LA-Test Project Overview and Research Questions ............................................................. 1
Research Questions and Description of Research Teams ........................................................ 5
Overview of Activities ............................................................................................................ 6
K-12 Team: Observations of First Year Teachers ................................................................ 11
K-12 Team: Recruitment to K-12 Teaching Careers ............................................................... 13
Conceptions of Teaching & Learning Team: Development of the FASCI Instrument .............. 23
Conceptions of Teaching & Learning Team: Characteristics of LAs ....................................... 33
Discipline-Based Educational Research Team: Physics .......................................................... 48
Discipline-Based Educational Research Team: Applied Mathematics ................................. 58
Discipline-Based Educational Research Team: MCD Biology (BCI) ....................................... 65
Discipline-Based Educational Research Team: MCD Biology (Content & Attitudes) .......... 70
Discipline-Based Educational Research Team: Chemistry ..................................................... 77
References .............................................................................................................................. 93
The Learning Assistant Model at the University of Colorado at Boulder uses the transformation of large-enrollment science courses as a mechanism for achieving three goals:

- To recruit and prepare talented science majors for careers in teaching
- To engage science faculty in the recruitment and preparation of future teachers
- To improve the quality science education for all undergraduates

The LA Program is a Teacher Recruitment Program, therefore students hired to be LAs typically have not already decided to become teachers.

One of the main purposes of the LA program is to find talented mathematics and science majors who might become interested in teaching as a result of their experiences working as an LA. Therefore, we hire a large number of LAs with the expectation that a fraction of them (typically 15%) will become interested in the complexities of teaching and decide to become K12 teachers.

The transformation of large-enrollment courses involves creating environments in which students can interact with one another, engage in collaborative problem solving, and articulate and defend their ideas. To accomplish this, undergraduate LAs are hired to facilitate small group interaction in our large-enrollment courses at the same time they make up the pool from which we recruit new K-12 teachers.

The Difference

The differences between the LA model at the University of Colorado, Boulder and other standard models for undergraduate teaching assistants are:

- The focus on teacher recruitment and preparation
- A special seminar targeted at helping LAs integrate content, pedagogy, and practice
- A collaborative educational research program to evaluate the effects of the LA model
- The involvement of science research faculty in recruitment and preparation of teachers

Course Transformation and the Role of LAs

LAs are paid a modest stipend (currently $1,500/semester) to work approximately 10 hours per week in various aspects of course transformation. Examples of courses that have been transformed using LAs are listed in table 1.
Table 1. Learning Assistant Supported Courses

<table>
<thead>
<tr>
<th>Department</th>
<th>Transformed Course</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics</td>
<td>PHYS 1010/1020: Physics of Everyday Life I and II</td>
</tr>
<tr>
<td></td>
<td>PHYS 1110/1120: General Physics with calculus</td>
</tr>
<tr>
<td></td>
<td>PHYS 2130: Modern Physics for Engineers</td>
</tr>
<tr>
<td></td>
<td>PHYS 4810/7810: Teaching and Learning Physics</td>
</tr>
<tr>
<td>Astrophysical and Planetary Sciences</td>
<td>ASTR 1010: Introductory Astronomy</td>
</tr>
<tr>
<td></td>
<td>ASTR 1120: General Astronomy Stars and Galaxies</td>
</tr>
<tr>
<td></td>
<td>ASTR 2000: Ancient Astronomies</td>
</tr>
<tr>
<td>Chemistry</td>
<td>CHEM 1021: Introductory Chemistry</td>
</tr>
<tr>
<td></td>
<td>CHEM 1111: General Chemistry</td>
</tr>
<tr>
<td></td>
<td>CHEM 4411: Physical Chemistry/Biochemistry Applications I</td>
</tr>
<tr>
<td>Molecular, Cellular, &amp; Developmental Biology</td>
<td>MCDB 1111: Biofundamentals</td>
</tr>
<tr>
<td></td>
<td>MCDB 2150: Principles of Genetics</td>
</tr>
<tr>
<td></td>
<td>MCDB 4650: Developmental Biology</td>
</tr>
<tr>
<td>Applied Mathematics</td>
<td>APPM 1340/1345: Two–semester Calculus course with Orals</td>
</tr>
<tr>
<td></td>
<td>APPM 1350: Calculus I for Engineers</td>
</tr>
<tr>
<td></td>
<td>APPM 1360: Calculus II for Engineers</td>
</tr>
<tr>
<td></td>
<td>APPM 3310: Matrix Methods</td>
</tr>
<tr>
<td></td>
<td>GEEN 1350: Calculus I Workshop-small group problem solving</td>
</tr>
<tr>
<td></td>
<td>GEEN 1360: Calculus II Workshop-small group problem solving</td>
</tr>
</tbody>
</table>

What is a Transformed Course?

There is no dictated design of what course transformation should look like. Instead, faculty members who are awarded LAs must agree to:

- use LAs to promote interaction and collaboration among students enrolled in the course
- meet in weekly planning sessions with the LAs who support their courses
- attend biweekly meetings with other faculty participating in the program
- attend a summer preparation session for using LAs
- evaluate transformations and assess learning in their own courses

Because there is little dictation as to what a transformed course should look like, there exist several models of course transformation among our participating departments. For example, one of two models of transformation in the physics department utilizes the University of Washington’s *Tutorials in Introductory Physics* in recitation sections each headed by one graduate TA and one undergraduate LA. The *Tutorials* involve conceptually-based group problem-solving activities which are based on research in physics education. LAs who work in *Tutorial sessions* formatively assess student understanding, ask guiding questions, and facilitate collaboration within groups. These tutorial sessions are supplemented by weekly lectures which are made interactive through infrared response systems and collaborative peer instruction.

A different model for course transformation is used in the Applied Mathematics department. In weekly LA-led problem-solving sessions, each small group of enrolled students uses a 2’x 3’ dry-erase board to collaboratively construct problem solutions.
A more radical form of transformation is found in the Astrophysical and Planetary Sciences department where one lecture per week is replaced by Learning Team sessions headed by LAs who facilitate collaborative problem solving as students analyze real astronomical data and generate and compare models to fit these data.

The LA Experience

Although the LA experience is somewhat different for each course, the experience for all LAs involves three related activities:

1. **Content**: they meet weekly with their faculty instructor to plan for the upcoming week, reflect on the previous week, and analyze assessment data;
2. **Practice**: LAs facilitate collaboration among learning teams by formatively assessing student understanding and asking guiding questions; and
3. **Pedagogy**: LAs from all departments attend a special Mathematics and Science Education seminar where they reflect on their own teaching and learning and make connections to relevant education literature.

The Mathematics and Science Education Course

The Mathematics and Science Education seminar is jointly conducted by a faculty member from the School of Education and a K12 teacher. In this course, new LAs reflect on their own teaching practice, reflect on the transformations of the course in which they are working, investigate relevant educational literature, and engage in in-depth discussions about their own teaching and learning. Seminar readings and discussions include topics such as questioning techniques, learning theory, cooperative learning, student epistemologies, metacognition, argumentation, multiple intelligences, the nature of science, national standards, teaching with technology, and qualities of an effective teacher. Students in this course try out new ideas each week in their learning teams and report their results in seminar. In many cases, LAs provide guidance to one another regarding managing issues that typically arise in their learning teams. Each week, LAs complete online reflections on their teaching and the learning of the students in their learning teams. In addition, throughout the semester LAs turn in two reflective essays that integrate the education literature with their own teaching and learning experiences. LAs often report that by studying and reflecting on student learning, they have become better learners themselves. At the
end of each semester, LAs in the seminar present a poster session attended by their lead instructors, School of Education faculty, University of Colorado administrators, graduate students, and their peers. Each LA or small group of LAs present a poster that focuses on aspects of the LA experience that influenced their thinking both as a learner and as a teacher.

**Focus on Teacher Recruitment**

Although the LA experience is valuable for undergraduates who continue to any career, our program is specifically designed to actively recruit talented undergraduate students to careers in teaching. Therefore, a student can continue to be an LA for a second semester only if he or she shows commitment to finding out more about teaching. This may be evidenced by taking an education course or participating in an early K-12 field experience. LAs can be hired for a third semester only if they have been accepted to a teacher certification program at which time they are eligible for NSF funded Noyce Teaching Fellowships of up to $10,000 per year\textsuperscript{ii}. As Noyce Teaching Fellows, students can become Lead LAs who mentor novice LAs, participate in the development course educational technology, or work with mathematics, science, and education faculty conducting educational research.

---

In STEM Colorado, students learn about teaching while they are teaching and while they are learning science/math content.
The LA-Test Project Research Questions

The Learning Assistant model for Teacher Education in Science and Technology (LA-Test) research project was designed to test the effectiveness of the LA model specifically in terms of LAs’ development of content knowledge, pedagogical knowledge, and their practice in K-12 schools. Faculty members from education, mathematics, and science, K-12 teachers, graduate students, and Noyce Fellows comprise three interacting research teams: the Discipline-Based Educational Research (DBER) team, the Conceptions of Teaching and Learning (CTL) team, and the K-12 team. These interacting research teams investigate teacher recruitment rates as well as the research questions shown in table 2 and synthesize results on an ongoing basis.

Each research team focuses on a specific set of questions (shown below). Answers to these questions will be synthesized by the end of the project in order to make inferences about the impact of the Learning Assistant (LA) Model on teachers’ pedagogical content knowledge.

The DBER team consists of graduate students from five departments: Physics, Chemistry, Applied Mathematics, Astrophysical and Planetary Sciences, and Molecular, Cellular, and Developmental Biology. The CTL team consists of faculty members from the School of Education in Research and Evaluation Methodology and from the Physics department, and graduate students in Science Education and in Physics Education Research. The K12 team consists of faculty members and graduate students in Science and Mathematics Education.

Research Questions:

**DBER Team:**
(a) How do LAs compare to other STEM majors in terms of their content understanding, beliefs about the discipline, and beliefs about learning in the discipline?
(b) What effects can be observed on student achievement in courses that are supported by LAs?

**CTL Team:**
(a) What is the effect of the LA model on the sophistication of LA pedagogical understanding?
(b) Does sophistication of pedagogical understanding vary by length of exposure to the LA model?
(c) How is the pedagogical sophistication of STEM LAs different from the sophistication of STEM non-LAs who become teachers?

**K-12 Team:**
How do teachers and teacher candidates who participated as LAs compare to those who did not in terms of:
(a) Practicum-based coursework
(b) Their teaching practices
(c) K-12 student attitudes and beliefs about mathematics and science
(d) Retention and attrition rates
LA-Test Project Overview of Activities for Year 1

**LA-TEST WHOLE PROJECT TEAM**

1. **Meetings.** Each research team met once per week individually, and all three research teams meet once per month. The DBER team meeting is rather large because several faculty members from participating departments have become interested in finding out more about research in science and mathematics education. There are approximately 25 faculty members at the DBER meetings. Faculty and graduate students that are actually funded through the project meet once per month (or more).

2. **Workshops.** We have held two well-attended workshops on the Learning Assistant program for faculty from universities across the nation and we will hold four more by the end of 2007. These workshops are listed below.

   Nuts and Bolts of Designing a Pedagogy Course for Learning Assistants. V. Otero and S. Iona. At the annual meeting of the Physics Teacher Education Coalition, Boulder, CO. March 1-3, 2007.

   The University of Colorado Learning Assistant Model. V. Otero and S. Pollock. At the annual meeting of the Physics Teacher Education Coalition, Boulder, CO. March 1-3, 2007.


   Learning Assistant Model for Recruiting Talented Math and Science Students to Careers in Teaching. For the University of North Carolina System. Each institution will be represented by a dean or equivalent from the college housing the physics department, a physics faculty member interested in teacher preparation, a dean or equivalent and a faculty member from the unit responsible for the pedagogy requirements of pre-service teachers. University of North Carolina, Chapel Hill, August 3, 2007.


   Up Close and Personal with the University of Colorado Learning Assistant Program. V. Otero, S. Pollock, S. Iona, N. Finkelstein, M. Klymkowsky. University of Colorado at Boulder – This workshop is for members of the Physics Teachers Education Coalition (PTEC) and is funded by PTEC (supported by the American Physical Society). October 26-27, 2007.
3. **Publications and Talks.** We have published 7 manuscripts associated with this grant and have given over 16 talks around the nation, mostly by invitation. We are working in small groups to continue to publish our research. We anticipate at least 12 publications within the upcoming year. Publications and talks are listed below.

**Refereed Journal Articles:**


**Refereed Conference Proceedings:**


**Conference Proceedings**


**Popular Publications**


**Presentations**


Publications Anticipated for 2007-2008


4. **Activities of each research team** (and smaller groups within teams) are reported individually in the following sections.
Activities

Two teachers (former LAs) entered the profession during year 1 of the project. One had majored in physics other had a double major in physics and math. We found one “match,” a student who attended the same teacher preparation program (at CU Boulder), had a similar GPA, and same major (Physics) but did not participate in the LA program. We met with these teachers and collected three types of data:

1. Observations of their teaching using the Reformed Teacher Observation Protocol (RTOP)
2. Interviews (Background and Post-Instructional). Interview protocols were developed on the basis of past research and educational theory.
3. Teacher Work Sample Portfolios from their student teaching semesters.

We will be collecting Classroom Artifact Packages (Scoop Notebooks) during Fall 2007. The K-12 team met once per week to determine a rubric that would be used to score all of the different types of data. We decided instead to continue to use the scales and rubrics associated with each instrument (RTOP, Scoop Notebooks, and Interview Protocol Categories). We created a mapping between the categories for each instrument.

Two members of the research team attended a three-day workshop in Washington D.C. on using and scoring the RTOP. When they returned they offered a 3-day workshop for other members of the K-12 team. We worked with video tapes of actual classroom teaching episodes in order to establish reliable coding among the group. The group was ultimately able to score within 10% of the one another. This is considered high rater reliability by the RTOP developers.

We conducted three interviews and two classroom observations with each of the three teachers. Results are described in the findings section.

In Fall 2007, we expect to have 10 former LAs teaching full time for their first year. We will also match this sample with approximately 10 control teachers. We will observe and interview each of these teachers and collect artifact packages from them three times per year. We are in the process of collecting Teacher Work Sample Portfolios from these teachers who student taught during the Spring 2007 semester.

In early September, 2007 we will hold workshops for the teachers who will participate in the project (former LAs and control teachers). During this workshop we will discuss compensation and what to expect throughout the year.
Findings

For the K-12 team’s group on First Year Teachers, most of year 1 was spent developing interview protocols, obtaining reliability in scoring with the Reform Teacher Observation Protocol (RTOP), and obtaining reliability in scoring Teacher Work Sample Portfolios using the Scoop Notebook (Artifact Packages) scoring guide. We only had two former LAs in the field during this year, so our data are not fully developed.

Most of our interviews with teachers took place near the end of the Spring 2007 semester and we are just beginning to code and analyze these interviews.

We have not collected any Scoop Notebooks yet, we will begin this in Fall 2007.

We used the RTOP for two observations with three different teachers, two former LAs and one control teacher. We found that the two former LAs had average RTOP scores of 71 and 61. The control teacher had an average RTOP score of 57. We look forward to collecting more data to get a better read on the situation.
K-12 Team
Recruitment to K-12 Teaching Careers
A Study that Compares LAs to their Peers
Laura Moin and Heidi Iverson

Activities

During year 1, a research assistant and a faculty member designed the research methods required to answer the big picture questions: Is the LA program indeed producing an effect? We compared the LA population with other populations of math and science majors. Our comparison population was drawn from students who attend the mid-semester LA Recruitment session but did not apply (non-applicants) or applied and were not accepted (non-LAs). In this study, we investigated whether the LA program is simply recruiting among highly interested candidates who would go into teaching careers regardless of the LA program, or whether the LA program indeed impacts students’ decisions to become teachers. We operationalized the big picture question into the following research questions:

1) What is the profile of the undergraduates who come to the mid-semester Open House Recruitment session? Are they more interested in teaching careers than the general math/science undergraduate population? Are those who apply to the program more interested in teaching careers than those who don’t apply?
2) Why do they attend the mid-semester Open House Recruitment session?
3) Why do some Open House attendants decide to apply and others do not?
4) What is the effect of participating in the LA program (pre-post measures) as opposed to just gaining a teaching experience in terms of teaching career interest?
5) What is the effect of participating in the LA program (pre-post measures) in terms of perceptions of teaching?

