Introduction
We were interested in collecting data regarding the impact of the LA program implemented by the department of Mathematics at the University of Colorado at Boulder. Does the use of LAs acting as guides to other undergraduates taking MATH 1300 (Calculus 1) improve the students’ perception of mathematics? Also, what is the impact of new conceptual worksheets on classroom culture and social dynamics among students?

The research we conducted during the 2009-10 school year involved the documentation and development of the learning assistant (LA) program in the mathematics department. Although the LA program has found success in science departments at CU-Boulder and other campuses (Otero, Finkelstein, McCray, & Pollock, 2006), there is relatively little research regarding its implementation in undergraduate mathematics. Due to the oft-recognized status of calculus as a “gatekeeper” to STEM majors and careers, our research focused specifically on the design and implementation of the LA program in undergraduate calculus. The trajectory of this research is part of a larger developmental research program with several potential contributions to STEM education:

- Improving student learning through selection, adaptation and design of mathematics tasks that promote student engagement in familiar and challenging problems;
- Design and facilitation of learning environments that support socially-mediated learning;
- Determining effective ways to utilize learning assistants in undergraduate calculus.

As part of this research, we developed interview protocols based on the instructional tasks used in tutorials to support and document student learning and understanding of essential calculus concepts. To document changes in student dispositions towards mathematics we administered
the Colorado Learning Attitudes about STEM Survey (CLASS) to students at the beginning and end of the Spring 2010 semester, as they completed their first semester of calculus. To document instructional practices and student interaction in calculus recitations we observed tutorials and completed the Reformed Teaching Observation Protocol (RTOP).

**Development of Think-Aloud Protocols**

Using two criteria for student reasoning that could be generally identified as recall/memorization and connections/justification, we independently analyzed all instructional tasks that were used in Math 1300 recitations in spring 2008, fall 2008, and spring 2009. The initial inter-rater agreement regarding the type of reasoning elicited by tasks was 71%. After discussing several tasks exemplifying recall/memorization vs. connections/justification we re-rated the tasks and improved the inter-rater agreement was 93%. The results from the second round of our task analysis are summarized in Figure 1.

![Figure 1. Percentage of task types used in Math 1300 recitations](image)

This summary reveals two trends in the use of tasks over the past three semesters: a modest increase in the use of application of derivative problems and a significant increase in the percentage of tasks that promote mathematical connections and justification. Yet these trends do not reflect intentional planning on the part of the calculus supervisor or the TAs. These data suggest the need for a deeper analysis of student reasoning elicited by these tasks so that we have actual evidence of reasoning elicited, for those students who are making satisfactory progress in calculus and those who are not. The following sections describe specific design features of tasks identified as eliciting “reasoning beyond recall.”

**Worksheet Development (2009-10 school year)**

Four different instructor committees were in charge of developing the worksheets. The first committee designed the worksheets up to the first midterm, the second committee worked on the worksheets between the first and second midterms, the third committee was in charge from the second midterm to the third midterm, and the fourth took over up until the final.

The committees were made of three to four instructors who would meet approximately once a week to develop worksheets and quizzes for each week. They would work on the projects week by week, and reflected on what material was most important for students to cover. They would then write questions meant to elicit exploratory thinking from students.

Since each committee was different and committees did not talk to one another, the worksheets differ significantly in design and purpose. They range from step-by-step instructions to arrive at an end result to open-ended questions that require more guessing and testing. The worksheets built upon material covered in the text and were meant to go beyond what was covered in class. For instance, they might ask students to prove a theorem that they had seen in class, or find counterexamples to a rule students have covered in class. Sometimes, they developed a concept that had only been introduced in lecture.

Clearly, the trend over the prior three semesters showed that the implicit goals of a more
collaborative learning environment in tutorials influenced the selection and design of tasks used in these sessions. These goals continued to inform the design of worksheets during fall 2009. To more explicitly study the reasoning elicited by such tasks, however, we developed think aloud protocols (based on these worksheets) to support the conduct of individual interviews with students.

