ruled out.

Persistence in Chemistry

Grades are not the only factor that students consider when choosing to persist in a STEM program. Strenta et al. (1994) and Rask (2010) found that not only low grades during introductory STEM coursework but also student perception of poor quality teaching, and discouragement of students from asking questions in STEM classrooms played a part in well-qualified students' decisions to leave an intended STEM major. In this study, results showed that positively impacting these two aspects of the STEM course experience may have the opposite effect by leading to stronger retention. CHEM 201 students reported stronger UTA rapport compared to GTA rapport, which included aspects of the experience such as being comfortable to ask questions. Students also reported a perception of stronger positive UTA impact on their academic achievement compared to GTAs, from which one might infer a student perception of higher teaching quality from the UTAs. Results from the logistic regression of the persistence outcome variable showed that students are three times more likely to persist in CHEM 202 if they had a UTA in CHEM 201. Other positive predictors of retention included having strong college grades, and being well-prepared (e.g. strong ACT math scores) to take on STEM coursework. Coupled with the HLM analysis result that UTAs were more effective at helping students with higher college GPAs achieve higher grades, the stronger persistence of UTA-led students showed that the UTA program is an effective program for retention of introductory-level students in STEM majors. These specific positive predictors for retention of UTA-led students offers evidence for why the UTA program may have had this desirable outcome.

Research Question 4 - Influence of the UTA Program on Student Identity

The purpose of Research Question 4 was to explore the relationships between aspects of the UTA program and the reported STEM identity of the CHEM 201 students. As in the work of Hazari et al. (2010) and Carlone and Johnson (2007), results from this study show a correlation between students' math competency, student STEM interest, and student STEM recognition.

Math competency was measured before students came to university, so was not affected by this program at the university level. The average math competency for students enrolled in CHEM 201 was high enough so as not to be a concern that students were ill-prepared for college STEM study. By including a pre-existing math competency measure that students brought to the program, the analyses controlled for any potential positive impact of that variable in order to more carefully investigate the impact of STEM recognition and STEM interest above and beyond math competency.

It was not surprising to find little difference between the perceived STEM interest for students in UTA-led sections and students in GTA-led sections. Interest in STEM was probably initiated in childhood or early adolescence (Maltese & Tai, 2010; Tai, Liu. Maltese, & Fan, 2006). STEM interest would be difficult to change in young adulthood without intensive influence from some significant others. The UTA program did not provide that intense influence, nor did the traditional recitation sections with the GTAs.

By contrast, there did seem to be a UTA impact on STEM recognition. Student STEM recognition as operationalized in this study included not only their own perception of themselves as a 'science person,' but also a belief that meaningful others' also saw the student as a 'science person' (Carlone& Johnson, 2007; Hazari et al., 2010). For this study, a meaningful 'other' person seemed to be a role played by the UTAs – successful STEM students in programs only a few years ahead of the student him/herself which likely puts the UTA in the role of a credible proxy for social comparison to assess ability to succeed in a STEM program. The strength of the social comparison students may have made with UTAs was likely closely related to the reported stronger rapport these students had with UTAs over GTAs. This triangulates with the results reported in Research Question 2 documenting this stronger rapport, which collectively strengthens interpretations of how the UTAs may have differentially impacted students.

Conclusions

There are several possible ways to investigate strengthening STEM student retention when conducting a semester-long examination of students in a retention improvement program. One way is to investigate grades, which are important for student progression in a STEM program. An equally or perhaps even more relevant consideration is to examine enrollment in the next course as a measure of persistence, particularly if that enrollment could be increased among those students who achieved at least adequate grades in the first course. While most universities do not require or even encourage students to declare a major during the first year of college, students who intend to major in a STEM discipline enroll in STEM major-required two course sequences like CHEM 201 and CHEM 202 in their first year or two at university. 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. For increased retention of students in STEM programs, it is important for students to persist into that second course.

Based on the results of this study, there seemed to have been a set of mutually reinforcing elements of the UTA-led experiences that combined to positively influence their students' retention on a STEM program track as documented by their stronger persistence into the subsequent STEM course. These program elements included stronger UTA rapport with students, greater student perceived UTA impact on academic achievement, higher UTA-led student STEM recognition, and the positive UTA impact on final exam grades for students with strong college GPA. Any one of these elements alone may or may not have had a measurable impact on persistence, but combined they demonstrated a substantial (more than three times more likely) influence of the UTA program on students to persist progress in a STEM program of study.

Therefore, the UTA program described here positively impacted persistence to the next chemistry course for students required to take that course for their intended or declared major. This was a measureable outcome within a one-semester timeframe and could be viewed as part of an overall retention goal that would span several years. Given this positive outcome on the first steps of the student STEM experience, and given the critical role of the first year college experiences, decisions, and behaviors for launching students successfully toward a STEM degree, longer term program goals for increased

STEM student retention may be achieved. Following students from this course as they progress through upcoming course work and possible future interaction with UTAs in second year courses would provide more longitudinal information about the effectiveness of the UTA program for increased STEM student retention toward degree completion.