Our research design was as follows:

a) Identify populations of interest to compare (shown in Chart 1 at the end of this report)
b) Modify existing survey to adjust to our setting, get feedback from colleagues, revise.
c) Administer different formats but same core of the survey at four Open Houses: Fall ’05, Spring ’06, Fall ’06, Spring ’07. These data were finally analyzed through this project.
d) Design an on-line survey for attendants to the Open House sessions who decided not to apply (non-applicants) or who applied but were not accepted (non-LAs).
e) Send emails to non-applicants and non-LAs (those who applied but did not get hired) to ask them to fill out an on-line survey.
f) Gather information about applicants’ profiles before starting the program in application.
g) Collect responses to the survey.
h) Examine the application forms and extracted data to answer our research questions.
i) Conduct quantitative data analysis using SPSS and Excel.
Findings

Profile of Open House LA Recruitment session attendants.

Table 1 below shows the number of surveys collected in each term’s information session and the academic year distribution of respondents. On average about 27% of new attendants to the information sessions were females. The ethnic diversity of new attendants was: 87% White non-Hispanic, 7% Asians, 4% other and mixed ethnicities, and 1% Hispanic. In any given term, about half of new attendants to the information sessions applied to the program and the other half did not. On the aggregate, half of all the applicants got accepted to the program but there was some variation in the terms.

Table 1: Distribution of information session attendants by academic year and term

<table>
<thead>
<tr>
<th>Term or on-line survey</th>
<th>Academic year</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Freshmen</td>
<td>Sophomores</td>
</tr>
<tr>
<td>Fall ’05</td>
<td>23</td>
<td>29</td>
</tr>
<tr>
<td>Spring ’06</td>
<td>46</td>
<td>14</td>
</tr>
<tr>
<td>Fall ’06</td>
<td>22</td>
<td>35</td>
</tr>
<tr>
<td>Spring ’07</td>
<td>43</td>
<td>26</td>
</tr>
<tr>
<td>Information session totals</td>
<td>134</td>
<td>104</td>
</tr>
</tbody>
</table>

Figure 1 shows the histogram of teaching career interest among the 298 new attendants who responded the information session surveys. The percentages inside the columns indicate the percentage of respondents in each interest level. The distributions for applicants and non-applicants did not show significant differences.

The distribution of interest in K-12 math and science teaching careers among our information session attendants was within the distribution range found in comparable universities where the
general math and science undergraduate population was surveyed (Moin, Dorfield, & Schunn, 2005). Neither did we find a significant effect on teaching career interest for gender or academic year.

Attendants indicated two reasons to attend the information session from the following from a list: a) I had an [intern] in one of my course and thought that what s/he did was cool; b) An instructor told me and I thought it would be cool; c) I was told I would be good; d) I was looking for an opportunity to try teaching; e) I was looking for an opportunity to earn money; d) I was looking for an opportunity to get more involved in my discipline; e) I just tag along with some friends; f) Because of the free refreshments; g) Other reasons, please specify. There were no significant differences in the reasons for attending between applicants the information session. Responses are depicted in Figure 2 below.

Other reasons included: “Extra-curricular activity,” “I want to learn,” “Teaching is one way of understanding the material better,” “1st astronomy class and I really enjoy it,” My sister [name of sister] told me it was a great learning experience to help others,” “I came to see [name of female attendant],” “I like tutoring and want to use this as a resume builder for med school,” “It is a good work experience on my resume,” “Improve my odds for grad school,” “My [illegible] professor told me it would help develop better teaching techniques.”

Except for the reasons “I am looking for an opportunity to try teaching” and “I am looking for an opportunity to earn money,” we found no significant correlation between attendants’ level of interest in math/science K-12 teaching careers and selection of a reason to attend. For the reasons “I am looking for an opportunity to teach” the correlation coefficient was significant (r = .24, n = 291, p<.01) indicating that those with high interest in teaching career tended to select this option and those with low interest in teaching careers tended not to select it. For the reason “I am looking for an opportunity to earn money,” the correlation was also significant but in the opposite direction; those with high interest in teaching career tended not to select this option and those with low interest in teaching careers tended to select it. (r = -.20, n = 291, p<.01).
Figure 1. K-12 math/science teaching interest for new attendants to the information sessions
New attendees to the Open House session expressed their most appreciated aspect of teaching contained in the table 2 below. Applicants and non-applicants did not significantly differ in their most appreciated teaching factor, neither did those who later became LAs vs. non-LAs.

Table 2: Most appreciated aspects of teaching mentioned by new attendees

<table>
<thead>
<tr>
<th>Most appreciated teaching factor</th>
<th>Percent of respondents indicating factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impact on people’s life</td>
<td>76.3%</td>
</tr>
<tr>
<td>Affective exchange</td>
<td>30%</td>
</tr>
<tr>
<td>Creativity</td>
<td>25%</td>
</tr>
<tr>
<td>Being an authority figure or role model</td>
<td>21.3%</td>
</tr>
<tr>
<td>Possibilities of professional development</td>
<td>15%</td>
</tr>
<tr>
<td>School schedule (summer vacations, etc.)</td>
<td>9.4%</td>
</tr>
<tr>
<td>Job opportunities</td>
<td>3.8%</td>
</tr>
<tr>
<td>Prestige</td>
<td>2.5%</td>
</tr>
<tr>
<td>Low risk</td>
<td>.6%</td>
</tr>
<tr>
<td>Not demanding job was not mentioned</td>
<td>0%</td>
</tr>
</tbody>
</table>

New attendees also expressed their least appreciated aspect of teaching. However, this question
was asked with marked different wordings in the various sessions so data cannot be combined. Table 3 below shows only the factors that were similarly worded in the various sessions.

Table 3: Least appreciated aspects of teaching mentioned by new attendees

<table>
<thead>
<tr>
<th>Least appreciated teaching factor</th>
<th>% responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salary</td>
<td>46.3%</td>
</tr>
<tr>
<td>Discipline and behavior</td>
<td>30%</td>
</tr>
<tr>
<td>Dealing with parents and administrators</td>
<td>25%</td>
</tr>
<tr>
<td>Routine work</td>
<td>18.8%</td>
</tr>
<tr>
<td>Low prestige</td>
<td>5%</td>
</tr>
<tr>
<td>Being in front of the class, authority figure, role model</td>
<td>3.8%</td>
</tr>
</tbody>
</table>

Why do they attend the Open House session?
The reasons that new attended indicated for attending the Open House are summarized in table 4 below. Each respondent could indicated as many reasons as appropriate.

Table 4: Reasons for attending the Open House session

<table>
<thead>
<tr>
<th>Reason</th>
<th>Percent of respondents indicating the reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>I had an LA and though that what s/he did was cool</td>
<td>18.3%</td>
</tr>
<tr>
<td>An instructor told me and I thought it would be cool</td>
<td>28.0%</td>
</tr>
<tr>
<td>I was told I’d be good</td>
<td>29.3%</td>
</tr>
<tr>
<td>I want to try teaching</td>
<td>50.0%</td>
</tr>
<tr>
<td>Opportunity to earn $</td>
<td>19.5%</td>
</tr>
<tr>
<td>To get more involved in the discipline</td>
<td>47.6%</td>
</tr>
<tr>
<td>I tag along with a friend</td>
<td>1.2%</td>
</tr>
<tr>
<td>Free refreshments</td>
<td>2.4%</td>
</tr>
<tr>
<td>Other</td>
<td>2.4%</td>
</tr>
</tbody>
</table>

Reasons for applying to the program
Some responses to the question “Why would you like to be an LA?” indicated just one reason and others detailed several reasons. The most salient reasons included: a) helping others
understand the content, b) exploring teaching as a potential career (although not necessarily at the K-12 level), and c) learning the discipline more thoroughly. Other reasons listed less frequently were: d) learning about teaching and learning which would help in any profession, e) getting more involved in the discipline, the department, and with faculty, f) getting a convenient on-campus job related to their field of studies, g) thinking that they would be good and enjoy it, and h) desire to emulate what an intern did in a course they took.

*Reasons for not apply to the program*

The great majority of the 27 non-applicants who responded to the on-line survey referred to lack of time or schedule conflicts with other activities as the main deterrent to apply to the program. Another reason expressed was lack of confidence in being accepted due to low grades. Finally, a few respondents indicated that they were not really interested in teaching.

It seems that the main reasons for not applying were similar to the reasons for not reapplying to participate in the program a second time. In a given term, the instructor teaching the pedagogy seminar associated with the program, who was also the director of the program, posed the following question to the interns: “Are you planning on reapplying to [the program] next semester. Why or why not?” From a total of 22 responses, 9 indicated intention to reapply for consecutive terms. However, 3 indicated intention to reapply for future terms, 6 indicated they do not intent to reapply and 4 withheld responses. In the follow-up conversation that endured, those offering affirmative responses manifested increased interest in teaching careers, learning about teaching and learning, satisfaction in helping others understand difficult concepts, and enjoyment. The great majority of negative responses (either not applying for consecutive term or ever) reflected scheduling conflicts with other courses and activities.

*How does not participation in the program affect teaching interest and career paths*

The 50 non-program participant who responded to the on-line survey did not show a statistically significant change in K-12 math/science teaching career interest in the pre (Open House survey) and post (on-line) measure. There was no increase in teaching interest due to advancement in the career or maturity, but there seem to have an increase in teaching interest due to exposure to teaching. The time elapsed between pre and post responses to the teaching interest question depended upon the term of the Open Houses attended. Gain in other early teaching experiences
was significantly correlated with time elapsed between responses ($r = -.35, p < .01, n=88$).

About 14 on-line survey respondents sought a teaching experience of some sort after not participating in the program.

*What is the effect of participating in the LA program (pre-post measures) as opposed to just gaining a teaching experience in terms of teaching career interest?*

We were able to trace 42 respondents who filled out two Open House surveys, one before participating in the program and one as a continuing LA. They show a significant increase in teaching interest from 3.0 to 2.7 on a 5 point scale (1=highest interest, 5=lowest interest) but small effect size. Paired sample t-test ($p<.05$).

We conducted a one-way ANOVA on change in teaching interest among those who did not seek any teaching experience, those who sought any type of teaching experience outside of the program, and those participating in the program. We found statistically significant differences ($p < .05$). Post-hoc comparisons showed that the difference between participation in the program and not seeking any teaching experience were significant. Figure 3 below shows these results.

![Figure 3: One-way ANOVA change in teaching career interest for the 3 groups](image-url)
What is the effect of participating in the LA program in terms of perceptions of teaching?

New attendees to the Open House sessions and continuing LAs significantly differed in their perception of one of the most attractive teaching factors: 34% of new attendees indicated “affective exchange with students” among the top 2 most appreciated factors vs. 10% of the LAs. The difference may be mostly attributed to the setting of the teaching experience. There was a statistically significant difference between those who experienced K-12 teaching (n=33) 42% of whom indicated this factor vs. those with College teaching (n = 19) among whom only 5% indicated this factor (p < .01). However, appreciation of the affective exchange aspect of teaching was not correlated with measures of teaching interest.

New attendees and continuing LAs also significantly differed one of the least appealing teaching factors: no choice in what to teach or how to teach it. LAs seemed to be less bothered or to have had more choice (p<.05). Only 3% of continuing LAs (n=31) indicated this factor as opposed to 15% of new attendees (n=82).
The universe: math, science, and engineering undergraduate population

Attendants:
- Heard and attended the information session
- Heard of the program and the information session but did not attend the information session
- Haven’t heard of the program and/or the information session

Applicants:
- Applied to the program
- Did not apply to the program

Participants:
- LAs: Applied, were accepted, and participate in the program
- Non-LAs: Applied but did not become interns (were not accepted or declined)

Fellows:
- Program graduates who enroll in teacher education
- Program graduates who do not enroll in teacher education

Other candidates: Prospective teachers who didn’t participate in the program
Conceptions of Teaching and Learning Team (CTL)
FASCI Instrument

Development of the Flexible Application of Student-Centered Instruction (FASCI) Instrument to Measure Sophistication of Pedagogical Knowledge

Robert Talbot and Derek Briggs

Activities

Introduction

The FASCI instrument is intended to measure what we view as two critical dimensions of pedagogical sophistication (i.e. strategic knowledge) in science and math instruction: (1) the ability to structure classroom activities from a student-centered perspective, and (2) the ability to apply the teaching strategies that underlie these activities flexibly in context. We call the combination of these two dimensions “Flexible Application of Student-Centered Instruction” (FASCI). Our approach is to measure teachers on these two dimensions by scoring their responses to scenario-based items in which we attempt to “observe” their classroom practices indirectly. Our work is framed by what Wilson (2005) has called the four building blocks of measurement:

1. establishing a theory for the construct to be measured (the construct map),
2. specifying a design for the items that are instantiations of this theory (the items design),
3. specifying rules for scoring the items (the outcome space), and
4. specifying a model that relates scored item responses back to the construct of interest (the measurement model).

During the 2006-2007 academic year our activities have focused on work in the following areas:

- Theory Development
- Item Design
- Pilot Administration
- Item Scoring
- Presentation of Research

Each of these activities is discussed in detail below.

Theory Development

A starting point (i.e., first building block) for instrument development is to establish a theory for the construct one wishes to measure. Through ongoing weekly meetings, the FASCI team has worked to develop a theory of pedagogical sophistication of science and math teachers. It is our theory—situated within the larger conceptual framework that results from combining Shulman’s model of pedagogical reasoning (Shulman, 1986) with Schwartz et al’s model for the transfer of learning (Schwartz et al., 2005)—that helps us to hypothesize a continuum of qualitative distinctions among teachers who vary in their pedagogical sophistication. We make this
hypothesis explicit through the use of a construct map, shown in Figures 1, 2, and 3. Note that Figure 1 shows the composite FASCI construct, based on Schwartz et al’s model of transfer in learning and their dimensions of efficiency and innovation, which correspond to our dimensions of Student-Centered Instruction (SCI) and Flexible Application (FA) respectively. The construct map represented by Figures 2 and 3 represents two dimensions, analogous to the two axes in Figure 1. These dimensions also represent what we view as the key features of the ways that teachers regularly apply their strategic knowledge in classroom settings. For each of these two dimensions of pedagogical sophistication, we distinguish three levels, and the typical characteristics of teachers at each level. First, when implementing a learning activity for a classroom session, we hypothesize that a teacher’s pedagogical sophistication will vary as a function of that teacher’s ability to: 1.) view classroom activities as a means of eliciting information about student prior knowledge (which may include misconceptions) and 2.) use this information to shape subsequent teaching acts. We represent this dimension as the ability to structure classroom activities through a student-centered lens (i.e., student-centered instruction; SCI). Second, during or following a given learning activity, a teacher’s pedagogical sophistication will vary as a function of that teacher’s ability to flexibly modify classroom activities with new teaching strategies when the current activity has not produced the desired results (i.e. flexible application; FA).