**Report of worksheet content, level of reasoning, and think aloud structure**

There were fourteen worksheets used in MATH1300 (Calculus 1) during the spring 2010 semester. These worksheets covered a wide range of topics, including: logarithmic functions, continuity and limits, the derivative function, differentiability and non differentiability, the product rule, logarithmic differentiation, extrema, optimization problems, related rates, Riemann sums and definite integrals, the fundamental theorem of calculus, differential equations, and integration by substitution. The last worksheet did not introduce new concepts but instead was a review for the final.

The first two committees developed more level 2 and 3 questions than the last two committees. As the semester progressed, questions became more and more a way to spell out steps for the students as opposed to learning through exploration. For instance, in worksheet 3, students are asked to explain how the limit definition of the derivative describes the rate of change (question 2a). The last question on worksheet 4 asked students to “[d]escribe a real-world situation that might be modeled by a continuous function that is nondifferentiable (at one or more points in its domain).” However, worksheet 13 is a step by step guide to using integration in order to find velocity and position from acceleration and initial values.

We used these worksheets to develop think aloud protocols. We developed six think aloud protocols based on the first six worksheets. For each question on the worksheets, we thought of what we could ask to lead the student to the correct answer (“scaffold”), and then thought of questions that we could ask to take the student to the next level of reasoning and make connections with other mathematical concepts they had seen in previous classes or in lecture (“reflection”). For instance, in order to have students reflect on the relationship between the degree of a polynomial and the graph of the same polynomial we asked the following:

**Scaffold:**

- Can you think about the graph of $x^2$ and its behavior? What does this tell you about the relationship between the degree of a polynomial and the number of “turns” in its graph?

**Reflection:**

- Why do you think the degree is one more than the number of turns (when all the zeroes are real? I think it might be a little harder to think about when there are imaginary roots)
- What does it mean graphically when a function as a real root?

A complete version of this pairing and think-aloud scaffolds and reflection questions with the worksheet tasks has been included in the appendix for this report.

**Conduct of think aloud protocols**

We sent out emails to every student enrolled in MATH 1300 to recruit them for think
aloud protocols. Four students responded, and we conducted think aloud sessions with three students, because the fourth broke her leg and had to cancel the interview. The students who responded were very motivated and used the session as a tutoring session. They were willing to participate in more think aloud protocols, although we did not conduct any more. Even though we sent the recruitment emails twice, we did not get as much participation as we first expected. We might have to offer gift cards, or possibly extra-credit, for participation, as the idea of a free tutoring session was not enough to encourage students to participate.

All three interviews were approximately one hour long. We conducted the interviews as prompted by the protocols, but adapted the questions if the student needed more guidance, or if the students answered several of our questions at once. We asked all our reflection questions. We are in the process of transcribing and coding the interviews for most prominent themes.

**Interviews**

As summer break began, we were in the process of scheduling interviews with all TAs and LAs for MATH1300 in the Spring of 2010. So far, we have received one response from one of the TAs, and two from the LAs.

**Observation of Classroom Practice in Tutorials**

In this section we report the impact of the new projects on class structure and content. Specifically, we are interested in the impact on lesson design and implementation, content and propositional knowledge, procedural knowledge, and classroom culture, in particular communicative interactions and student teacher relationships.

We hypothesized that the implementation of the new instructional materials would provide both TAs and LAs opportunities to support mathematical investigations initiated by students working in small groups, as well as guided explorations of new concepts before they were presented in lecture. The lesson design incorporated proof of theorems used in calculus 1, special cases of limits, derivatives, and integration, as well as real-life related rates problems. Both TA and LA were supposed to act as a guide and help students work through the worksheet, giving students enough guidance to successfully complete the worksheet without giving away the solution.

**Method:** Both observers attended a workshop in which they compared ratings on all the items on the Reformed Teacher Observation Protocol (RTOP; Sawada et al, 2000). This process was repeated three times, and in each case, the ratings matched closely. The observers then set up observation times for all recitation sections, ensuring that each TA and each LA was observed at least once, and on separate occasions. Each TA and LA was observed three times, and in each case, the dyad was different.