Implications

Although the trained and supported UTAs in this study did not help college students with average or below average GPAs increase their final exam grades, the influence UTAs had on their students seemed to encourage persistence to the next chemistry course required for many STEM majors. The encouragement to persist may be far more valuable for STEM student retention than increasing grades. Further research, including a deeper investigation of student perceptions of UTAs and persistence in STEM majors using more extensive surveys of students or interviews with students, UTAs and GTAs, is warranted. More information is also needed to fully characterize the types of interactions that UTAs tend to have with their students that are different from those between GTAs and their students. For instance, UTAs did mention during conversations held in seminars that students often came to them for course or professor selection recommendation, as well as career advice. Did students ask GTAs these kinds of questions? If not, why not?

This study described the UTA selection process, semester-long training workshops and seminars, and STEM support given to the UTAs. Are all these program elements critical? Are there some elements more critical than others? Could the elements be improved so that not only encouragement to persist but also STEM academic achievement improves for more students?

Although this was the first research investigation into the PRIMES program, the results showed that the way in which the UTAs are being supported and used in the Chemistry Department is effective for student STEM persistence. Are other departments' UTAs as successful in encouraging STEM persistence? Are UTAs in other chemistry courses, such as general chemistry lab or organic chemistry, as effective at encouraging STEM persistence?

Therefore, although many questions about the effectiveness of trained and supported UTAs on student performance and persistence remain to be answered, this initial investigation into the PRIMES program implementation in general chemistry recitation sections suggests promising results for retention of STEM students.

REFERENCES

ACT, Inc. (May, 2012). *ACT benchmarks for college readiness*. Retrieved from http://www.act.org/research/policymakers/pdf/benchmarks.pdf

ACT, Inc. (2012). *ACT profile report-national*. Retrieved from http://www.act.org/newsroom/data/2012/pdf/profile/National2012.pdf

Amaral, K. E, & Vala, M. (2009). What teaching teaches: Mentoring and the performance gains of mentors. *Journal of Chemical Education 86*(5), 630-633.

Aud, S., Hussar, W., Planty, M., Snyder, T., Bianco, K., Fox, M.,… Drake, L. (2010). *The Condition of Education 2010.* (NCES 2010-028). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Augustine, N. R. (2005). *Rising above the gathering storm: Energizing and employing America for a brighter economic future.* Washington, DC: National Academy Press.

Bargh, J. A., & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology*, *72*(5), 593-604.

Benware, C. A., & Deci, E. L. (1984). Quality of learning with an active versus passive motivational set. *American Educational Research Journal, 21*(4), 755-765.

Berkner, L., He, S., and Cataldi, E.F. (2002). *Descriptive summary of 1995–96 beginning postsecondary students: Six years later.* (NCES 2003-151). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Berryman, S. E. (1983). *Who will do science? Minority and female attainment of science and mathematics degrees: Trends and causes.* New York: The Rockefeller Foundation.

Black, A. E. & Deci, E. L. (2000). The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: A self-determination theory perspective. *Science Education, 84*, 740–756.

Brainard, S. G., & Carlin, L. (1998). A six-year longitudinal study of undergraduate women in engineering and science. *Journal of Engineering Education, 87,* 369-376.

Braxton, J. M., Bray, N. J., & Berger, J. B. (2000). Faculty teaching skills and their influence on the college student departure process. *Journal of College Student Development*, *41*(2), 215-227.

Business-Higher Education Forum. (2007). *An American imperative: Transforming the recruitment, retention, and renewal of our nation's mathematics and science teaching workforce*. Washington, D.C.: Author.

Business-Higher Education Forum. (2010*). Increasing the number of STEM graduates: Insights from the U.S. STEM education & modeling project*. Washington, D.C.: Author.

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching, 44*(8), 1187-1218.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.).* Hillsdale, NJ: Erlbaum.

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

Conciatore, J. (1990). From flunking to mastering calculus: Treisman's retention model proves to be "too good" on some campuses. *Black Issues in Higher Education*, *6*(22), 5-6.

Cooper, M. M., Grove, N., Underwood, S.M., &. Klymkowsky, M. W. (2010). Lost in Lewis structures: An investigation of student difficulties in developing representational competence. *Journal of Chemical Education*, *87*(8), 869-874.

Cooper, M. M., Underwood, S. M. Hilley, C., & Klymkowsky, M. W. (2012). Development and assessment of a molecular structure and properties learning progression. *Journal of Chemical Education*, *89*(11), 1351-1357.

DeBoer, G. (1984). Factors related to the decision of men and women to continue taking science courses in college. *Journal of Research in Science Teaching*, *21*(3), 325 – 329.

Deci, E. L. (1975). *Intrinsic motivation*. New York: Plenum Press.

Education Trust. (2009). *Education watch national report*. Washington, DC: Education Trust.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117-140.

Fingerson, L. and Culley, A.B. (2001). Collaborators in teaching and learning: Undergraduate teaching assistants in the classroom. *Teaching Sociology*, *29*(3), 299-315.