![Construct Map](image)

**Figure 1.** FASCI construct, based on a Model for Learning Transfer (Schwartz et al., 2005)
<table>
<thead>
<tr>
<th>Level</th>
<th>Respondent Characteristics</th>
</tr>
</thead>
</table>
| 3     | • Teacher primarily views classroom activities as ways to help students make sense of new ideas, but for this to happen the activity should elicit information about current levels of conceptual understanding.  
• Teacher elicits students’ prior conceptions and uses this information to shape subsequent teaching strategies. |
|       | ![Teacher ↔ Students]        |
| 2     | • Teacher primarily views classroom activities as ways to help students make sense of new ideas, but for this to happen the activity should elicit information about current levels of conceptual understanding.  
• Teacher elicits students’ prior conceptions but does not use this information to shape subsequent teaching strategies. |
|       | ![Teacher → Students]        |
| 1     | • Teacher primarily views classroom activities as ways to help students make sense of new ideas. Information goes from teacher to student or from more knowledgeable student to less knowledgeable student.  
• Students may volunteer information about current levels of conceptual understanding during these activities, but the teacher does not see this as central to any given activity. |
|       | ![Teacher → Students]        |

*Figure 2. Pedagogical Sophistication: The Student Centered Instruction (SCI) Dimension*
A clarification of our definition of what constitutes a distinct teaching strategy: In presenting Shulman’s model of pedagogical reasoning, the teaching “acts” were described as distinct moves a teacher makes within a classroom session. We conceptualize a teaching strategy as an act chosen purposefully to engage students with the content or skills they are expected to learn. So while all strategies are acts as presented in Shulman’s model, not all acts necessarily constitute strategies as we are defining the term. For example, the use of humorous anecdotes by a teacher can go a long way towards establishing a classroom environment that is conducive to learning, but we would not characterize this as a teaching strategy. In the Handbook of Research on Science Education, Treagust (2007) distinguishes between six categories of teaching strategies:

1. Demonstrations
2. Explanations (e.g., telling students the answer, interactive explanations)
3. Questioning (e.g., dialog, debate, discussion, use of wait-time)
4. Representations (e.g., concrete, verbal, mathematical, visual, symbolic, gestural)
5. Group and Cooperative Learning (e.g., jigsaw, peer-tutoring, group investigations)
6. Inquiry (e.g., exploration, conceptual invention, conceptual expansion)

Treagust presents these categories as existing on a continuum from teacher-centered to student-centered, with Demonstrations being the least student-centered and Inquiry being the most student-centered. While it is easy to imagine that a demonstration can be made to be very student-centered or that an inquiry could be conducted in a teacher-centered fashion, Treagust’s conceptualization makes it easier to make distinctions among strategies.
With respect to our development of scenarios in the FASCI instrument, each item scenario (which we discuss in more detail in the next section) consists of a specified classroom activity, and implicit in each of these activities is an emphasis on one of the teaching strategies primarily characterized by one of the six categories above. While any given classroom activity will usually involve some combination of these different strategies, one will usually be most prominent. When we say that a classroom activity has been *modified*, this implies a shift in emphasis for the activity from one teaching strategy to another.

**Item Design**

A construct map is a first step in operationalizing hypothesized characteristics of pedagogically sophisticated teachers. The items design is the means by which the theory of the construct is instantiated into the prompts given to survey respondents. The items design answers the questions: Why were these particular items developed? How will they elicit information that will distinguish one teacher from another in terms of their pedagogical sophistication? Our approach has been to develop items that present respondents with a variety of classroom-based scenarios with the following general structure:

- Present a classroom scenario that involves a specified learning activity. The scenarios should vary to the extent that they appear more student-centered or teacher-centered.
  
  a) Item Prompt: How might this activity facilitate student learning?
  
  - Present information elicited from the learning activity that represents a potential obstacle to student learning.
  
  b) Item Prompt: Describe both what would you do and what you would expect to happen as a result.
  
  c) Item Prompt: If the approach you described above in (b) didn’t produce the result(s) you anticipated by the end of that class session, what would you do in the next class session?

These item prompts are the latest and most refined version. Work throughout the last year has taken us through 2 previous versions of item prompts. The item below illustrates how the template above is applied:
Students are working in groups of four to discuss a conceptual question you provided them at the beginning of class.

b) How might this activity facilitate student learning?

As the activity proceeds, one group gets frustrated and approaches you—they’ve come up with two solutions but can’t agree on which one is correct. You see that one solution is right, while the other is not.

c) Describe both what would you do and what you would expect to happen as a result.

d) If the approach you described above in (b) didn’t produce the result(s) you anticipated by the end of that class session, what would you do in the next class session?

While we have developed 10 different scenario-based items, due to time constraints the piloted version of the FASCI only contains five, along with a set of background questions about previous teaching experiences. Many of the scenarios used in the FASCI instrument were inspired by the actual experiences recounted to us by Learning Assistants through our evaluations of the STEM Colorado programs. All scenarios were written by a team of faculty and graduate students with expertise in both measurement and science education. The context for each of the scenarios is given the following constraint: “Please assume (unless it is otherwise specified) that you are teaching a high school course in physics, chemistry, biology, earth science or math to a class of 25-30 students.” While information elicited from the respondent in all three item prompts could be used to arrive at a single holistic score on both the FA and SCI dimensions, prompt (a) is primarily aligned with the SCI dimension and prompts (b) and (c) with the FA dimension.

Pilot Administration

The current version of the FASCI is currently being administered to a stratified sample consisting of 20 Learning Assistants in the STEM Colorado program, 10 students enrolled in a Masters Degree program in math or science education, 10 faculty members or graduate students with content expertise and teaching experience in math or science, and 10 practicing secondary science teachers. Pilot administration began in March 2007 and continues today. As of this writing, approximately 40 sets of responses have been collected from participants and are being scored by members of the instrument development team. The survey is being administered via a secure online survey service (www.questionpro.com).

Item Scoring

Before the current pilot administration of the FASCI, one scoring moderation session was held with members of the FASCI and LATEST teams. This occurred in October 2006. Team members scored item responses beforehand and then met to discuss their scoring. During this session, the FASCI development team received feedback which helped to inform future versions.

---

1 We have found that it takes the average respondent about 40 minutes to complete the most recent version of the FASCI.
of both the Theory of pedagogical sophistication and the scoring guides with which item responses are scored. Later in the year (February 2007) another smaller scoring moderation session was held in which four researchers met to further refine the scoring guides. The current version of each of the scoring guides is shown in Figures 4 and 5. These guides are based on the construct maps, and serve to relate the outcome space (which is defined by the expected and actual item responses) back to the construct map.

<table>
<thead>
<tr>
<th>Level</th>
<th>Explicit/active elicitation of student ideas</th>
<th>Uses ideas to shape instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>2</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>1</td>
<td>NO</td>
<td>N/A</td>
</tr>
</tbody>
</table>

*Figure 4. SCI scoring guide*

<table>
<thead>
<tr>
<th>Level</th>
<th>Modification of teaching approach</th>
<th>Discussion of contextual factors that bear on the modification of the teaching approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>2</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>1</td>
<td>NO</td>
<td>NO</td>
</tr>
</tbody>
</table>

*Figure 5. FA scoring guide*

Research Presentation

Our work on the FASCI development was shared with the educational research community at the American Educational Research Association annual meeting in April, 2007 (Briggs et al., 2007). The paper presented was part of a session on learning progressions. As we continue to work on this particular aspect of the instrument development, we plan to prepare a similar manuscript for publication.

Conceptions of Teaching and Learning Team
Development of the FASCI Instrument (continued)

Findings

Construct Explication

The development of our theory of pedagogical sophistication (and the resulting FASCI construct) is an ongoing process which is central to our efforts. While we feel that our current working theory is a good foundation for the constructs, we know that it will continually be in development. We have found that this theory development is an iterative process, as is the larger process of instrument development. As we continue to work on the FASCI, we expect our theory
of pedagogical sophistication to become more refined and better supported by the data. At this point, we feel that our theory of pedagogical sophistication of science and math teachers seems to be able to capture a range of teacher sophistications. In other words, we do see distinction between novice and expert teachers on both the Flexible Application (FA) and Student-Centered Instruction (SCI) dimensions.

**Item Development**

Scenario-based item development is difficult, but also very “fine-tunable.” Both item stems and question prompts can be changed slightly to elicit subtly different information from the respondents. While we want the items to elicit useful and scoreable responses, we do not want them to be too leading such that qualitatively different responses are scored similarly. We are finding that our pilot administration and scoring is informing us about item difficulties, and also about the specific types of item stems and prompts that are most appropriate for the FASCI. As mentioned above, our item prompts have been through multiple versions in the last year. We expect further revision of our item stems and prompts as more data is gathered and scored.

**Respondent Reactions/Feedback**

Respondent reaction to answering the scenario-based questions has varied widely. At the end of the instrument, there is an opportunity for respondents to provide us with any comments. Three examples of these comments are given below:

ID=2002667: “The questions were very vague and it made some very hard to answer.”

ID=1991969: “I felt that, by design, I was repeating some of the answers for the different parts, but then I realized that there are important subtle differences in each part. I think many participating will feel that it is repetitious and redundant, and not recognize the subtle differences in the parts, but they may have lots to say.”

ID=1976723: “This was an effective survey in highlight which methods of teaching are more effective. It also brought to my attention that I should follow up on my students understanding more so than I do now. Sometimes, the last question of each section was not applicable. Thanks!”

As expected, some respondents have expressed frustration at repeatedly answering the same item prompts for different stems. Further, the “vague” nature of scenarios is frustrating to some (e.g. respondent #2002667), but as mentioned above we feel that too much detail or context could make the questions too leading. Finding this balance between the openness of the question and the specificity has proven to be challenging indeed. Also as expected, some respondents (e.g. respondent # 1976723) found participating to be a useful exercise which helped them to think about their teaching. Along this line, we intend to explore the idea of using the FASCI scenarios as discussions in a teaching methods course next year. As we continue to gather more data, we will be interested to see respondent reactions to participation and we will continue to take these comments into account in further design efforts.
Example Data

We are finding a range of sophistication in responses to the FASCI. Below are examples of responses which we consider to be highly sophisticated on the SCI and FA dimensions respectively. (Refer to Figures 1 through 5 for construct maps and scoring guide locations which further define these responses.)

SCI score = 3 (High)
ID=2195315: “[I would] think of an application in which the students’ idea obviously fails. Let the student deal with that application and confront the failure of their idea.”

This respondent clearly uses the students’ idea to shape subsequent teaching action: presenting them with an idea to struggle with in which their conceptions would fail.

FA score = 3 (High)
ID=2191029: “Depends on the question and its role in the class. If the question was fundamental and something we needed to build on right away, it's conceivable that I might *discuss* the two solutions, have OTHER groups help argue out which was right (and why) and which was wrong (and why). If time was short, I might just present such a 'discussion' myself. If there was time and other groups were also struggling on the same point, I might built a secondary activity to approach the same concept from a new angle (but that seems pretty unlikely, given how rushed and structured school tends to be!”

This respondent mentions the context dependence (questions’ role in the class, time available) as determining in part what teaching action they will take next. The qualifier “it depends” often leads to a response which is indicative of high FA. It also appears that this respondent has a large repertoire of teaching activities from which to draw (teacher-centered discussion, student-centered group discussions, or a new “secondary activity”), thereby increasing his/her ability to be flexible.

Following are example responses in which SCI and FA were rated as low.

SCI score = 1 (Low)
ID=1961454: “I would do this [solve an example problem on the board] so that students would have a record in their notes of how to solve the problem the correct way. Also if they see the problem done on the board and they copy it down, that is two chances for them to remember the correct way to solve the problem.”

This response shows an individual who (in this particular situation) views the activity as an opportunity for students to gain the knowledge from an all-knowing expert. No attempt is made to elicit students’ ideas about the topic under study or to incorporate active student participation in the presentation of this example problem. It is also important to note that we are finding that this particular scenario (which deals with working an example problem on the board) is very difficult for respondents to score highly on in the SCI dimension.

FA score = 1 (Low)
ID=2003336: “I would give a brief but thorough explanation of the topic at hand, and then offer a 'make-up' quiz. Then I would repeat the previous process, if necessary.”

This respondent does not modify their teaching approach based on prompt c (“the approach you tried still doesn’t work…”). They try the same thing over again and do not discuss any contextual factors which may bear upon their decision.

Inferences

Although it is too early for us to make very many strong inferences about the existence of the FASCI construct in our respondents, we expect to be able to do so within the next year. As mentioned above, we do see a range of the amount of the FASCI construct present in our respondents. As we continue to gather data and further develop the FASCI instrument, we will be in a position to make stronger claims about the pedagogical sophistication of science and math teachers whom we have surveyed.
Activities

In addition to the development of an instrument that can measure pedagogical sophistication, a faculty member and graduate student from the Physics Education Research group work with the CTL team mine existing data to determine various characteristics of students who were hired to become Learning Assistants and those who applied and were not hired or those who did not apply.

Existing data that were investigated are listed below. The assessments listed in (1) and (2) below were given to all the students enrolled in LA-supported courses as well as to the Learning Assistants.

1. Assessment data from conceptual instruments (such as the Force and Motion Conceptual Evaluation) that were give pre/post-at the beginning of the semester and at the end.

2. Colorado Learning Attitudes about Science Survey (CLASS) data given pre/post. This instrument measures students’ views about science, science learning, and their interest in the topic.

3. Learning Assistant weekly reflections on their own teaching – findings from this study are not complete and will not be reported in the findings section.

In this study, Learning Assistants’ conceptual understanding (as measured by the conceptual instruments) and their views of science, science learning, and interest in physics were measured both when they were a participant in the course as a student (they were enrolled in the course) and in later semesters after they became learning assistants. As discussed in the findings section, post-data for future-LAs reflect student characteristics after they were informed that they were accepted (or not accepted to be an LA).
Characteristics of Learning Assistants

Findings

Introduction

The learning assistant experience is a unique opportunity for undergraduates to explore the practice of teaching. We would like to understand what effect this experience has on the undergraduates involved: how it impacts their content knowledge, beliefs about teaching and learning physics, and about the nature of physics and science in general. We are taking a two-step approach to address this query. First, we attempt to establish who applies to and is hired to be a learning assistant. This provides some insight into the background of our learning assistants. Second, we use quantitative measures, as well as qualitative evidence, to understand what changes occur in the factors previously named. Our focus is on Physics 1110 and 1120, the students who take these courses and apply to be learning assistants, as well as those students who are learning assistants in these courses. The majority of learning assistants in the physics department work in these two courses.

Findings

Applicants to LATEST

A. Physics 1110


<table>
<thead>
<tr>
<th>Semester</th>
<th>N (# students)</th>
<th>Pretest (%)</th>
<th>Posttest (%)</th>
<th>Gain of averages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>~ +/-1%</td>
<td>~ +/-1-2%</td>
<td></td>
</tr>
<tr>
<td>Sp07: Non-Applicants</td>
<td>361</td>
<td>29.9</td>
<td>62.1</td>
<td>.46</td>
</tr>
<tr>
<td>Sp07: Applicants*</td>
<td>11</td>
<td>51.2</td>
<td>84.8</td>
<td>.69</td>
</tr>
<tr>
<td>Fa06: Non-Applicants</td>
<td>325</td>
<td>33.6</td>
<td>67.1</td>
<td>.50</td>
</tr>
<tr>
<td>Fa06: Applicants*</td>
<td>14</td>
<td>62.6</td>
<td>90.3</td>
<td>.74</td>
</tr>
<tr>
<td>Sp06: Non-Applicants</td>
<td>262</td>
<td>27.0</td>
<td>58.7</td>
<td>.43</td>
</tr>
<tr>
<td>Sp06: Applicants**</td>
<td>21</td>
<td>37.5</td>
<td>77.8</td>
<td>.64</td>
</tr>
<tr>
<td>Fa05: Non-Applicants</td>
<td>291</td>
<td>34.1</td>
<td>59.0</td>
<td>.38</td>
</tr>
<tr>
<td>Fa05: Applicants**</td>
<td>12</td>
<td>34.6</td>
<td>74.5</td>
<td>.61</td>
</tr>
<tr>
<td>Fa04: Non-Applicants</td>
<td>392</td>
<td>33.2</td>
<td>69.0</td>
<td>.54</td>
</tr>
<tr>
<td>Fa04: Applicants^</td>
<td>6</td>
<td>39.4</td>
<td>92.9</td>
<td>.88</td>
</tr>
<tr>
<td>Sp04: Non-Applicants</td>
<td>333</td>
<td>28.9</td>
<td>73.9</td>
<td>.63</td>
</tr>
<tr>
<td>Sp04: Applicants</td>
<td>4</td>
<td>32.6</td>
<td>87.1</td>
<td>.81</td>
</tr>
</tbody>
</table>

“Applicants” includes all students in 1110 in the specified semester who at some point in their undergraduate career applied to be a learning assistant. *Significant (p<0.05) difference between pretest scores of non-applicants and applicants. **Moderately significant (0.05<p<0.1) difference between pretest scores of non-applicants and applicants.
While there is a significant difference between FMCE gains of non-applicants and applicants, there is no significant difference between applicants who are hired and those who are not hired on this measure.