**Data Source:** The RTOP was completed for all observations of tutorials. The RTOP is a tool designed to document characteristics of classroom practice that define reformed teaching. Our observers focused on five categories of the RTOP: (1) lesson design and implementation, (2) propositional knowledge, (3) procedural knowledge, (4) communicative interactions, and (5) student/teacher relationships. The RTOP includes five items for each category that are rated on a scale from 0 to 4. Each recitation and comparison class was observed using the RTOP.

**Results:** Pairs of TAs and LAs collaborated to create the instructional environment for undergraduate calculus students. However, there were multiple combinations of TA and LA pairings so that no TA-LA pair was unique to each tutorial. To test the independent
contributions of TAs and LAs to observed classroom practice, we found RTOP category means for each TA and did the same for each LA. In Figures 2 and 3, “mean of means” for each category is summarized for each TA. The table in Figure 2 also includes the standard deviation and range for each category. A cursory glance at the graph in Figure 3 shows that a range in practice does exist across TAs, from a high overall score in the mid 70s to a low overall score in the low 40s. This is a significant range in observed practice given that an overall score of less than 45 is comparable to a traditional physics lecture and an overall score of greater than 65 is comparable to an approach that promotes student inquiry and modeling. The overall scores for each TA displayed in Figure 3 were, respectively, 58.0, 58.3, 52.0, 44.0, 59.0, and 74.5. Therefore, the other four TAs were very similar in observed practice, situated somewhere between traditional and reformed mathematics instruction. However, it is hard to separate TAs from LAs, since they were observed in dyads. This problem can be remedied by using the recitation section as the unit of analysis, in which case we can say recitations are still inconsistent in offering a non-traditional and more exploratory learning environment.

Figure 2. Summary data for Math 1300 Teaching Assistants

In Figures 4 and 5, “mean of means” for each category is summarized for each learning assistant. A review of these data summaries suggest a similar range in observed practice for LAs, which is not surprising considering that LAs and TAs were instructional collaborators. The range, however, has a slight downward shift with two high overall scores in the low 60s to a low overall score in the low 40s. The overall scores for each LA displayed in Figure 5 were, respectively, 41.0, 59.0, 61.8, 61.7, and 54.3. Therefore, three LAs contributed to classroom environments that were very similar (overall score between 59 and 62) with one LA contributing to more traditional mathematics instruction.

For both TAs and LAs, ratings were highest for classroom culture-student/teacher relationship and content-propositional knowledge. This might indicate that the use of TAs and LAs who have experience with the material as guides fosters communication in the classroom as well as discussion. Ratings for content-procedural knowledge were lowest, which indicates that students had fewer opportunities to make conjectures or reflect on what they learned.

Figure 4. Summary data for Math 1300 Learning Assistants

Figure 5. Individual RTOP Category Data for Learning Assistants (n = 6)

Change in Students’ Dispositions Towards Mathematics

In addition to observing calculus tutorial sessions, we were interested in the students’ attitudes and beliefs towards mathematics. The Colorado Learning Attitudes about Science Survey (CLASS ; Adams et al, 2004; Perkins et al, 2005) is an adaptation of a survey designed for
monitoring undergraduate attitudes towards science. Historical use of the survey with undergraduates before and after completing large lecture-based science courses has typically demonstrated declines in students’ attitudes over time. The few exceptions to this result were brought about with more student-centered interactive discussions and opportunities for students to propose explanatory models.

CLASS was adapted so that the same survey could be used with undergraduate calculus students. These adapted surveys also underwent validation studies so that the analyses of student responses for each prompt and construct was compared with responses offered by an expert panel that included STEM faculty and educators at the University of Colorado.

Method

Every calculus student in the Spring 2010 semester was encouraged to participate with CLASS via extra credit. The survey was administered both at the beginning and end of the semester in order to see how student opinions may have changed over the course of the school year. Although student names were gathered in order for them to receive extra credit and to match pre data to post data, their answers were anonymous.