Gafney, L., & Varma-Nelson, P. (2007). Evaluating peer-led team learning: A study of long-term effects on former workshop peer leaders. *Journal of Chemical Education, 84*(3), 535-539.

Galama, T. & Hosek, J. (2008). *U.S. Competitiveness in Science and Technology.* Santa Monica, CA: RAND Corporation. Retrieved from http://www.rand.org/pubs/monographs/MG674.

George M. 1996. *Shaping the future: New expectations for undergraduate education in science, mathematics, engineering, and technology*. Publication 96–139. Arlington, VA: National Science Foundation.

Goldman, R. D. & Slaughter, R. E. (1976). Why college grade point average is difficult to predict. *Journal of Educational Psychology, 68*(1), 9-14.

Gonzales, L.M. & Keane, C.M. (2010). Who will fill the geoscience workforce supply gap? *Environmental Science and Technology, 44*(2), 550-555.

Gosser, D., Roth, V., Gafney, L., Kampmeier, J.,Strozak, V., Varma-Nelson, P., Radel, S., & Weiner, M. (1996). Workshop chemistry: Overcoming the barriers to student success. *The Chemical Educator, 1*(1), 1-17.

Gosser, D. K., Jr.; Roth, V. (1998). The workshop chemistry project: Peer-led team learning. *Journal of Chemical Education, 75,* 185–187.

Grice, A., Peer, J.M., & Morris, G.T. (2011). *Today's aging workforce: Who will fill their shoes?* Proceedings of the 2011 64th Annual Conference of Protective Relay Engineers, pp.483-491, DOI: 10.1109/CPRE.2011.6035641. Retrieved from URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6035641 &isnumber=6035494.

Hake, R.R. (1998). Interactive-engagement vs traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1):64-74.

Hazari, Z., Sonnert, G., Sadler, P. M., & Shanahan, M. C. (2010). Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study. *Journal of Research in Science Teaching, 47*(8), 978-1003.

Henderson, C., Beach, A., & Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. *Journal of Research in Science Teaching*, *48*(8), 952-984.

Hewitt, N. M., and Seymour, E. (1991). *Factors contributing to high attrition rates among science, mathematics, and engineering undergraduate majors*. Report to the Alfred P. Sloan Foundation. Boulder, CO: Bureau of Sociological Research, University of Colorado.

Hilton, T. L. & Lee, V. E. (1988). Student interest and persistence in science: Changes in the educational pipeline in the last decade. *The Journal of Higher Education*, *59*(5), 510-526.

Holahan, C. J., & Moos, R. H. (1981). Social support and psychological distress: A longitudinal analysis. *Journal of Abnormal Psychology, 90*, 365–370.

Horn, L., and Berger, R. (2004). *College persistence on the rise? Changes in 5-year degree completion and postsecondary persistence rates between 1994 and 2000* (NCES 2005–156). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

House, J. D., & Wohlt, V. (1990). The effect of tutoring program participation on the performance of academically underprepared college freshmen. *Journal of College Student Development, 31*(4), 365-370.

Hug,S., Thiry, H. & Tedford, P.(2011). Learning to love computer science: Peer leaders gain teaching skill, communicative ability and content. In *Proceedings of the 42nd ACM Technical Symposium on Computer Science Education* (pp. 201-206). ACM.

Johnson, A., Brown, J., Carlone, H., & Cuevas, A. K. (2011). Authoring identity amidst the treacherous terrain of science: A multiracial feminist examination of the journeys of three women of color in science. *Journal of Research in Science Teaching, 48*(4), 339-366.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1998). Cooperative learning returns to college. *Change, 30*(4), 26-35.

Kentucky Council on Postsecondary Education. (2007). Kentucky's STEM imperative: Competing in the global economy. Frankfort, Kentucky, Retrieved from http://cpe.ky.gov/NR/rdonlyres/F42E412A-8508-4269-A50B-1E5F896CD42F/0/STEMreportFINALDRAFTwCovers.pdf

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

Lent, R. W., Brown, S. D., & Hackett, G. (2002). Social cognitive career theory. *Career Choice and Development, 4,* 255-311.

Lewis, S. E. (2011). Retention and reform: an evaluation of peer-led team learning. *Journal of Chemical Education, 88*(6), 703-707.

Lewis, S.E, & Lewis, J.E. (2008). Seeking effectiveness and equity in a large college chemistry course: *Journal of Research in Science Teaching, 45*(7), 794–811.

Maltese, A. V.. & Tai, R.H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in stem among U.S. students. *Science Education*, 95(5), 877-907.

Moss, C. M. & Brookhart, S. M. (2009). The lay of the land: Essential elements of the formative assessment process. In C.M. Moss & S.M. Brookhart (Eds.), *Advancing formative assessment in every classroom: A guide for instructional leaders.* Alexandria, VA: ASCD Press.

National Academy of Sciences - National Research Council. (1999). *Transforming undergraduate education in science, mathematics, engineering, and technology*. Washington, D.C.: National Academies Press.

National Academy of Sciences. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Washington DC: Author.

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (2011*). Rising above the gathering storm revisited: Rapidly approaching category 5. Condensed version*. Washington, DC: The National Academies Press.