<table>
<thead>
<tr>
<th>Semester</th>
<th>N (# students)</th>
<th>Pretest (%)</th>
<th>Posttest (%)</th>
<th>Gain of averages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fa06: Not Hired</td>
<td>9</td>
<td>66.3</td>
<td>88.9</td>
<td>.67</td>
</tr>
<tr>
<td>Fa06: Hired</td>
<td>5</td>
<td>55.8</td>
<td>92.7</td>
<td>.84</td>
</tr>
<tr>
<td>Sp06: Not Hired</td>
<td>15</td>
<td>38.6</td>
<td>77.0</td>
<td>.63</td>
</tr>
<tr>
<td>Sp06: Hired</td>
<td>6</td>
<td>34.8</td>
<td>79.8</td>
<td>.69</td>
</tr>
<tr>
<td>Fa05: Not Hired</td>
<td>9</td>
<td>32.3</td>
<td>77.8</td>
<td>.67</td>
</tr>
<tr>
<td>Fa05: Hired</td>
<td>3</td>
<td>41.4</td>
<td>64.6</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Using a student’s t-test on the pretests, the difference between those applicants who were not hired and those who were are not found to be significant. Using error propagation, differences between the gains of the two populations are not found to be significant.

The data shows us that students who apply to be a learning assistant have a strong understanding of physics content when compared to other introductory physics students. Interestingly, the data also shows that in recent semesters applicants also have higher pretest scores, suggesting that these students may be entering the introductory course with some background in physics (most likely from high school physics or physical science). We compared the pretest distributions of all students in the Sp04 and Fa06 semesters to explore the possibility that the population of students in this course is changing. We found no significant difference between these two courses, suggesting that we are drawing from a different part of the population, while the population as a whole does not change.
Pretest scores on FMCE for two semesters. In Sp04, there was no significant difference between the pretests of applicants and non-applicants. In Fa06, the difference was highly significant (p = 0.001). A comparison of the two courses suggests that the general population of students is not changing.

2. Grades
Since there was no measurable difference between hired and not hired applicants on the FMCE, we next looked at final course grades to see if there are differences in the populations.

Average grades in 1110. In Fa04 and Sp07, the number of hired applicants (N = 2) was too small to represent graphically. Final course grades are not available for Fa05. Note that this does not differentiate between students who apply while enrolled in 1110 and those who apply after completing 1110.
There appears to be a strong correlation between whether or not a student applies to be a learning assistant and their final course grade. This is an expected result – the program actively recruits high performing students. It appears that applicants who are hired have higher grades than those who are not, however there are not enough semesters represented to know if this is a recurring trend. In Fa06, there does not appear to be a difference between the two groups’ grades.

3. Attitudes and Beliefs: CLASS (Colorado Learning Attitudes about Science Survey)  
Shifts in overall CLASS, personal interest, and conceptual understanding (as measured by the CLASS) were calculated. Data is combined from Sp06, Fa06, and Sp07. The population labeled “Same Grades” is a group of students from the Non-Applicants who have the same grade distribution as the applicants. Applicants have been limited to those students who applied to be a learning assistant while enrolled in 1110.

Overall CLASS shifts in 1110. There is a significant difference (p<0.05) between the pretests of applicants (both hired and not hired) and non-applicants, and between applicants and the “Same Grades” population.
Personal Interest shifts in 1110 (as measured by CLASS). There are significant differences in the pretests of applicants (both hired and not) and non-applicants, as well as between applicants and “Same Grades.”

Conceptual Understanding shift in 1110. There is a significant difference between the pretests of those not hired and the non-applicants, but not between those hired and the non-applicants.
Applicants to the learning assistant program appear to have more expert-like beliefs about physics than those who do not apply, even in their first semester of physics at the university level. This is not simply a relic of these being high performing students – even when compared with students who do as well in the course, they have significantly more expert-like attitudes and beliefs.

The CLASS is structured such that students are given a statement and students must indicate to what extent they agree with the statement on a five point Likert-scale. Statements on which there was a significant difference between the responses of applicants and non-applicants are:

- In physics, mathematical formulas express meaningful relationships among measurable quantities.
- When I am solving a physics problem, I try to decide what would be a reasonable value for the answer.
- After I study a topic in physics and feel that I understand it, I have difficulty solving problems on the same topic.
- I find carefully analyzing only a few problems in detail is a good way for me to learn physics.
- I can usually figure out a way to solve physics problems.
- I think about the physics I experience in everyday life.
- I am not satisfied until I understand why something works the way it does.
- I study physics to learn knowledge that will be useful in my life outside of school.
- To understand physics I discuss it with friends and other students.
- I enjoy solving physics problems.
- Reasoning skills used to understand physics can be helpful to me in my everyday life.

One of the main themes these statements is the incorporation of physics into the student’s life. Applicants also had significantly greater interest in physics.

Applicants who were and who were not hired did not differ significantly in their responses.
1. Conceptual Learning: BEMA (Brief Electricity and Magnetism Assignment) Pre and post, matched only.

<table>
<thead>
<tr>
<th>Semester</th>
<th>N (# students)</th>
<th>Pretest (%)</th>
<th>Posttest (%)</th>
<th>Gain of Averages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sp07: Non-Applicants</td>
<td>243</td>
<td>N/A</td>
<td>51.7</td>
<td>N/A</td>
</tr>
<tr>
<td>Sp07: Applicants</td>
<td>15</td>
<td>N/A</td>
<td>62.8</td>
<td>N/A</td>
</tr>
<tr>
<td>Fa06: Non-Applicants</td>
<td>338</td>
<td>25.8</td>
<td>54.9</td>
<td>.39</td>
</tr>
<tr>
<td>Fa06: Applicants</td>
<td>18</td>
<td>33.2</td>
<td>70.9</td>
<td>.56</td>
</tr>
<tr>
<td>Sp06: Non-Applicants</td>
<td>201</td>
<td>26.4</td>
<td>53.0</td>
<td>.36</td>
</tr>
<tr>
<td>Sp06: Applicants</td>
<td>15</td>
<td>30.0</td>
<td>63.0</td>
<td>.47</td>
</tr>
<tr>
<td>Fa05: Non-Applicants</td>
<td>320</td>
<td>25.1</td>
<td>49.8</td>
<td>.33</td>
</tr>
<tr>
<td>Fa05: Applicants</td>
<td>22</td>
<td>29.6</td>
<td>57.0</td>
<td>.39</td>
</tr>
<tr>
<td>Sp05: Non-Applicants</td>
<td>224</td>
<td>27.0</td>
<td>58.3</td>
<td>.43</td>
</tr>
<tr>
<td>Sp05: Applicants</td>
<td>8</td>
<td>37.1</td>
<td>69.7</td>
<td>.52</td>
</tr>
</tbody>
</table>

“Applicants” includes all students in 1120 in the specified semester who at some point in their undergraduate career applied to be a learning assistant. Pretests were not given in Sp07. There is no significant difference in pretests of applicants and non-applicants. There is a highly significant difference (p < 0.05) in posttest between the two populations in Sp07, Fa06, and Sp06, and a moderately significant (0.05<p<0.1) difference in Fa05 and Sp05.

Re-presentation of above data. BEMA gains (gain of the average pre and post tests). “Applicants” includes “Hired Applicants”.

While there is a significant difference between the gains made on the BEMA between applicants and non-applicants, there does not appear to be (in the limit that we are working with small numbers of students) a difference between applicants who are and who are not hired. Similarly to applicants from 1110, applicants that take 1120 appear to do better on standard measures of conceptual learning in these introductory courses. It is important to note, however, that students who are in the 1110 applicant pool may also be in the 1120 applicant pool.
2. Grades

Average grades in 1120. In Sp05, the number of applicants who were not hired was too small (N = 2) to represent independently. This does not differentiate between those students who applied during 1120 and applied after 1120, though information on applicants’ progress in the course to date was available during the application process.

As with the 1110 data, there appears to be a strong correlation between course grade and whether or not a student applies to be a learning assistant. There is not, however, evidence in the data that students are chosen to be learning assistant based on their 1120 grades. In all semesters for which there is data, applicants who are not hired and those who are had comparable grades in course.

3. Attitudes and Beliefs: CLASS (Colorado Learning Attitudes about Science Survey)

Shifts in overall CLASS, personal interest, and conceptual understanding (as measured by the CLASS) were calculated. Data is combined from Sp06, Fa06, and Sp07. The population labeled “Same Grades” is a group of students from the Non-Applicants who have the same grade distribution as the applicants. Applicants have been limited to those students who applied to be a learning assistant while enrolled in 1120, though it is noted that eventually attention must be paid to students who apply after completing 1120 (or who apply without taking either 1110 or 1120).
Overall CLASS shifts in 1120. There is a significant difference between the pretests of applicants and non-applicants, however there is only a moderately significant ($p = 0.053$) difference between the “Same Grade” population and applicants.

Personal Interest shifts in 1120 (as measured by CLASS). There are significant differences in the pretests of applicants (both hired and not) and non-applicants, as well as between applicants and “Same Grades.”
Conceptual Understanding shift in 1120. There is a significant difference between the pretests of those hired and non-applicants, but not between non-applicants and those who are not hired. There is not a significant difference between the pretests of applicants and the “Same Grade” population.

As in 1110, we observe an applicant population with more “expert-like” beliefs about physics. This difference is most noticeable in the category of personal interest, and much less noticeable in the conceptual understanding category. Of interest is the fact that applicants show positive shifts on the CLASS overall during their 1120 semester, while non-applicants show negative shifts.

Statements on which applicants and non-applicants differed significantly are:

- Knowledge in physics consists of many disconnected topics.
- Understanding physics basically means being able to recall something you've read or been shown.
- Spending a lot of time understanding where formulas come from is a waste of time.
- If I get stuck on a physics problem on my first try, I usually try to figure out a different way that works.
- If I don't remember a particular equation needed to solve a problem on an exam, there's nothing much I can do (legally!) to come up with it.
- If I want to apply a method used for solving one physics problem to another problem, the problems must involve very similar situations.
- There are times I solve a physics problem more than one way to help my understanding.
- To understand physics, I sometimes think about my personal experiences and relate them to the topic being analyzed.
- I enjoy solving physics problems.
- I am not satisfied until I understand why something works the way it does.

As in 1110, there is one theme in these statements of relating and incorporating physics into one’s life. Another theme, however, is related to how the students actually do physics. In the 1110 statements, the two populations differed in their confidence in their ability to do physics. In the 1120 statements, a big difference appears to be how the two populations believe physics should be done, as well as what the role of formulas or equations is in physics.

**Learning Assistant Instructors**

1. **Conceptual Understanding**: It has been shown (see Y1_Physics_LATEST) that Learning Assistants have conceptual gains on the FMCE and BEMA when they are working as learning assistants for 1110 and 1120 respectively.

2. **Attitudes and Beliefs: CLASS**

   ![Chart showing shifts on CLASS categories for learning assistants working in 1110. The difference between pre and post of conceptual understanding and overall is significant using a matched student’s T-test.](chart.png)
Shifts in CLASS for learning assistants working in 1120. Using a matched student's T-test, the differences between pre and post are not significant.

It is difficult to judge the significance of these shifts without comparing to some “control” group. Comparing students who applied while enrolled in 1110 and went on to take 1120, we can use students who are not hired as a control to see if the learning assistant experience affects the attitudes and beliefs of students.

Comparison of students who were working as learning assistants in 1110 while enrolled in 1120 (“Hired”) to students who were not hired to work in 1110. During the application process, there was no significant difference between the two populations. There is no difference in the pretests of the two groups, however there is a significant difference (p <0.05) between the overall and personal interest post tests. Furthermore, shifts for the learning assistants were all positive, while shifts for the other group were all negative.
These initial results suggest that the learning assistant experience uniquely affects the students involved. To explore this more, we would like to eventually look at CLASS performance of LAs in upper division physics courses.

3. **Qualitative Data**

There are sources of data that may shed light on the changes that occur in the beliefs of learning assistants. These are more qualitative in nature – they include surveys which the learning assistants fill out at the end of each semester, as well as reflections which learning assistants are expected to submit each week. While I have not yet been able to obtain access to the surveys, I do have one semester of reflections submitted online. The reflections do little to highlight changes the learning assistants go through, however they do suggest that each learning assistant has their own idea of what their role in the classroom is, and the reflections may provide some insight into what those roles are. Below are a few examples of LA statements (direct quotes) from these reflections, and how they might be interpreted.

**Student 1:**

*(In response to prompt 1: Main goal for the day and plan for achieving that goal)*
- Asking better questions
- Fostering discussion
- by joining in on tangent conversations and then steering back to the topic at hand.
- Asking questions that were less clearly leading, but meant to foster discussion.
- ask more open questions
- Trying to listen to what the students truly think

*(In response to prompt 4: Did all the students appear to be participating? Any good discussion? Any problems with the group, and if so how do you plan on addressing that problem?)*
- Some students just go off and do their own thing. Sometimes asking them what they think about something will bring them back in, but sometimes only for a few minutes.
- ... I made an effort to be more of a person, and less of a teacher.

This LA seems to see their role as encouraging and developing discussion of physics in the tutorials. They attempt to be “less of a teacher,” and instead they focus on listening to students and fostering discussion.

**Student 2:**

*(In response to prompt 1)*
- To adequately address the misconceptions that people had... and demonstrate the logical conclusions that could be drawn
- Students were generally able to work through the tutorial... My focus for today was to treat problems on an individual basis... until I was sure they had the right idea.

*(In response to prompt 2: What Worked & What Didn't Work)*
there were times when I would directly address one person, and ask them questions
Some students responded better to a logical approach to why forces must have an equal and opposite reaction force. Other students responded better to examples, but some just gave me blank stares.

(In response to prompt 5: My Reflection : I am really glad that I... Why didn't I... Next time I am going to... My biggest problem during session was...)
- I really wanted to skip ahead and tell students the conclusions of the tutorial, because so many just had no idea where it was going.
- I was able to directly get up and address the whole class, because the TA was sick. This gave me a chance to focus and emphasize the subjects that I felt were important to emphasize, and therefore it was easier to do it on a global scale rather than individually.

This LA sees their role as a teacher, whose responsibility is to get students to understand concepts. This LA will talk to students one-on-one to help them understand (even at the expense of other students, as suggested in other comments). This student likes to “demonstrate” and explain concepts to help students understand. The goal of the class seems to be to work through the tutorial, conceptual understanding is supposed to come with this end.

Next Steps

Applicants

It is still not clear what differentiates students who are hired and those who are not. A quick analysis of application essays does not reveal any strong trends between the two populations. A next step to shed light on this question would be to interview faculty members involved in the application process to see what factors they value.