We conducted the experiment at the University of Colorado. All 14 calculus instructors participated and informed their students about the surveys. There were 15 sections of calculus and all employed learning assistants once a week during tutorial sessions. A total of 350 students were enrolled in calculus by the end of the Spring 2010 semester.

Data Sources

227 students took the pre survey and 191 the post survey. The total number of completed CLASS surveys was then reduced by the elimination of forms that were invalidated due to either an incorrect response to a key question to monitor student engagement or incomplete forms. In addition, each survey with the same number on the Likert scale chosen a significant number of times was closely inspected in order to discern those students who exhibit no effort. The number students then became 194 pre and 166 post. However, only 91 students were able to be matched with their pre and post results.

Results

Figure 6 shows the percentage of students with a favorable or unfavorable view in each category. Most shifts from pre to post are within the standard error. The largest shift occurs within the Sense Making/Effort category at -6.3% less students having favorable beliefs. The unfavorable beliefs in that same category jumped up by only 1.1%. This is the only category that the 95% confidence interval shows statistical significance.

Although the aforementioned category is the only one in the 95% confidence interval, there are a few other remarkable aspects to the CLASS results. According to historical trends, it is rare to find any positive shift in undergraduate STEM courses; however, we do see positive shifts close to the standard errors in PS Confidence, PS Sophistication, and Applied Conceptual Understanding. Although this shift does not fall within a 95% confidence interval, the shift in Applied Conceptual Understanding is significant using a 75% confidence interval. That is, these “close to statistical significance” shifts are not worthy of reportable findings but
are certainly worth monitoring in subsequent semesters.  

*Figure 6: CLASS summary*

In order to see just how close together most results were, Figure 7 shows a scatterplot of the table in Figure 6. The purpose of this scatterplot is to be able to visually compare the shifts from blue to red. A shift in the northwest direction is towards expert opinion, and a shift in the southeast direction is away from expert opinion.  

*Figure 7: CLASS summary graph*

A similar approach to comparing a shift in student beliefs towards expert beliefs is given in Figures 8 and 9. Figure 8 shows the results of how students agree with experts between pre and post. Figure 9 shows the results of how students disagree with experts between pre and post.  

*Figure 8: Graph of student agreement with experts*

*Figure 9: Graph of student disagreement with experts*

**Discussion of CLASS Results**

Even though most of the results were not statistically significant, there was one category in which many students no longer had a favorable belief. Moreover, the three aforementioned categories consisting of positive shifts in the students’ beliefs do warrant a closer inspection. Unfortunately, there is no control group to compare this to in order to see how changes in the calculus curriculum have affected students’ attitudes. In addition, students’ attitudes in either direction may not necessarily correlate with student performance.

**Conclusions from Year 1 study**

The purpose of this research is to improve instruction in the mathematics department and to support student learning. We use Think-Aloud Protocols to identify where students struggle and relay that information to the instructors. For example, when one of the reflection questions to a certain step in the proof of the product rule is, “Why did you do this?” and the students do not know, the instructors should be informed. We hope that the knowledge we gain from Think-Aloud encourages the committees creating the weekly worksheets to change
their tactics.  

The RTOP data show significant extremes in means, which is remarkable given that all the classes were structured the same: hand out a worksheet and answer questions about it. The variance in means came from how each TA approached the worksheet. Some TAs felt there was a need to briefly go over the ideas of the worksheets before handing them out. They often did this through a traditional mini-lecture. However, some TAs wanted their students to learn as much as possible through their own exploration, and only when there were a lot of students stuck did they go to the board to explain. There was also one particular observation where the TA was asked by his instructor to teach a topic during recitation because the class was falling behind. So, instead of filling out that week’s worksheet, he ended up lecturing for 50 minutes.

The CLASS data gave us an interesting phenomenon that involved an unprecedented positive shift in student feelings towards mathematics, especially in problem-solving. Not only did more students have favorable positions in certain categories, but fewer students had unfavorable positions as well.

All this data we collected will be used for instruction improvement and students understanding. It is important to know their opinions about math in general and see how those opinions may change based on the type of instruction they receive, which can be helped by information we collected from RTOP and the Think-Aloud Protocols.

References