National Center for Education Statistics (2009). NCES 2009-161. *Students who study science, technology, engineering, and mathematics (STEM) in postsecondary education.* Washington, DC: U.S Department of Education, Institute for Education Sciences.

National Research Council. (2011). *Expanding underrepresented minority participation: America's science and technology talent at the crossroads.* Washington, DC: The National Academies Press, 2011.

National Science Board. (2010). *Science and engineering indicators 2010*. Arlington, VA: Author.

Ong, M., Wright, C., Espinosa, L. L., & Orfield, G. (2011). Inside the double bind: A synthesis of empirical research on undergraduate and graduate women of color in science, technology, engineering, and mathematics. *Harvard Educational Review, 81*(2), 172-209.

Otero, V., Finkelstein, N., McCray, R., & Pollock,S.(2006). Who is responsible for preparing science teachers? *Science, 313*(5786): 445-446.

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.

Packer, J.E. & Woodgate, S.D. (1991). Lewis structures, formal charge, and oxidation numbers: A more user-friendly approach. *Journal of Chemical Education, 68*(6), 456-458.

Rask, K. (2010). *Attrition in STEM fields at a liberal arts college: The importance of grades and pre-collegiate preferen*ces [Electronic version]. Retrieved November1, 2012, from Cornell University, School of Industrial and Labor Relations site: http://digitalcommons.ilr.cornell.edu/workingpapers/118/

Ratelle, C. F., Larose, S., Guay, F., & Senécal, C. (2005). Perceptions of parental involvement and support as predictors of college students' persistence in a science curriculum. *Journal of Family Psychology, 19*(2), 286.

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

Raudenbush, S.W., Bryk, A.S, & Congdon, R. (2004). *HLM 6 for Windows* [Computer software]. Skokie, IL: Scientific Software International, Inc.

Romm, I., Gordon-Messer, S., & Kosinski-Collins, M. (2010). Educating young educators: A pedagogical internship for undergraduate teaching assistants. *CBE-Life Sciences Education, 9,* 80-86.

Rosser, S. V. (1995). *Teaching the Majority: Breaking the Gender Barrier in Science, Mathematics, and Engineering*. New York City, NY: Teachers College Press, Columbia University.

Rutherford, F. J., & Ahlgren, A. (1990). *Science for all Americans*: New York City, New York: Oxford University Press.

Schalk, K., McGinnis, J., Harring, J., Hendrickson, A., & Smith, A. (2009). The undergraduate teaching assistant experience offers opportunities similar to the undergraduate research experience. *Journal of Microbiology & Biology Education, 10*(1). doi:10.1128/jmbe.v10i1.97

Schoenfeld, A.H. (1987). What's all the fuss about metacognition? In A. H. Schoenfeld, (ed.) *Cognitive science and mathematics education* (pp. 189-215). Hillsdale, NJ: Lawrence Erlbaum Publishers.

Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: Why undergraduates leave the sciences*. Boulder, CO: Westview Press.

Snijders, T.A.D. (2005). Power and sample size in multilevel linear models. In B.S. Everitt & D.C Howell (Eds.), *Encyclopedia of Statistics in Behavioral Sciences* (Vol. 3, 1570-1573*)*. Chicester, U.K.: Wiley.

Stevens, J. R. (2009). *Applied multivariate statistics for the social sciences*. New York, NY: Routledge.

Strenta, A. C., & Elliott, R. (1987). Differential grading standards revisited. *Journal of Educational Measurement, 24*(4), 281-291.

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*(5): 513-547.

Strike, K. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. *Philosophy of science, cognitive psychology, and educational theory and practice*. Albany: SUNY Press.

Suidan, L., Badenhoop, J.K., Glendening, E.D., & Weinhold, F. (1995). Common textbook and teaching misrepresentations of Lewis structures. *Journal of Chemical Education. 72* (7), 583-589.

Swarat, S., Drane, D., Smith, H. D., Light, G., & Pinto, L. (2004). Opening the gateway: Increasing minority student retention in introductory science classes. *Journal of College Science Teaching, 34*, 18-23.

Tai, R.H., Liu, C.Q., Maltese, A.V., & Fan, X. (2006). Planning early for careers in science. *Science, 312*, 1143–1144.

Teddlie, C. & Tashkkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA: Sage Publications.

Thornton, R. K., & Sokoloff, D. R. (1998). Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula. *American Journal of Physics, 66*, 338-351.

Tien, L. T., Roth, V., & Kampmeier, J. A. (2002). Implementation of a peer-led team learning instructional approach in an undergraduate organic chemistry course. *Journal of Research in Science Teaching*, *39*(7), 606-632.

Tinto, V. (2001). *Rethinking the first year of college*. Higher Education Monograph Series, Syracuse, N.Y.: Syracuse University.

Tinto, V. (2006). Research and practice of student retention: what next? *Journal of College Student Retention: Research, Theory and Practice, 8*(1), 1-19.

Tobias, S. (1990). *They're not dumb, they're different: Stalking the second tier*. Tucson, AZ: Research Corporation.