LA Instructors

It is apparent that being a learning assistant does have some impact on the conceptual understanding, as well as the attitudes and beliefs of learning assistants. It is less obvious how this change occur, and how the LA’s understanding of their role in the classroom may impact these changes. With the addition of more reflections, as well as a more detailed analysis of these, I believe that we can come to some conclusions about the various roles LAs believe they are expected to fill. After that, it may be insightful to compare CLASS shifts to the roles LAs embody. The surveys that learning assistants fill out at the end of the semesters may also shed light on what is happening with the learning assistants. It may be useful to conduct interviews throughout the semester, and to have the learning assistants fill out paper copies of their reflections. I believe that one reason for the thinness of the reflections (they tend to be repetitive and not always completed) is that they are completed online, and the LAs do not believe anyone is reading and responding to them.
Discipline-Based Educational Research (DBER) Team

Physics

A study on attitudes, conceptual understanding, gender, and a longitudinal study

Steve Pollock and Noah Finkelstein

Activities

The DBER Physics group taught the LA-Supported Courses listed below and collected pre and post Conceptual and additudinal data in all courses. We used tested instruments such as the Force and Motion Conceptual Evaluation, the Brief Electricity and Magnetism Survey, and the Colorado Learning Attitudes about Science Survey. The courses that we worked with during year 1 of the project are listed below.

*Physics 1110* (Physics I, mechanics). Enrollment typically 5-600 each semester. There are three 50 minute lectures a week (using Peer Instruction) and one 50-minute Tutorial section for ~24 students, led by a grad physics TA and 1-2 LAs.

*Physics 1120* (Physics II, E&M). Enrollment typically 350-500 each semester. There are three 50 minute lectures a week (using Peer Instruction) and one 50-minute Tutorial section for ~24 students, led by a grad physics TA and 1-2 LAs.

*Physics 4810/7810* (Teaching and Learning Physics). Enrollment typically 12-24, mixed graduate and advanced undergraduate. Meets twice a week for 75 minutes.

Current course syllabi can be found at

http://www.colorado/edu/physics/phys  [1110, or 1120, or 4810]

For earlier semesters, the naming convention is e.g.
http://www.colorado/edu/physics/phys1110/phys1110_fa04

NOTE: LAs have been used in other physics courses (especially Phys 1010/1020, the Physics of Everyday Life, and Phys 2130, Modern Physics for Engineers) but we do not have the same kind of systematic data collected for these courses as we do for the ones presented in the Findings section of our year 1 LA Test report.
Findings

Course development and transformations have been going on since before the LA-TEST grant, with our first Tutorials implemented in 2003, and data collection beginning in 2004. We have observed and documented the following:

- improve student content mastery in:
  - mechanics (Physics 1)
  - electricity and magnetism (Physics 2)
  - teaching and learning physics (Phys 4810/7810)
- often succeed at supporting favorable attitudes and beliefs:
  - there are some instances of degradation on the CLASS (phys 1), similar to other published work (See Adams, Phys Rev 2, 010101)
  - several of the implementations have seen no degradation on the CLASS, which we consider significant success
- provide lasting impact on content mastery
  - in upper division physics we observe difference between those students have had Tutorials and those who have not
  - there is differential positive impact for those students who were LAs
- depend upon which / how faculty implement these curricula

A. Physics I:

The FMCE (Force and Motion Concept Evaluation) were given pre and post in Phys 1110 (introductory, calculus based physics):

1. Conceptual Learning: Learning Assistant data
   (not all matched - typically missing ~1 for pre and/or post)
   Phys 1110, Sp 2007. For 7 LAs, pre ave was 87.4, post ave was 94.8, gain of ave = .59
   Phys 1110, Fa 2006. For 7 LAs, pre ave was 88.7, post ave was 98.5, gain of ave = .87
   Phys 1110, Sp 2006. For 5 LAs, pre ave was 98.3, post ave was 100.0, gain of ave = 1.0
   (No LAs were used for the 3 semesters preceding, no data for pilot period before that.)

2. Conceptual Learning: Data for classes as whole:

<table>
<thead>
<tr>
<th>Semester</th>
<th>N (# students)</th>
<th>Pretest (%)</th>
<th>Posttest (%)</th>
<th>Ave normalized gain</th>
<th>Gain of averages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(matched only)</td>
<td>~ +/-1%</td>
<td>~ +/-1-2%</td>
<td>~ +/- .02</td>
<td></td>
</tr>
<tr>
<td>Sp 2007</td>
<td>363</td>
<td>28.6</td>
<td>61.7</td>
<td>.49</td>
<td>.46</td>
</tr>
<tr>
<td>Fa 2006</td>
<td>331</td>
<td>33.2</td>
<td>67.1</td>
<td>.53</td>
<td>.51</td>
</tr>
<tr>
<td>Sp 2006</td>
<td>278</td>
<td>26.5</td>
<td>59.5</td>
<td>.47</td>
<td>.45</td>
</tr>
</tbody>
</table>
Fa 2005 *  293  32.1  58.3  .42  .39  
Sp 2005 *  213  28.4  58.5  .45  .42  
Fa 2004 **  392  33.2  69.0  .59  .54  
Sp 2004  335  28.2  73.9  .66  .64  
* means no Tutorials (and no LAs): traditional recitations, a "control" of sorts.
** means no Tutorials (and no LAs) - but they utilized small-group recitations centered around research-based
workbook materials.

All the rest used UW Tutorials and LAs (Sp04 was taught by one of the Co-PIs) Note that although all our courses demonstrate solid learning gains, the two semesters with traditional recitations show the lowest two normalized gains. The best terms show very high normalized gains, nearly triple national averages for traditional classes (Hake (1998), "Interactive Engagement vs …" Am J. Phys 66,1, 64)

![Normalized gain](image)

We re-present these data in the figure above, plotting a histogram of individual students' normalized gains, averaging over all 4 of our Tutorial courses, and separately, averaging over both of our traditional-recitation courses.

3. Attitudes and Beliefs: CLASS Data:
The above graph shows shifts (Post - Pre) for "overall favorable" CLASS score in Phys 1110. (Scores to the right are pre-tests) Note that, after the first term, all these classes show overall negative shifts. (Sp04 was the only one shown here taught by one of the CO-PIs)
These results are typical for attitude and belief shifts (as demonstrated e.g. by the MPEX, refxx) Note that Fa04 through Fa05 were non-Tutorial courses: it does not appear from these data that the curricular reforms in the recitation plays a significant role on CLASS shifts, in either direction.

B. Physics II:

The BEMA(Brief Electricity and Magnetism Survey) was given pre and post in Phys 1120 (introductory, calculus-based second semester physics) for all terms since we implemented Tutorials in Fa 2004:

1. Conceptual Learning: Learning Assistant data (not all matched, but close)
   Phys 1120, Sp 2007. For 4 LAs, pre average was .71, post average was .80 (g=.31)
   Phys 1120, Fa 2006. For 6 LAs, pre average was .64, post average was .71 (g=.19)
   Phys 1120, Sp 2006. For 5 LAs, pre average was .73, post average was .81 (g=.31)
   Phys 1120, Fa 2005. For 8 LAs, pre average was .66, post average was .76 (g=.31)
   Phys 1120, Sp 2005. For 6 LAs, pre average was .75, post average was .90 (g=.61)
   Phys 1120, Fa 2004. For 8 LAs, pre average was .54, post average was .73 (g=.42)

2. Conceptual Learning: Data for classes as whole:

<table>
<thead>
<tr>
<th>Semester</th>
<th>N (# students)</th>
<th>Pretest (%)</th>
<th>Posttest (%)</th>
<th>Ave normalized gain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(matched only)</td>
<td>~ +/-1%</td>
<td>~ +/- 1-2%</td>
<td>~ +/- .02</td>
</tr>
<tr>
<td>Sp 2007</td>
<td>260</td>
<td>N/A</td>
<td>52.8</td>
<td>N/A</td>
</tr>
<tr>
<td>Fa 2006</td>
<td>351</td>
<td>26.3</td>
<td>55.7</td>
<td>.40</td>
</tr>
<tr>
<td>Sp 2006</td>
<td>196</td>
<td>26.4</td>
<td>53.4</td>
<td>.37</td>
</tr>
<tr>
<td>Fa 2005</td>
<td>325</td>
<td>25.2</td>
<td>49.5</td>
<td>.33</td>
</tr>
<tr>
<td>Sp 2005</td>
<td>232</td>
<td>27.4</td>
<td>58.7</td>
<td>.43</td>
</tr>
<tr>
<td>Fa 2004</td>
<td>319</td>
<td>26.2</td>
<td>58.9</td>
<td>.44</td>
</tr>
</tbody>
</table>

The Table above shows remarkable consistency of pretests from term to term.
Fa04 and Sp05, with the highest gains, were taught by Co-PI's on the grant. Histograms for pre and post data look relatively normally distributed, and are not plotted for all semesters.

In the figure above, we illustrate an interesting faculty development effect. The histogram (barred) data show pre- and post- distributions for the two terms (averaged) taught by co-PI's. In the foreground, plotted as lines, we show Fa05 data for Prof. X (dashed, pre on the left, post is clearly shifted to the right) and again for the same Prof. X, teaching again in Fa06. Prof X is regular faculty, Fa05 was his first experience teaching a large introductory class. The second time he had some support (from a Co-PI) in running Tutorials, but otherwise ran the class himself. His class' average post-BEMA score has improved by 6 points.

3. Attitudes and Beliefs: CLASS Data:

The above graph shows shifts (Post - Pre) for "overall favorable" CLASS score in Phys 1120. (Scores to the right are pre-tests at the start of the term). Note that these shifts are all negative, but smaller than many we observed in Phys 1110 (and not always significantly negative) CO-Pis taught in Fa04 and Sp05.

C. Physics 4810/7810:
We have matched pre-post BEMA scores for the start and end of the terms in which students took our new upper division course on Teaching and Learning Physics. Some of these students were explicitly engaged in LA activities in Physics II as the "practicum" component of the course, others were not 1120 LA's (their teaching activities were in other environments, primarily high school tutoring or mechanics Tutorials)

For N=8 (over 2 semesters) who served as LAs in E&M Tutorial environment:

- BEMA pre ave was 72%, post ave was 85% , gain of ave = .46

For N=13 (over 2 semesters) who did not serve as LA's:

- BEMA pre ave was 70%, post ave was 77% , gain of ave = .24

The above results combine graduate and undergraduates - although our graduate students average considerably higher on the BEMA pretest (77%, compared to 53% for the undergrads), and compose about 2/3 of this sample.

Here we conclude the following: The pretest scores of undergraduates students in this upper division course are typical, comparable to what we see from other upper-division physics majors (see next section). All students engaged in this course post gains on the BEMA, but those directly involved with teaching the particular material being tested by the instrument show significantly higher gains.

D. Upper Division Physics

Associated with this project, we have conducted a longitudinal study to investigate the longer term effects of these introductory reforms on our upper-division majors. Below is a plot of BEMA scores measured for physics majors after completing upper-division Electricity and Magnetism (either I or II, data is combined).

For the 1st three terms of data collection (Fa04-Fa05), none of the students had themselves taken a freshman physics class with Tutorials, since those were just starting up. This is our "control
group”, shown in the first bin in the figure, labeled "No Tutorials". (Total N=75 unique students, average is 54 +/- 2%) In the following three terms, from Sp06-Sp07, some of the students had come from other freshman experiences and had not had Tutorials when they were freshmen (again labeled "No Tutorials". This bin has N=23, average is 57 +/- 4%, please note this is not significantly different from this same category in the earlier bin.) Next we see students labeled "Tutorials", this indicates students in these same classes (Sp06-Sp07 combined, once again), but who had themselves taken their Freshman physics II several terms earlier at CU in a class using Tutorials and Learning Assistants. (This group contains N=33 students, average = 72 +/- 3%) The final bin is for only those students who had been an LA at some point in their career for the Physics II course, this has N=7 unique students, average = 79 +/- 4%) We thus see evidence that the Tutorial experience in freshman year translates to significantly improved BEMA scores at the upper division level, and that being an LA moves scores up to levels comparable to what we measure for incoming graduate Teaching assistants whose average BEMA score (before teaching) is 80 +/- 4% (N=15)
Discipline-Based Educational Research (DBER) Team
Astrophysical & Planetary Science

Creation of a Diagnostic Instrument for Introductory Astronomy Classes

Douglas Duncan

Activities

The LA-Test grant calls for assessment of the learning taking place in transformed undergraduate classes that make use of LAs. In the field of astronomy, there is only one existing Astronomy Diagnostic Test (ADT). It is good but limited in subject matter. Some Astrophysical and Planetary Sciences (APS) faculty members use it, reasoning that it gives feedback on the teaching effectiveness of perhaps 20% of their course. Many faculty members, however, refuse to use it because it does not cover enough of the modern curriculum.

The DBER APS group has been working on developing a more appropriate diagnostic test (conceptual evaluation). The goals in creating a new diagnostic test were therefore multiple:

1. Create an instrument that tests conceptual or deep learning, not just the learning of facts.
2. Create an instrument that covers enough of the curriculum taught in APS, and is amenable to APS faculty, that they will use it.
3. Make it in multiple-choice and format so that it is easy to administer and grade.
4. Complete the work in one semester, so that it could be tested in summertime and used in fall 2007.

The existing ADT has no questions about the solar system. Introductory astronomy in APS is taught in two semesters and the first semester (ASTR1010) is mostly about the solar system. In the past, ASTR1010 has used the largest number of LAs of any astronomy course. Therefore the need was particularly great for a diagnostic that includes questions about the solar system.

Procedure

Good questions from oral interviews

Duncan and learning assistant Mary Beth Sechler started by surveying APS faculty members who teach solar system astronomy as to what topics they thought were most important. Faculty were invited to suggest open-ended questions on these topics. We then gave oral interviews to a number of ASTR 1010 students, asking them the open-ended questions and transcribing both correct and incorrect answers. Common incorrect answers were then used as the distractors when composing multiple choice questions. This procedure has been shown to result in a very effective test (e.g. the Force Concept Inventory in physics, Hestenes et al 1992). It is also true that faculty members usually believe that oral exams probe students’ knowledge deeply, and we hope that they will be inclined to trust the diagnostic potential of questions constructed in this manner.
Unfortunately, creating questions based on oral interviews is very time consuming. We interviewed 15 students, creating 60 pages of transcripts. From these we obtained about a dozen good questions. For a comprehensive diagnostic we needed more questions.

**Good questions from faculty members**

Since we anticipated that we would not have enough time or resources to create a full solar system diagnostic test based solely on student interviews, we also asked faculty members to give us multiple-choice questions they had used in the past that they believed were good at measuring students’ conceptual understanding. We used a fraction of these and combined them with the questions generated from oral interviews to form a Draft Solar System Diagnostic Test.

**Testing and perfecting the draft**

We then gave the Draft Diagnostic Test to 42 students drawn from two different introductory classes. We identified questions that elicited student misconceptions as well as a few (about 15%) that were answered correctly by almost none of the students or by all of the students. The latter were eliminated. We also gave the draft diagnostic to eight students who took the test in our presence and were asked to identify any questions that were unclear. We used this information to eliminate some questions and clarify the wording of others. The result is a diagnostic test of 35 questions covering all aspects of solar system astronomy identified as very important by APS faculty who teach ASTR 1010. We also timed how long it took students to complete the diagnostic test. Most took 30-40 minutes.

Unlike some diagnostic tests that are designed to be given at the start and the end of the semester, this diagnostic test should probably be given only at the end. Few of the topics in solar system astronomy would have been encountered by an introductory student prior to taking the class. The relevant comparison should be between classes which have been transformed and those that have not. We expect to use the diagnostic test in this way in fall 2007 and succeeding semesters.