Treisman, U. (1985). A model academic support system. *Improving the Retention and Graduation of Minorities in Engineering*, 55-65. Retrieved from http://www.discovery-press.com/retentionhandbook/Chapter8.pdf (September 25, 2012).

Treisman, U. (1992). Studying students studying calculus: A look at the lives of minority mathematics students in college. *The College Mathematics Journal*, *23*(5), 362-372.

U. S. Congress, Senate Committee on Commerce, Transportation. Subcommittee on Science, & Space. (1990). *Shortage of engineers and scientists*. Hearing before the Subcommittee on Science, Technology, and Space of the Committee on Commerce, Science, and Transportation, United States Senate, One Hundred First Congress, second session, on training scientists and engineers for the year 2000--the National Science Foundation's role, May 8, 1990 (Vol. 4). USGPO.

University of Surrey Psychology Department (2007). *How do I test the normality of a variable's distribution?* Guildford, Surrey, UK: The University of Surrey, Department of Psychology. Retrieved from http://www.psy.surrey.ac.uk/cfs/p8.htm

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

Walpole, M. (2003). Socioeconomic status and college: How SES affects college experiences and outcomes. *The Review of Higher Education, 27*(1), 45-73.

Ware, N., and D. Dill. (1986). Persistence in science among mathematically able male and female college students with pre-college plans for a scientific major. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, Calif.

Weidert, J. M., Wendorf, A. R., Gurung, R. A .R., & Filz, T. (2012). A survey of graduate and undergraduate teaching assistants. *College Teaching, 60*(3), 95-103.

Wheeler, L., Martin, R., & Suls, J. (1997). The proxy model of social comparison for self-assessment of ability. *Personality and Social Psychology Review, 1*(1), 54-61.

White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological review*, *66*(5), 297.

For IRB Approval Stamp

Appendix A

The Role of Undergraduate Teaching Assistants in STEM Student Achievement and Identity Development

November 7, 2012

Dear UofL student:

You are being invited to participate in a research study by answering the attached survey about your experiences interacting with the teaching assistant (TA) in this course. The goal of this study is to learn how to most effectively support the use of TAs to enhance your success in the course. There are no known risks for your participation in this research study. The information collected may not benefit you directly. The information learned in this study may be helpful to others. The information you provide will help the project team strengthen our efforts to shape the TA program to most effectively support all students' success in courses like this. Your completed survey will be stored at the offices of project faculty. The survey will take approximately 10 minutes to complete.

Individuals from the Department of Middle & Secondary Education, the Institutional Review Board (IRB), the Human Subjects Protection Program Office (HSPPO), and other regulatory agencies may inspect these records. In all other respects, however, the data will be held in confidence to the extent permitted by law. Should the data be published, your identity will not be disclosed.

Taking part in this study is voluntary. By completing this survey you agree to take part in this research study. You do not have to answer any questions that make you uncomfortable**.** You may choose not to take part at all. If you decide to be in this study you may stop taking part at any time. If you decide not to be in this study or if you stop taking part at any time, you will not lose any benefits for which you may qualify.

If you have any questions, concerns, or complaints about the research study, please contact: Dr. Thomas Tretter, 852-0595.

If you have any questions about your rights as a research subject, you may call the Human Subjects Protection Program Office at (502) 852-5188. You can discuss any questions about your rights as a research subject, in private, with a member of the Institutional Review Board (IRB). You may also call this number if you have other questions about the research, and you cannot reach the research staff, or want to talk to someone else. The IRB is an independent committee made up of people from the University community, staff of the institutions, as well as people from the community not connected with these institutions. The IRB has reviewed this research study.

If you have concerns or complaints about the research or research staff and you do not wish to give your name, you may call 1-877-852-1167. This is a 24 hour hot line answered by people who do not work at the University of Louisville.

Sincerely,

Thomas R Tretter

TA COURSE SURVEY FOR UNDERGRADUATES

Please respond to each of the following items in terms of how true it is for you with respect to interactions with your Teaching Assistant (TA) and the course section in which the TA was present. Thank you for your input. We are striving to improve the educational experiences of undergraduate in this course.

1. I found the course led by the TA to be enjoyable.
   Strongly Agree Agree Neutral Disagree Strongly Disagree
2. I found the course led by the TA to be a valuable learning experience.
   Strongly Agree Agree Neutral Disagree Strongly Disagree
3. My TA had a strong knowledge of the course content.
   Strongly Agree Agree Neutral Disagree Strongly Disagree
4. My TA explained the material very clearly.
   Strongly Agree Agree Neutral Disagree Strongly Disagree
5. My TA was able to lead effective discussions.
   Strongly Agree Agree Neutral Disagree Strongly Disagree
6. Overall, my TA was an excellent instructor and a valuable resource.
   Strongly Agree Agree Neutral Disagree Strongly Disagree
7. I feel my TA provides me with choices and options for strengthening my learning.
   Strongly Agree Agree Neutral Disagree Strongly Disagree
8. I am able to be open with my TA during class.
   Strongly Agree Agree Neutral Disagree Strongly Disagree
9. My TA conveyed confidence in my ability to do well in the course.
   Strongly Agree Agree Neutral Disagree Strongly Disagree
10. My TA encouraged me to ask questions.
    Strongly Agree Agree Neutral Disagree Strongly Disagree
11. I feel my TA cares about me as a person.
    Strongly Agree Agree Neutral Disagree Strongly Disagree
12. I don’t feel very good about how my TA talks to me.
    Strongly Agree Agree Neutral Disagree Strongly Disagree
13. My TA tries to understand how I see things before suggesting a new way to do things.

Strongly Agree Agree Neutral Disagree Strongly Disagree

14. Because of the TA support I got in this course, I am more confident in being successful in a future course I might take in this discipline.
    Strongly Agree Agree Neutral Disagree Strongly Disagree

15. I think my grade in this course is higher because of the help I got from the TA.
    Strongly Agree Agree Neutral Disagree Strongly Disagree

16. I think I understand the material in the course better than I would have otherwise because of the help I got from the TA.
    Strongly Agree Agree Neutral Disagree Strongly Disagree

17. I would do well in this course even if I did not have TA –led recitation section.
    Strongly Agree Agree Neutral Disagree Strongly Disagree

18. I regularly <u>sought</u> help from the TA.
    Strongly Agree Agree Neutral Disagree Strongly Disagree

19. I regularly <u>received</u> help from the TA.
    Strongly Agree Agree Neutral Disagree Strongly Disagree

20. How much time in minutes do you spend <u>interacting</u> with the TA each week?
    ____

21. Is attendance at this course (with TA interaction) voluntary or mandatory?
    ____

22. I think I am a "science person" or "math person".
    Strongly Agree Agree Neutral Disagree Strongly Disagree

23. Most of my family and friends think of me as a "science person" or "math person".
    Strongly Agree Agree Neutral Disagree Strongly Disagree

24. I want others to think of me as a "science person" or "math person".
    Strongly Agree Agree Neutral Disagree Strongly Disagree

Please rate your interest in the following areas:

25. I am interested in conducting my own experiments
    Very interested somewhat interested neutral not interested not interested at all

26. I am interested in talking with others about science concepts or issues
    Very interested somewhat interested neutral not interested not interested at all

27. I am interested in using mathematics to solve problems or answer questions
Very interested somewhat interested neutral not interested not interested at all

28. I am interested in knowing more about science or mathematics
Very interested somewhat interested neutral not interested not interested at all

29. Please answer the following ONLY if you intend to or have declared a major in science (e.g., biology, chemistry, geosciences, physics), mathematics or engineering:
My career plans to pursue a science/engineering/math degree have been strengthened , based in part or all on the influence of the TA on my learning in this course.
Strongly Agree Agree Neutral Disagree Strongly Disagree

30. Please answer the following ONLY if you intend to or have declared a major in a program **other than** science (e.g., biology, chemistry, geosciences, physics), mathematics or engineering:
I am planning to change my major to one in a science (e.g., biology, chemistry, geosciences, or physics), mathematics or engineering program, based in part or all on the influence of the TA in this course.
Strongly Agree Agree Neutral Disagree Strongly Disagree

31. What was your ACT Math score?___________________________

32. What was your ACT Science score? _________________________

33. What was your ACT Reading Score?_________________________

34. How many AP courses with science and math content did you complete in high school?_________________________

35. How many other AP courses did you complete in high school (Not math or science)__________________________________

36. Please list the educational level your parents have attained

Mom: HS diploma Bachelor’s degree Graduate degree Other (describe)_________

Dad: HS diploma Bachelor’s degree Graduate degree Other (describe)_________

37. How would you identify your ethnicity?_______________________________________

38. What is your gender?

____

39. What is the letter grade you expect to receive in this course?____________________

APPENDIX B

Reflection on your UTA experience (Fall 2012)

In reflecting on all of your experiences as a UTA this past semester, please consider the following questions carefully and submit your responses to Blackboard. Your responses will help the PRIMES Leadership Team improve UTA experiences in the future. Please include enough detail in your responses to communicate your ideas clearly but concisely. Please answer the questions with complete sentences.

1. Consider the statement, “Anyone can learn in a science (mathematics or engineering) class” and offer your response as to whether you consider this statement to be generally true.
    a. Why or why not?
    b. If you believe it is true that all students can learn in a science (mathematics or engineering) class, what does it take for a student to successfully learn science (mathematics or engineering) concepts and processes?

Describe how your UTA experiences contributed (maybe confirmed, maybe caused change, maybe both) to your perspective reflected in responses above.

2. Recalling the project goal to support undergraduates to have successful experiences in their science/mathematics/engineering classes, in your work with undergraduates:
    a. What went well?
    b. What was not as effective as you had hoped?

3. In your opinion, what are the most important characteristics of an effective mathematics, science or engineering instructor?

4. As an undergraduate teaching assistant
    a. What strengths did the program help you improve?

b. What strengths did you bring to the program that were helpful for the UTA task but essentially remained unchanged?

c. What attributes would you like to improve as a UTA or as a future college faculty instructor or mentor?