**Other introductory classes**

Inspired by the results of constructing the solar system Astronomy Diagnostic Test we realized how useful a better diagnostic test would be for the second semester introductory course which covers topics including stars, galaxies, and cosmology. Both semesters of introductory astronomy are supposed to teach scientific or critical thinking, but this is rarely tested. We therefore began construction of the second diagnostic test. This was deliberately done in sections or modules, so that faculty members could choose to use some or all of the modules. The modules are:

- Sky Motions
- Scale of the Universe
- Physics, Light, and Gravity
Stellar Astronomy  
Galactic and Extragalactic Astronomy  
Cosmology  
Scientific Reasoning and Attitudes

Each section contains typically eight questions. The sections on sky motions, and physics light and gravity, are filled with questions that have already been used in the original Astronomy Diagnostic Test were tested in physics, so we view them as complete.

The section on Scientific Reasoning and Attitudes includes four questions that deal with science and pseudoscience, and a number of true false questions taken from the Colorado Learning about Science (CLASS) Survey.

We are at the stage of gathering open ended questions and good multiple-choice questions on the other topics from faculty members. We hope to construct a useful diagnostic test on these topics during fall 2007. When this is done, there will be diagnostic instruments suitable for use in all introductory astronomy classes. (APS offers four different classes.)

We realize that making a diagnostic test modular means that not all faculty members will give the same exact test. It seems more important to convince faculty members who are currently not doing any assessment to start assessing their classes, and we accept that they might only use, say, 2/3 of the modules. This will still provide very useful assessment information to them. Additionally, since many introductory classes are taught each semester, we should still build up assessment data for all of the modules.
Discipline-Based Educational Research (DBER) Team

Applied Mathematics

A Study on Learning Assistant-Supported Oral Exams

Mary Nelson and James Curry

Activities

The DBER Applied Mathematics (APPM) group taught the LA-supported courses listed below and used oral assessments to improve student understanding in Fall APPM 1350 and Spring APPM 1360 classes. We plan to increase the participation of students in Oral examinations in APPM 1350/60 and expand orals to APPM 3310 in Fall 2007. We believe that LA-supported oral exams and calculus workshops have an impact on student learning.

**APPM 1340/45** (Calculus I, two-semesters). Enrollment typically 60-80 each semester. There are three 50 minute lectures a week and and one 50-minute recitation for ~24 students, led by a grad physics TA and 1-2 LAs. 90-100% of students participate in oral assessments before each test and before the final.

**APPM 1350** (Calculus I, one semester). Enrollment typically 600-650 in the fall and 150-200 in the spring. There are three 50 minute lectures a week and one 50 minute recitation (sometimes run by Noyce Scholars who were once learning assistants) for ~25 students, typically led by a grad APPM TA. 20-30% of students participate in oral assessments before each test.

**APPM 1360** (Calculus II, one semester). Enrollment typically 400-500 in the spring and 150-200 in the fall. There are three 50 minute lectures a week and one 50 minute recitation for ~25 students, typically led by a grad APPM TA. 20-30% of students participate in oral assessments before each test.

**GEEN 1350** (Calculus I Workshop). Enrollment is typically 50-60 in the fall and 10-30 in the spring. The students meet for one hour forty minutes with a TA and 1-2 LAs and work in small groups to solve difficult problems and discuss concepts.

**GEEN 1360** (Calculus II Workshop). Enrollment is typically 40-50 in the spring and 10-30 in the fall. The students meet for one hour forty minutes with a TA and 1-2 LAs and work in small groups to solve difficult problems and discuss concepts.

**APPM 3310** (Matrix Methods). Enrollment is typically 60-100 in the fall and 50-70 in the spring.

Current course syllabi can be found at  [http://amath.colorado.edu/courses/](http://amath.colorado.edu/courses/)

NOTE: LAs have been used in another applied mathematics course (Fourier Series) but we did not use LAs in this course in 2006-07.
DBER-Applied Mathematics (continued)

Findings

Course development and transformations have been going on since before the LA-TEST grant, with our first oral assessments implemented in 2003 and our workshops were begun even earlier. Data collection began in 2003. We have observed and documented the following:

- improved student conceptual understanding and grades in:
  - APPM 1340/45 (Calculus I, two-semester course.)
  - APPM 1350 (Calculus I, one-semester course.)

- improved student performance on procedural questions on a common final
  - APPM 1340/45 students

- use of orals in APPM 1360, data not yet analyzed

- often succeed at supporting favorable attitudes and beliefs:
  - evidenced by responses to FCQ questions
  - evidenced by surveys, one-minute papers and enthusiasm of students after participating in orals

- provide lasting impact on retention
  - students who attended orals in APPM 1340/45 showed significantly higher rates of retention at the university
  - APPM 1340/45 students who benefited from orals attended and succeeded at Calculus II at significantly higher rates than their counterparts

A. APPM 1340/45

All incoming freshmen taking Calculus I are given a placement test consisting of algebra and trigonometry questions which has been shown to be an excellent predictor of success in Calculus I. The placement test results for 10 years have been analyzed and it has been shown that there is no difference in the regression equations obtained for each of the 10 years.

Oral assessments were offered to all APPM 1340 (treatment) students, and the mean placement score of those treatment students was shown to be 1 standard deviation lower than that of the control group. In spite of that, when the treatment and control groups were given a common final, the treatment group did significantly better on three measures. The treatment group’s mean final exam grade was significantly better than the control group’s. Since the treatment concentrated on improving conceptual understanding, it was not surprising that the treatment students did significantly better on the conceptual questions on the final exam, but they also did significantly better on the procedural questions.

In the table below, students are defined as “at-risk” if they earned a 17 or less on the placement test. The Table describes the scores of the aggregate treatment and controls groups and the scores of those groups broken down into both at-risk and not-at-risk. The treatment students did considerably better in every category.
Table 1: Comparisons of grades on final exam.

<table>
<thead>
<tr>
<th>Group</th>
<th>Subgroup</th>
<th>Final Exam</th>
<th>Conceptual</th>
<th>Procedural</th>
<th>Placement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>At-risk</td>
<td>93</td>
<td>9.2</td>
<td>42.6</td>
<td>14.15</td>
</tr>
<tr>
<td></td>
<td>Not-at-risk</td>
<td>98.2</td>
<td>8.4</td>
<td>44.7</td>
<td>20.3</td>
</tr>
<tr>
<td></td>
<td>All students</td>
<td>95.4</td>
<td>8.9</td>
<td>43.4</td>
<td>16.5</td>
</tr>
<tr>
<td>Control</td>
<td>At-risk</td>
<td>52</td>
<td>3.25</td>
<td>22</td>
<td>13.99</td>
</tr>
<tr>
<td></td>
<td>Not-at-risk</td>
<td>79</td>
<td>6.7</td>
<td>33.3</td>
<td>23.4</td>
</tr>
<tr>
<td></td>
<td>All students</td>
<td>74</td>
<td>6</td>
<td>31</td>
<td>21</td>
</tr>
</tbody>
</table>

As described above placement scores have been shown to be excellent predictors of course grades for control students at the p = .001 level, that is, the following regression equation is a very good predictor of a control students course grades (see Table 3).

Control: \( G = -1.556 + .164 \times \text{PLACE} \)

Table 2: Calculus I course grades are predicted by placement scores for control students

<table>
<thead>
<tr>
<th>CONTROL</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>244.174</td>
<td>1</td>
<td>244.174</td>
<td>225.382</td>
<td>.000</td>
</tr>
<tr>
<td>Residual</td>
<td>513.521</td>
<td>474</td>
<td>1.083</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>757.695</td>
<td>475</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(a) \( R \text{ Squared} = .568 \)

Table 3 establishes the coefficients for predicting control group course grades, explaining the equation above.

Table 3: Regression coefficients: Course grades predicted by placement scores

<table>
<thead>
<tr>
<th>CONTROL</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Constant)</td>
<td>-1.556</td>
<td>.239</td>
<td>-6.500</td>
<td>.000</td>
</tr>
<tr>
<td>place score</td>
<td>.164</td>
<td>.011</td>
<td>.568</td>
<td>.000</td>
</tr>
</tbody>
</table>

Table 4 shows that for treatment students, placement scores were not predictive.

Table 4: Course grades are not predicted by placement scores for treatment students.

<table>
<thead>
<tr>
<th>TREATMENT</th>
<th>Sum of Square</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>.140</td>
<td>1</td>
<td>.140</td>
<td>.075</td>
<td>.786</td>
</tr>
<tr>
<td>Residual</td>
<td>48.517</td>
<td>26</td>
<td>1.866</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>48.657</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The regression equation would be: \( G = 2.623 - .018 \times \text{PLACE} \) for treatment students. The regression equation for the control group demonstrates that the coefficient of student’s placement score is .164 having a significant effect on their predicted grade, whereas for the treatment students,
the coefficient is only -.018. The regression analysis in Table 4 demonstrates that placement scores are not predictive for treatment students.

Table 5 shows the placement score and the grade that the placement test predicted for each treatment student compared to grade they actually earned.

Table 5: Treatment students’ grades: predicted by the regression versus actual grades

<table>
<thead>
<tr>
<th>Student Number</th>
<th>Placement Score</th>
<th>Predicted Final Grade</th>
<th>Actual Final Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19</td>
<td>1.6</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>1.9</td>
<td>2.7</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>.8</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>1.6</td>
<td>3.0</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>.5</td>
<td>2.3</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>1.6</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td>.5</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>19</td>
<td>1.6</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>14</td>
<td>.8</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>23</td>
<td>2.2</td>
<td>.7</td>
</tr>
<tr>
<td>12</td>
<td>20</td>
<td>1.75</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>7</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>16</td>
<td>1.1</td>
<td>3.7</td>
</tr>
<tr>
<td>15</td>
<td>18</td>
<td>1.4</td>
<td>3.7</td>
</tr>
<tr>
<td>16</td>
<td>21</td>
<td>1.9</td>
<td>3</td>
</tr>
<tr>
<td>17</td>
<td>15</td>
<td>1</td>
<td>3.3</td>
</tr>
<tr>
<td>18</td>
<td>24</td>
<td>2.4</td>
<td>3.3</td>
</tr>
<tr>
<td>19</td>
<td>15</td>
<td>1</td>
<td>2.3</td>
</tr>
<tr>
<td>20</td>
<td>11</td>
<td>.3</td>
<td>3</td>
</tr>
<tr>
<td>21</td>
<td>15</td>
<td>1</td>
<td>3.3</td>
</tr>
<tr>
<td>22</td>
<td>16</td>
<td>1.1</td>
<td>4</td>
</tr>
<tr>
<td>23</td>
<td>16</td>
<td>1.1</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>17</td>
<td>1.3</td>
<td>3.3</td>
</tr>
<tr>
<td>25</td>
<td>12</td>
<td>.5</td>
<td>2.3</td>
</tr>
<tr>
<td>26</td>
<td>18</td>
<td>1.4</td>
<td>4</td>
</tr>
<tr>
<td>27</td>
<td>21</td>
<td>1.9</td>
<td>3.3</td>
</tr>
<tr>
<td>28</td>
<td>17</td>
<td>1.3</td>
<td>3</td>
</tr>
<tr>
<td>29</td>
<td>17</td>
<td>1.3</td>
<td>0</td>
</tr>
</tbody>
</table>

Yellow highlighting in the figure above indicates students who were predicted to “fail” (D, F or W) by the regression line but who earned a grade of C or better. Turquoise highlighting indicates a student predicted to pass who does better than the prediction, and green highlighting is a student predicted to fail who earns a marginally higher grade.
Below are Scatter plots for the control and treatment groups. Figure one demonstrates that increases in control students’ placement scores correlates to increased final exam scores in that group. Figure two shows that increases in treatment students’ placement scores does not result in a consistent increase in exam scores. It appears that oral assessments were able to overcome the negative effect of poor placement scores, i.e. poor mathematical preparation.

**Figure 1.** A scatter plot of final exam scores of students in the control classes.

![Scatter plot of final exam scores of students in the control classes.](image)

Table 7 shows the effect size of the difference in final exam scores. An effect size of 1.5 standard deviations is quite unusual in social science. The fact that the students in the treatment group with the lowest placement scores (<18/30, defined as at-risk of failing the course) did more than a standard deviation better than even the control students who were not-at-risk, a substantial difference.
Figure 2. A scatter plot of final exam scores of students in the treatment class reveals no regression line.

Table 7: Effect size comparisons

<table>
<thead>
<tr>
<th>Comparison groups</th>
<th>Mean Exam Score</th>
<th>Standard deviation</th>
<th>Mean difference</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment At-risk</td>
<td>93.13</td>
<td>27.414</td>
<td>40.9</td>
<td>1.49</td>
</tr>
<tr>
<td>Control At-risk</td>
<td>52.23</td>
<td>24.411</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment Not-at-risk</td>
<td>98.20</td>
<td>25.302</td>
<td>20.59</td>
<td>.77</td>
</tr>
<tr>
<td>Control Not-at-risk</td>
<td>77.61</td>
<td>26.691</td>
<td></td>
<td>.57</td>
</tr>
<tr>
<td>Treatment At-risk</td>
<td>93.13</td>
<td>27.414</td>
<td>15.52</td>
<td>.57</td>
</tr>
<tr>
<td>Control Not-at-risk</td>
<td>77.61</td>
<td>26.691</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There could be some bias in the final exam scores, if the treatment instructor “tutored” the treatment students for the final exam. To eliminate subjectivity, we looked at more objective measures, the students’ retention rates at the university and their success rates in Calculus II. We found that while 30% of the treatment at-risk left the university, 45% of the control at-risk left.

None of the students in Calculus II had orals available. They were all integrated into the same classes without orals. Table 9 shows that while only 16% of the control at-risk students passed Calculus II, 50% of the treatment at-risk students were able to do so. This indicates that the treatment students did far better after the treatment than the control students.
Table 9. Performance of at-risk on final and enrollment in Calculus II

<table>
<thead>
<tr>
<th>GROUP</th>
<th>Number of at-risk students taking the final</th>
<th>Mean course grade</th>
<th>Standard deviation</th>
<th>% at-risk who took Calculus II (excluding dropped students)</th>
<th>% of the at-risk students who passed Calculus II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>16</td>
<td>2.34 (C+)</td>
<td>1.30</td>
<td>56%</td>
<td>50%</td>
</tr>
<tr>
<td>Control</td>
<td>61</td>
<td>.79 (D-)</td>
<td>1.06</td>
<td>20%</td>
<td>16%</td>
</tr>
</tbody>
</table>

Calculus I (APPM 1350):

Analysis of data is ongoing for the other treatment classes. Preliminary Answer Tree results show that on the first exam, placement scores were the best predictor of test 1 grades. Then by placement score, the next best predictor was whether the students participated in orals or not. Within individual placement scores, students averaged 6-12% better on the first test if they participated in orals. Analysis of data for the other two semester tests has yet to be analyzed.

Calculus II (APPM 1360): Appropriate data has been collected but has not yet been analyzed.

Learning Assistants: We have 4 of our former learning assistants who student taught in the spring semester and who all have teaching positions in the fall. Three will be teaching in Colorado and we will follow their progress in the new year. Two learning assistants will student teach this year and we will follow them in their first year of teaching.

We have not yet completed the content test for Calculus I. We hope to bring it on line in the fall for students and learning assistants and will also give both learning assistants and students the CLASS in the coming year.

We have become more adept at data collection over the past year, and hope to provide more elaborate analyses after the 2007-08 academic year.
Discipline-Based Educational Research (DBER) Team
Molecular, Cellular, and Developmental Biology

Developing an instrument to measure conceptual understanding of core concepts in biology

Mike Klymkowsky, Kathy Garvin-Doxas, and Will Leary

Activities

The DBER Molecular, Cellular, and Developmental Biology (MCDB) group has been focusing on developing a conceptual instrument to measure students’ conceptual understanding of fundamental concepts in biology, the Biology Concept Inventory (BCI).

Three learning assistants (LAs) were involved with the Spring 2007 version of MCDB 1111: Biofundamentals, a 4 credit lecture/virtual laboratory course for majors in MCDB described as part of an Point of View article for Life Science Education (see Klymkowsky, 2007). The course minimizes formal lecture, and each class period is organized around analyzing student responses to questions, embedded within the web-based course materials and the tutorial exercises assigned at the end of the preceding class period. Interactive response systems (clickers) are used as a means of polling student opinions, driving discussion, and determining which ideas need further discussion. The course began with the administration of the Biology Concept Inventory (BCI), a 30 question instrument designed to examine student understanding of key areas of basic biology. This version of the BCI was released in the Fall of 2006, and this was the first time it was given to students in MCDB 1111.

LAs ran weekly sessions with enrolled students centered around constructing concept maps of previous weeks materials and in working with tutorials assigned after each class period. Students were required to upload their answers to these tutorials through the CU Learn course website before the next class. The tutorials, which are still in the early stages of development, evaluation, and revision, are modeled after similar exercises in physics. They are designed to encourage students into metacognitive reflection as part of approaching complex questions.

We are also developing a set of inquiry-based Tutorials for this course. These tutorials will be modeled after the University of Washington Tutorials which have proven to be successful in physics. In the upcoming year, we hope to continue with the development of the BCI and to examine ways to effectively lead students to useful understandings of concepts such as emergent behavior associated with random processes and interconnected networks. We also need to track how interventions introduced into one class persist or erode in following classes depending upon whether effective pedagogical practices are used in subsequent classes. More discussion of the BCI can be found in the Findings section of our year 1 report.

For more information about this course see: http://www.colorado.edu/MCDB/MCDB1111/
Findings

Through the development of the BCI, we have found that we are in the position of being able to recognize areas/concepts that are problematic for students. During year 1, we identified these areas and are beginning to develop tutorials to address these difficulties. Details are described below.

Both through the LAs analysis of student thought processes, and student answers to tutorial questions, we have been able to identify a number areas/concepts that are problematic for students. Some of these were unexpected and new to us (i.e. not previous identified either during previous versions of the course or through the process of question development for the BCI). For example, Biofundamentals deals extensively with evolutionary mechanisms. In the process of asking whether the genetic code is a homologous or an analogous trait of organisms, we discovered student confusion between the genetic code as the algorithm by which genetic information encoded in DNA is read and the genetic information (the genotype) held within the DNA. We then confirmed and extended this observation through student responses to questions administered through our Ed's Tools site. We are currently working to devise tutorial and LA proctored exercises aimed at bring students to a better understanding of the terms "algorithm" and "genetic code", and we expect that questions on these topics will become part of future versions of the BCI.

Research associated with the development of BCI lead to the revelation that the majority of students appear to believe that molecules bind to one another much like jig-saw puzzle pieces. It is our working hypothesis that such a "jig-saw puzzle" viewpoint, which is likely to be actively instilled into students, leads to destructive interference with subsequent understanding that the evolutionary processes by small changes of structure can lead to functional changes in proteins - a hypothesis described in detail in Klymkowsky (2007b). A similar issue, which also impacts
student understanding of evolutionary processes, involves the specificity of catalytic reactions. Analysis of students in MCDB 1111 using the BCI indicates that the percentage of students picking the correct response, namely that correct binding is associated with the lower total energy, was not changed by instruction (29% pre- and post-).

To test and extend our hypothesis Will Leary, a Noyce Fellow, carried out interviews with student engaged in a simple tutorial that asked students to rank interactions between molecules. We are currently in the process of analyzing these interviews, and working to devise an extended tutorial exercise that directly addresses this point. We plan to incorporate LAs in the administration, analysis, and revision of this "Molecular Interaction and Specificity" tutorial during the 2008 version of MCDB 1111.

At the same time, previous work and current tutorials used in MCDB 1111 appear to have had some impact on student understanding of the role of random processes, such as diffusion, in biological systems. During the development of the BCI, we determined that student understanding of random processes, such as diffusion and genetic drift, is relatively weak. More to the point, students have a conception (again probably inculcated by past instruction) of the perfection and efficiency of biological processes and system that leads to conceptual discord when considering the random processes, assumed to be inefficient, that underlie many biological events. In this context, however, instructional methods in MCDB 1111 do seem to have had some effect, leading to an increase in student's ability to recognize the similarity between genetic drift and diffusion, from pre- to post-instruction and compared to students entering a second semester genetic course.

While the change from 30% to 54% is encouraging, there is still clearly room for further improvement and a higher resolution analysis of student understanding of the role of randomness in biological systems. The importance of this concept is further underlined by a recent paper by Lynch (2007 PNAS 104:8597-8604) which points out the fact that not all working scientists
appreciate the power of purely random processes as drivers of evolutionary change. We are currently considering how to adapt the virtual lab that deals with diffusion to include the question of genetic drift and a deeper understanding of the power and limits of random processes. As part of this project, we plan to explore further student understanding of related issue through LA reports, and coordination with the alternative introductory courses in MCDB, MCDB 1150 (Introduction to Molecular Biology) and its associated laboratory course (MCDB 1151). This is the first year that this much large course, typically offered in the Fall Semester, will employ LAs. This will also give us an opportunity to directly compare student BCI responses between the MCDB 1111 and MCDB 1150. Furthermore, we will examine BCI responses from three courses that follow our introductory course, i.e. MCDB 2150 (Genetics), MCDB 3120 (Cell Biology), and MCDB 3500 (Molecular Biology) to determine the extent to time and subsequent instruction had influenced student ideas.

There are a number of challenges associated with the biology aspect of our project. The first is that lack of a commonly accepted core introductory curriculum, analogous to that in physics. For example, many biology projects (e.g. both MCDB and Integrated Physiology), do not formal teach evolutionary processes in any required course; evolutionary processes are an integral component of MCDB 1111. This makes the development and acceptance of assessment instruments more difficult that has been the case in physics, where this is a focus on Newtonian mechanics (first semester) and Electricity and Magnetism (second semester). In terms of instruction, the most difficult challenge is to uncover the persistence or preexisting, or the instructor-initiated generation of counterproductive student premises; unless recognized by students (and instructors) these actively interfere with an attainment of a robust and facile working understanding of the subject matter. This is associated with the challenges associated with devising effective interventions or environments through which students can recognize and revise their approaches.

As an illustration of persistent problem, uncovered by through the use of the BCI, consider the response to the following question.

If two parents display distinct forms of a trait and all of their offspring (of which there are hundreds) display a new form of the trait, you would be justified in concluding that…

- a. both parents were heterozygous.
- b. both parents were homozygous.
- c. one parent was heterozygous, the other was homozygous.
- d. a recombination event has occurred in one or both parents.

What is interesting about student responses to this question is that they reveal an unexpected inability to transfer the simple rules of Mendelian genetics from a conventional to a (slightly) unconventional context. A non-majors genetics course produce no significant difference in the frequency of the correct response, while a major's course produced only a small increase, together with a dramatic increase in an irrelevant, but technical sounding alternative (d). While these observations need to be replicated in other contexts, they suggest a wider problem with student's ability to apply basic genetic understanding to new scenarios. How this skill, which is
central confident analysis of novel scenarios, can be cultivated needs to be explored more rigorously.
Discipline-Based Educational Research (DBER) Team
Molecular, Cellular, and Developmental Biology

Conceptual Understanding and Attitudes
Michelle Smith, Jia Shi, and Jenny Knight

Activities

LA-supported courses
MCDB 1041 (Fall 2006). Enrollment 80. Two 50-minute lectures per week using peer instruction (clickers and in-class discussions) and one 50-minute interactive activity session per week (the class is divided into four sections of ~20 students), each session led by one of four LAs. Instructor: Jenny Knight

MCDB 2150 (Fall 2006). Enrollment 150. Three 50-minute lectures a week using peer instruction (clickers). Five separate “activity” sessions were also held during the course, in which students worked with one of two LAs on projects or presentations. Instructor: Sylvia Fromherz

MCDB 2150 (Spring 2007). Enrollment 400. Three 50-minute lectures per week using peer instruction (clickers). Two voluntary study groups (~25 students each) were led by LAs (two out of three LAs present at each session) each week. In these sessions, students used peer learning to work on problem set questions, with help from LAs, who used a Socratic method style of questioning. LAs also worked with the instructors to help rewrite problem set questions from previous semester so that the question had a human disease focus. Instructors: Mark Winey, Nancy Guild, Michelle Smith

Content mastery
Development of assessment tools
We are in the process of developing both Introductory and Genetics conceptual pre-post multiple choice assessment tools. The initial development process included identifying the learning goals for each course (in consultation with the instructors who teach the courses). We encouraged the creation of learning goals that included higher level reasoning skills (Allen and Tanner, 2002) rather than memorization alone, and then wrote questions for the assessment that address these goals, and also addressed student misconceptions that instructors were already aware of. The questions were then modified several times after receiving feedback from many faculty members in the department before being administered (respectively) to the Fall 2006 MCDB 1150 students and the Spring 2007 MCDB 2150 students, both pre and post (learning gain data shown below). Since administering the assessment tools, both have been considerably modified, and will be used again this upcoming year. The tools have been modified to include questions on topics that students clearly struggle with (from student interviews, clicker questions, exams and homework), and then validated through multiple rounds of student interviews (12-20
students interviewed for each assessment tool). Student feedback from the interview process has led to new versions of these assessment tools that use natural language instead of scientific jargon so it is easily understood by the students, use distractors derived from students’ misconceptions, are not based on memorization of facts, and are universally understood by the students (the students know what the question is asking even if they don’t know the answer). We anticipate further small changes in these tools over the years, but are ready to use them to help measure the effects on content learning gain as a result of using interactive engagement techniques and LAs in these courses. The genetics tool is also being administered at other institutions this year.

Attitude and beliefs

We have adapted the CLASS attitudinal tool developed in Physics so that it contains questions relevant for biology. The biology CLASS substitutes the word “biology” for “physics” where necessary, and includes several other small but relevant changes to the statements. This year, we will validate the Biology CLASS tool by interviewing students on the version we have been using, and modifying the tool as is necessary. We have already collected the views of biology experts who span multiple subdisciplines in biology including molecular, ecology, and physiology (30 faculty at a variety of institutions including community colleges, liberal arts colleges, and universities) on the current version of the Biology CLASS. The expert opinions have already allowed us to target a handful of questions as problematic (experts themselves do not agree), and have also allowed us to compare student and LA beliefs to those of experts (see Findings). Because the tool is still in development, we have only snapshots of our students’ and LA’s attitudes, rather than pre-post data that allows us to measure change in attitude. In the future, after we have a validated Biology CLASS, we will collect systematic pre and post data.

Findings

Course development and transformations on the two introductory courses in the major (Introduction to Molecular and Cellular Biology MCDB 1150 and Genetics MCDB 2150) began this past year. Work on Genetics for non-majors (MCDB 1041) has been ongoing for four years, but has not included systematic data collection. This past year, only MCDB 1041 and 2150 used LAs, but in the upcoming years, all three courses will do so. An upper division course, Developmental Biology MCDB 4650, has also undergone significant transformations, and uses LAs, but the LA-Test program does not fund these LAs. We include some data here from this course, since it has used LAs for several years.

Content mastery

MCDB 1150

Our initial pre/post assessment has given us insight into what students understand and what they learn during our Introduction to Molecular and Cellular Biology class. Performance on the introductory biology assessment tool is shown below. On average, students make substantial gains over the semester, but some students appear to make little or no gain. We found that students who enter the class with a high pretest score either were not challenged by the class, or did not make any effort on the post test, and thus ended up with low or no learning gain. These realizations encouraged us to carefully analyze our assessment tool, which we did in part through student interviews. We have found that some topics are easy, while some remain
difficult for students despite considerable instruction on the topic. Using results from this first administration of the assessment, as well as student interview responses, we have been rewriting the assessment questions, and will use the new version of this assessment in the fall. Questions that had high pretest scores or were confusing to students were dropped or substantially rewritten.

The results of this assessment will, in part, help to drive the way the course is taught next semester. Topics that remain challenging to students will be given priority. Activities and/or conceptually rich homework problems will be designed on these topics, and peer learning guided with LAs will allow us to measure whether specifically addressing the challenges will help students to better make sense of the topics.

**Content Learning Gains**

*in MCDB 1150 (Intro)*

**Fall ‘06**

<table>
<thead>
<tr>
<th>Student normalized learning gain</th>
<th>30.4% (SE +/- 2.79)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>243</td>
</tr>
</tbody>
</table>

Pre (avg): 57%
SE +/- .95

Post (avg): 72%
SE +/- 1.06

MCDB 2150

This class used three LAs for the first time. They worked with a subset of the class (45 students) facilitating small group work on problem sets. Student performance on the Genetics assessment tool is shown below. Overall, we can show that students made substantial learning gains (67.6%), but the LAs make even more impressive learning gains, surpassing those of both the students and graduate TAs. In addition, the LAs for this course had an overall higher average score on the post test than did the graduate TAs.
Effect of LA-led small group work on student performance in MCDB 2150

To begin to measure the effectiveness of the peer study groups led by the LAs on student performance, the exam scores of students that participated in the peer study groups were compared to three other groups of students in the MCDB 2150 genetics course. The students in these groups began the course with similar content knowledge as the students that participate in the study groups, as measured by their pre-assessment scores, but showed varying levels of participation in extra help sessions offered through the course. The first group of students (n=35) either did not participate in any extra help sessions offered through the course or rarely came to a drop-in help room (fewer than 5 times during the semester). The second group (n=28) came to a drop-in help room between 6 and 10 times during the course of the semester and the third group (n=39) came to the help room 11 or more times during the course of the semester. The last group (n=37) is comprised of students who participated in at least half of the 11 total peer study group sessions.

The course average on the first exam was 78.4% and all four groups performed close to the course average. At this point in the semester, the study groups had met three times, 48.3% of the students who reported coming to the help room between 6-10 times signed in on the help room attendance log (average number of visits =1.9), and 82.1% of the students who reported coming to the help room more than 11 times signed in on the help room attendance log (average number of visits =2.2). Therefore, at this point in the semester, participation in the help room or in peer study groups does not seem to affect student grades.
However, students that participated in study groups facilitated by LA’s outperformed the other groups on exams 2, 3, and the comprehensive final exam. On all three exams, the performance of the students in the peer study groups was better than the other groups and statistically significantly better (p<0.05) than the students who began the course with similar content knowledge but either never or rarely came to the help room (fewer than 5 times).

When we asked the students at the end of the semester how useful for your learning were the LA led study groups an overwhelming number said “a great deal.”

Positive feeling about the usefulness of the study groups was also reflected in specific comments from the students:

“I wholeheartedly believe that the learning group has helped me in the course. Not only have I been able to find new approaches to problems, I find myself more comfortable discussing science with others.”

“I find it intriguing to hear how other people are interpreting the course material and it
helps me feel like I am not alone in my confusion and thus keeps me from getting frustrated.”

Impact on LAs themselves
The LA’s also submitted weekly reflections about their experience with the study groups. These comments reflected the growth in thinking about learning and an increase in confidence that occurred during the course of the semester. After the first study group meeting, one LA who is studying to become a high school teacher commented:

“At first I was really nervous because I felt like they might not take me seriously enough, but they were very nice and I just wonder what they thought of me?”

Later in the semester this same LA reflected on the teaching techniques s/he was using:

“I would ask them to work on the problems with each other first and then ask me if they had any questions. This seemed to work a lot better than just asking them what answer they got, and then going into working on the problems. I would also ask them what they thought the answer was, then ask them to explain why they thought that. I really liked this because it made them think more deeply about the question, and their answer.”

Finally at the end of the semester after the students in the study groups had a chance to evaluate the LA’s, this LA commented:

“They really like it when I ask them to explain the problem to another student, which is good to know because I was not sure!!”

Because of the positive impact of the peer study groups on students and LA’s, in the fall 2007 semester, one lecture per week will be modified so they students are working together in groups to solve genetics problems. LA’s will help facilitate group interaction and use the Socratic method to stimulate discussions.