5. How has your UTA experience helped you grow as a scholar?

6. How would you advise a student interested in being a UTA—for what reasons would you recommend or not recommend the experience? (could be aspects of both in a response)

7. Describe the characteristics that come to mind when you think of a scientist (mathematician or engineer)

8. How are you like a scientist (mathematician or engineer)?

9. How are you not like a scientist (mathematician or engineer)?

10. What would be your ideal job after college and why?

APPENDIX C

END OF SEMESTER INTERVIEW QUESTIONS FOR UTA MENTOR FACULTY

Purpose of these questions: Contextualizing quantitative data—did UTAs impact student learning, focusing on the faculty perception of UTA content knowledge and learning approaches.

1. What did you anticipate when working with UTAs in your course? With undergrad students in ____course? Probe for both positive and negative anticipations
2. Anyone can learn science (mathematics or engineering)—is this a true statement? Why or why not? If you believe it is true that all students can learn science (mathematics or engineering), what does it take for a student to successfully learn science (mathematics or engineering) concepts and processes?
3. What went well in working with the UTAs in _____course?
4. What was not as effective as you had hoped?
5. What topics, concepts, attitudes or activities seem easy for UTAs to understand/perform? Did you expect this?
6. What topics, concepts, attitudes or activities seemed difficult for UTAs to understand/perform? Did you expect this??
7. Characterize the UTAs as teachers—individually or as types.
8. If you have had experience in working with GTA's, how does that experience compare with the UTA experience this semester?

Purpose of these questions: to evaluate how the teaching and communication skills of the UTA grew over the semester.

1. What characteristics does an effective science (mathematics or engineering) teacher/professor have?
2. How did you expect the UTAs to assist the undergraduate students in learning? What would be some expected benefits for the UTAs and their less experienced peers?

3. What are some of the strengths of your UTAs?
4. What are some of the weaknesses of your UTAs?
5. Let's talk about your experience in mentoring the UTAs. Was there a benefit to you? Was it difficult and how?
6. How have the UTAs changed, especially in the following ways: teaching efficacy, communicative abilities and content knowledge? (e.g. knowledge of student's or their own alternative conceptions, multiple representational ability, pedagogical skill-set, open-mindedness, attitude about undergraduates)
7. How has your UTA experience helped you grow as a teacher and a scholar?
8. What surprises you most about your UTA experience?
9. How would you advise a student interested in being a UTA—would you recommend the experience, what do you wish you knew or what would you do differently ?

Purpose of these questions: to examine how the identity of the UTA as a science/education professional changed during the UTA experience.

1. Describe the characteristics of scientist (mathematician or engineer) identity in general and in terms of your own professional identity.
2. How are the UTAs like scientists?
3. How are the UTAs not like scientists?
4. How have the UTAs changed over the semester as a science professional?
5. What have you learned about **students** this semester as a result of your experience mentoring UTAs?
6. What have you learned about **yourself** this semester as a result of your experience mentoring UTAs?

CURRICULUM VITAE

STEPHANIE B. PHILIPP

ADDRESS College of Education and Human Development
Porter Building Room 266 (813) 748-8155
University of Louisville
Louisville, KY 40292 stephanie.philipp@louisville.edu

EDUCATION

2009-2013	Ph.D.	University of Louisville	Curriculum and Instruction—Science Education Focus

Advisor: Dr. Thomas Tretter
Dissertation: Strengthening Stem Performance and Persistence: Influence of Undergraduate Teaching Assistants on Entry-Level Stem Students

1988-1990	M.S.	University of North Carolina-Chapel Hill	Analytical Chemistry
1984-1988	B.S.	University of Florida	Chemistry

ACADEMIC EXPERIENCE

2012-present Key Personnel, NSF STEP Grant: Partnership for Retention Improvement in Mathematics, Engineering and Science (PRIMES), University of Louisville

2012-present Graduate Research Assistant, Louisville Science Center *Science in Play* Exhibit Evaluation, University of Louisville

2011-present Key Personnel, NSF Research Initiation Grant-Engineering Education,
University of Louisville

2009-2012 Graduate Teaching/Research Assistant, Department of Secondary Instruction,
University of Louisville

2008-2009 Science Teacher, Madison Consolidated High School, Madison, Indiana

2007-2012 Visiting Chemistry Instructor, Hanover College, Hanover, Indiana

2005-2007 A.P. Physics Teacher, Plant City High School, Plant City, Florida

2003-2005 Science Teacher, W.D. Sugg Middle School, Bradenton, Florida

CHEMISTRY POSITIONS

1994-1998 **Project Manager** Warner Robins, Georgia
Geophex, Ltd.

1991-1994 **Project Manager** Research Triangle Park, North Carolina
Radian Corporation

1990-1991 **Senior Laboratory Scientist** Gainesville, Florida
Environmental Science & Engineering, Inc.

PRESENTATIONS AND PUBLICATIONS

Philipp, S.B., Tretter, T.R., Rich, C. (April 2013). Impact of undergraduate teaching assistants on STEM students' course experience. Paper presented at the *National Association for Research in Science Teaching (NARST) Annual Conference,* Rio Grande, Puerto Rico.

Philipp, S. B., Tretter, T.R., & Rich, C. (January 2013). From UTA to PST: Two Students' Pathway from Scientist to Teacher. Paper presented at the *Association for Science Teacher Educators Annual Conference*, Charleston, S.C.

Tretter, T. R., Rich, C. V., & **Philipp, S. B.** (2012, June). Improving undergraduate STEM learning by preparing and supporting effective peer mentors. Poster presented at *Science & Mathematics Teacher Imperative (SMTI) National Conference*, Arlington, VA.

Philipp, S.B., Tretter, T.R. & Rich, C. (March, 2012). The role of undergraduate teaching assistants in STEM courses. Paper presented at the *Spring Research Conference*, University of Louisville, Louisville, KY.

Tretter, T. R., **Philipp, S.B.**, and Brown, S. L. (March, 2012). Characteristics of teachers and professional development that predict growth in life science content knowledge. Paper presented at the *National Association for Research in Science Teaching (NARST) Annual Conference*, Indianapolis, IN.

Harnett, C. L., Tretter, T.R., & **Philipp, S.B.** (March 2012). Research initiation grant: Can makerspaces develop undergraduates' research creativity and innovation? Poster presented at the *Annual NSF Engineering Education Awardees' Conference*, Arlington, VA.

Philipp, S.B. & Shirley, M.L. (January, 2012). Preservice teachers' plans for questioning practice development. Paper presented at the *Association for Science Teacher Educators (ASTE) Annual Conference*, Clearwater, FL.

Philipp, S. B. & Shirley, M. L. (September, 2011). Pre-service science teachers' reflections on questioning practice. A presentation at the *2011 Mid-Atlantic Association of Science Teacher Educators Regional Conference*, Olive Hill, KY.

Philipp, S.B. (November, 2011). Teaching science concepts using case studies. A workshop presented at *Kentucky Science Teachers Association Annual Conference*, Lexington, Kentucky.

Philipp, S.B. & Shirley, M.L. (April, 2011). Understanding aspects of pre-service teacher questioning skills. Interactive poster paper presented at the *National Association for Research in Science Teaching (NARST) Annual Conference*, Orlando, FL.

Shirley, M.L. **& Philipp, S.B.** (April, 2011). Qualities of pre-service teachers' classroom questioning. Paper presented at the *National Association for Research in Science Teaching (NARST) Annual Conference*, Orlando, FL.

Philipp, S.B. & Shirley, M.L. (January 2011). Pre-service teachers' self-analysis of questioning practices. Poster presented at the *Association for Science Teacher Educators (ASTE) Annual Conference*, Minneapolis. MN.

Philipp, S. B. & Shirley, M. L**.** (September, 2010.). Development of pre-service teachers' questioning practice through self-analysis. A presentation at the *2010 Mid-Atlantic Association of Science Teacher Educators Regional Conference*, Johnson City, TN.

Philipp, S.B. (November, 2009) Let's get physical: Physical science learning activities in elementary school. A workshop presented at *Kentucky Science Teachers Association Annual Conference*, Lexington, Kentucky.

Colgate, S.O., Simon, C.G., **Boggess, S.J**., Moini, M, D'Agostino, A.T., Ammons, J.M. (1991). Gold/carbon superfine particles produced in pulsed CO_2 laser stimulated plasmas. *Journal of Vacuum Science Technology B, 9*(3), 1577-1595.

Boggess, S.J. & Colgate, S.O. (April, 1988) Gold/methane/hydrogen reaction products formed in a laser-stimulated plasma. A paper presented at the *National Conference on Undergraduate Research*, Asheville, North Carolina.

MEMBERSHIP IN PROFESSIONAL ORGANIZATIONS

American Chemical Society
Association of Science Teacher Education (ASTE)
Kentucky Science Teachers Association (KSTA)
Mid-Atlantic Association of Science Teacher Education (MA-ASTE)
National Association for Research in Science Teaching (NARST)
National Science Teachers Association (NSTA)

PROFESSIONAL SERVICE

2012-2013	Awards Committee-JRST Outstanding Paper	NARST
2012-2013	Graduate Student Forum Co-President	ASTE
2010-present	Research and Faculty Development Committee	University of Louisville
2010-present	Conference Proposal Reviewer	ASTE and NARST
2011-2012	Professional Development Committee	ASTE
2011	Mid-Atlantic ASTE Conference Committee	MA-ASTE
2008-2009	Science Olympiad Co-Advisor	Madison High School
2007-2008	Teaching and Learning Committee	Hanover College

PROFESSIONAL QUALIFICATIONS AND CERTIFICATIONS

2010-2011	Graduate Teaching Academy 2010-2011, University of Louisville, Louisville, Kentucky.
2009-present	Collaborative IRB Training Initiative (CITI) Certificate of Protection of Human Research Subjects, Social and Behavioral Basic Course
2007-present	Indiana 10-year Accomplished Practitioner Teaching License in Chemistry and Physics