Attitudes and Beliefs
CLASS Data
Since we have not yet begun systematically using the Biology CLASS, we do not have a measure of change in attitudes and beliefs from the beginning to the end of our courses. We do, however, have “post” CLASS data (given at the culmination of the semester) from several classes demonstrating that LAs achieve a much more expert level of attitudes and beliefs than do the students in these classes. The graph below shows data from students in two different classes, 2150 and 4650 (an upper division biology course).
Using categories established from the Chemistry and Physics CLASS (see graph), the 7 LAs from three different classes (MCDB 1111, 2150 and 4650) are in agreement with experts much more often than the students in those classes. This trend is also visible in the overall average % agreement with experts among the students and LAs (2150: 58%; 4650: 65%; LAs: 83.4%). It is also interesting to note that the students in the upper division class (4650) are more expert-like in their beliefs than those in 2150. Work in physics has shown a similar trend and attributes this trend to student selection in the upper division courses rather than a change in attitude (personal communication, K. Perkins).

In the upcoming year, we will gather pre and post CLASS data in all courses using LAs, as well as gathering such data from other MCDB courses in which LAs are not used. We can also follow students who took MCDB 1111, 1150 and 2150 this past year, and whom we have now hired as LAs in these courses to see how their attitudes and beliefs about biology change.
**Discipline-Based Educational Research (DBER) Team – Chemistry**  
**Annual Report**  
**Academic Year 2006-2007**

**Laurie Langdon**

**Acknowledgments:** Dr. Margaret Asirvatham, Professor Veronica Bierbaum, and Professor Robert Parson taught CHEM 1111 and CHEM 1131 this past year. They all contributed to planning for the use of LA’s and in developing new conceptual-based homework activities. They also participated in the construction and revision of concept survey instruments, as did the rest of the General Chemistry Working Group. Dr. Tom Pentecost contributed to the CHEM 1111 learning gain analysis as well as characterizing students who joined LA-led learning groups in Spring 2007. Michelle Stachurski (a Noyce fellow) compiled attendance data for Fall 2006 LA help sessions.

**Activities**

This was the first year that LA’s were used in Chemistry courses at CU. The Science Education Initiative (SEI) provided funding. Learning Assistants were used mainly to assist students with new homework assignments (Chemistry Concept Challenges) that emphasized real-world connections, conceptual understanding, and development of visualization skills (visualizing matter at the particulate level). Summaries of their use are provided below.

**Fall 2006: General Chemistry I (CHEM 1111)**

*Course description:*
Enrollment typically 750 – 800 in the Fall; three lecture sections; team-taught by two faculty members; 20 – 24 graduate Teaching Assistants lead 40 – 50 recitation and lab sections  
Three 50-minute lectures (using clickers); one 50-minute recitation; one 3-hour lab per week

*Use of LA’s:*
Six LA’s  
LA’s attended lecture; worked with students during clicker questions.  
LA’s hosted 20 hours (total) of drop-in help room hours (for help with Concept Challenges)  
• By end of semester, ~120 students per week attended LA sessions

**Spring 2007: General Chemistry II (CHEM 1131)**

*Course description:*
Enrollment typically 450 – 550 in the Spring; two lecture sections; taught by one faculty member; 12 – 16 graduate Teaching Assistants lead recitation and lab sections  
Three 50-minute lectures (using clickers); one 50-minute recitation; one 3-hour lab per week

*Use of LA’s:*
Ten LA’s  
LA’s attended lecture; worked with students during clicker questions; assigned to a “zone”
LA’s hosted 18 (total) “learning groups” each week
  - 150 (out of 500) students chose to join a learning group at the beginning of semester
LA’s hosted 12 hours (total) of drop-in help room hours
  - 30 – 40 students per week went to LA help room

**Commentary on evolving LA structure in Chemistry courses:**
Fall 2006 was our first attempt to incorporate LA’s into General Chemistry. CLASS survey data from previous semesters indicated that students’ attitudes and beliefs about chemistry and chemistry learning decline over the course of a semester. The General Chemistry instructors were concerned about students’ views that chemistry is not connected to the real world, that it consists of disconnected bits of information, and that it is more about memorization than conceptual understanding. In addition, results from a pilot concept survey given at the end of Spring 2006 and again pre/post in Summer 2006 indicated that students struggled with translating chemical formulas into molecular-level representations, among other key chemistry concepts.

**CHEM 1111 (Fall 2006):** The creation of new conceptual homework assignments and the use of LA’s went hand-in-hand in planning for this course. It was desirable for LA’s to contribute in their own way, unique from what TA’s were already doing (primarily helping students with computerized homework and questions on labs). This led to a somewhat compartmentalized structure in which the LA’s role was to assist students with the new conceptual homework. Drop-in “help-room” hours was our first attempt at this. Over time, a few learning communities started to form, as small groups of the same students would come to the same LA session each week. Overall, though, this structure seemed to limit the LA’s abilities to apply some of what they were learning in their weekly education seminar and in our weekly training meeting. For instance, it was difficult for LA’s to facilitate student group work in this drop-in setting. Students came in needing help in different parts of the activity, or not having even looked at the assignment and expecting answers from LA’s. On busy days, it was difficult for LA’s to “stand their ground” by encouraging students to work together and to guide student thinking. In addition, most of the LA’s felt that they were not able to help students very much in class during clicker questions. One perceived problem was the lack of time to discuss student reasoning; if LA’s asked a lot of questions and tried to get the students to explain their reasoning, they ran out of time and then felt like they needed to just give students the answers. LA’s also felt very outnumbered in class; in one section there was one LA for 180 students; in another, there were three LA’s for 350 students.

**CHEM 1131 (Spring 2007):** In consultation with the CHEM 1111 LA’s, and based on the researcher’s experience with the Peer-Led-Team-Learning model at the University of New Hampshire, a modified structure was implemented in Spring 2007. We wanted to impact a larger percentage of students in the course and at the same time improve the LA’s experience by giving them more opportunities to try out what they are learning in their Education seminar.
  - We were able to hire more LA’s (for fewer students in the course). We now had one LA per 50 students, instead of the 1:120 ratio in the Fall.
  - LA’s still attended lecture (which was still deemed as the least effective use of their time), but we attempted to improve this aspect by assigning each LA to a specific zone within the large lecture hall. We posted documents with each LA’s photo and their
assigned zone within the course CULearn website so that students who wanted to interact with an LA during class could find them and sit by them. In general, LA’s still viewed this as the least effective use of their time. They still felt rushed when working with students and often resorted to giving them answers when time was up so students wouldn’t be frustrated.

• The biggest change was implementing LA-led learning groups. CHEM 1131 students were invited to join a learning group at the beginning of the semester. Each LA led one or two of these groups per week; 5 – 12 students were assigned to each group. The groups met for 12 weeks during the semester, and the main goal was to work through the Chemistry Concept Challenges. About 150 (of 500) students opted to join a learning group.

  o In establishing this learning group model, we were explicit about the expectations and potential problems with the students, the LA’s, and the TA’s in the course (since TA’s were still grading the Concept Challenges). Students who opted to join a learning group were expected to attend the majority of their group sessions; otherwise, it would be difficult to establish effective learning communities. Students were provided with a slight grade incentive for attending groups, which mainly worked as a “buffer” in their Chemistry Concept Challenge grade. We anticipated some problems arising from students completing the Concept Challenges in their learning groups and then NOT receiving full credit from TA’s who were grading them. Again, we were explicit with students, LA’s, and TA’s about the fact that even though students may think they “get it” and can explain their reasoning verbally within this group setting, that students are still responsible for communicating their understanding, in writing, to their TAs. This grade buffer (up to 20 points if students attended at least 9 or 12 learning group sessions) seemed to neutralize this potential problem. In fact, if this was ever an issue, it was never brought up by TA’s or LA’s during the semester.

  o The main focus of the learning groups was to work through that week’s Concept Challenge. LA’s were encouraged to lead the sessions so that students were working together, either in small groups or as a whole group, but in a way that students were at the center of the activity, not the LA. (Here’s where LAs had freedom in structuring the activities and trying different approaches based on their particular group of students.) Because we had ten LA’s (close in size to most learning groups), our weekly 1.5-hour meeting was a great place to model different ways of structuring the group activities, such as working through an activity using a jigsaw structure, or by doing some initial whole-group brainstorming. Early on, it was evident that just having the LAs engage in the Concept Challenge in a certain way wasn’t enough; we needed to explicitly reflect on how the activity was structured, and why, once we had worked through it. We reflected on how certain structures (such as a jigsaw) forced each student to talk and contribute and also moved the focus off of LAs as the authority. On two occasions, one of the LAs did not participate in the activity but instead served as an observer of the process. These discussions and observations allowed us to focus on both the content of the Concept Challenges and on approaches to structuring learning groups.
Students who did not join a learning group could still receive help from LAs by attending drop-in help room hours. A number of students took advantage of this opportunity, but fewer than in Fall 2006.

**Plans for Fall 2007:**

**CHEM 1131:** Our model keeps evolving into one that affects a greater number of students. We’re moving toward more of a “physics tutorial” model next semester in Gen Chem II. We’re integrating five LA’s with six TA’s in recitation. The Chemistry Concept Challenge materials are being integrated into recitation, and we’ll have joint weekly training sessions with TA’s and LA’s on the material. We will be able to impact all 240 students enrolled in the class. The smaller enrollment allows us to pilot this model before trying to scale up to either CHEM 1111 or CHEM 1131 in the Spring.

**CHEM 3311 (Organic Chemistry I):** One returning LA and one volunteer LA (returning from last semester) will work with an instructor in this course. They will attend class and hold hybrid “learning groups.” The first part of each learning group session will be structured by materials developed by the instructor and LA’s, and the section half will be directed by students’ questions about the course material.

**Results**

This was the first year that LA’s were used in Chemistry courses. They were used in conjunction with new materials focused on developing students’ conceptual understanding of chemistry. The results section combines qualitative data with some measures of student learning and attitudes/beliefs. Below is a short description of the status of data sources:

- **General Chemistry I Concept Survey:** Used to measure learning gains in CHEM 1111. This multiple-choice instrument was constructed prior to the researcher’s arrival at CU from two literature sources (Mulford, 1996 and Birks, 2003). It was given post-Spring 2006 (no learning gains were calculated) and pre/post Summer 2006. It was revised for use in Fall 2006, the first semester that LA’s were used in the course. There really are no baseline data to be able to determine what constitutes a “high” learning gain. It is still under development. However, we use it to report learning gains from Fall 2006. (CHEM 1111 students completed the survey pre/post; LA’s and TA’s only completed the post survey.)

- **General Chemistry II Concept Survey:** Used in CHEM 1131. Unlike first semester General Chemistry, no one in the chemical education research community has completed a concept survey for use in second semester General Chemistry. We constructed a 19-item survey focused only on acid/base concepts that was used at the beginning of CHEM 1131 in Spring 2007. Throughout the semester, we used clicker questions in class to gather data about students’ incoming ideas about other topics related to the course (solubility equilibria, redox reactions, electrochemistry, and kinetics). We then constructed a 21-item survey that addressed all of these topics for use at the end of the
semester. Thus, there is only a set of acid/base questions that was used in both administrations. We are in process of analyzing those results to approximate learning gains in the course. (CHEM 1131 students, LA’s, and TA’s completed the Acid/Base Concept Survey and the General Chemistry II Concept Survey.)

- **CLASS (Colorado Learning Attitudes and Science Survey):** Used in CHEM 1111 and 1131. Unlike the concept surveys, we do have some historical data from CLASS, which measures students’ attitudes and beliefs about chemistry and learning chemistry. Prior results prompted instructors to implement the use of LA’s and new homework in the course.

- **Extra CLASS questions, grades, and other data:** Once students complete the 50-item CLASS survey, they are asked additional questions related to their course experience. These deal with students’ interest in the course, the extent to which they think various course components were useful to their learning, as well as free response sections to justify their numerical choices. Some of these data are used, especially in determining student satisfaction with joining (or not joining) a learning group in CHEM 1131. We also use course grades to determine the impact of LA’s on those students who took advantage of this resource.

Because our particular LA structure did not allow every student to engage with an LA in the course, we can compare learning gains, attitudes, and course grades for students who interacted with LA’s to those who did not. At this time, several gross comparisons have been made and are reported here. However, there may be other ways to tease out data related to the questions of who most benefited from the use of LA’s by examining subgroups within the population. These analyses are ongoing.

**CHEM 1111 (Fall 2006)**

*Attendance at LA help sessions*
Six LA’s were available to assist ~720 students with the Chemistry Concept Challenges. LA’s hosted a total of 20 hours of drop-in help room hours. Attendance was tracked for nine weeks, starting in the fourth week. Some students attended more than one session per week. (The highest number of sessions attended was 17 during this nine-week period.)

<table>
<thead>
<tr>
<th>LA sessions attended</th>
<th>Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (as recorded after week 4)</td>
<td>558 (78.3%)</td>
</tr>
<tr>
<td>1 – 3 sessions</td>
<td>48 (6.7%)</td>
</tr>
<tr>
<td>4 – 8 sessions</td>
<td>79 (11.1%)</td>
</tr>
<tr>
<td>9 or more sessions</td>
<td>28 (3.9%)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>713 (100%)</td>
</tr>
</tbody>
</table>

Attendance increased as the semester progressed. For instance, 59 students attended LA-led sessions in Week 4; 79 students attended in Week 8; 103 students attended in Week 10.
**Conceptual Learning**

**Learning Assistant data:** LA’s only completed the General Chemistry I Concept Survey at the end of the semester. Therefore, no learning gains can be calculated. The average score for LA’s at the end of the course was 73.3% \((n = 5)\).

**Whole class data:** Normalized learning gains were calculated for CHEM 1111 students. We compared the entire class performance to those students who attended at least one LA help session:

<table>
<thead>
<tr>
<th></th>
<th>Pre avg</th>
<th>Post avg</th>
<th>Gain avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entire class ((n = 595))</td>
<td>41%</td>
<td>55%</td>
<td>24%</td>
</tr>
<tr>
<td>At least 1 LA session ((n = 137))</td>
<td>37%</td>
<td>52%</td>
<td>23%</td>
</tr>
</tbody>
</table>

The learning gain differences are not significant, but it does suggest that students who started out with less conceptual understanding made of LA’s in the course.

We also binned students who made use of LA’s by LA session attendance (1 – 3; 4 – 8; more than 9 sessions) and by learning gains (high, medium, low, or negative). A chi square analysis reveals no significant differences.

**Correlations among LA session attendance, exam / homework scores, and course grades**

**All CHEM 1111 students:** For the entire class \((n = 712)\), LA session attendance was significantly correlated to Chemistry Concept Challenge score \((r = .396; p = .000)\) and total course points \((r = .172; p = .032)\). This makes sense, as LA’s assisted students with these assignments, which then contributed 3% toward the total course points. A chi square analysis indicates a significant relationship between LA session attendance and grade in the course (which is of course, a function of total course points). Below is a graph of grade distributions by LA session attendance:
CHEM 1111 students who attended at least one LA session:
For these students \((n = 138)\), LA session attendance was significantly correlated to Chemistry Concept Challenge score \((r = .383; p = .000)\), Exam 2 score \((r = .182, p = .033)\), clicker score \((r = .177; p = .038)\), and total course points \((r = .221; p = .009)\). Students were binned by final course grade and LA session attendance. Although a chi square analysis reveals no significant differences \((p = .149)\), this graph does suggest that students who attended more often tended to earn A’s and B’s in the course.
**Attitudes and Beliefs: CLASS data**
Learning Assistant data: These data were not collected for LA’s.

**All CHEM 1111 students:** As mentioned previously, some baseline data have been collected using the CLASS. Below is a graph showing shifts towards “expert” attitudes and beliefs about chemistry and learning chemistry for three Fall semesters of CHEM 1111. Instructors were mostly the same during this period (the course is team-taught by two or three instructors). The main difference is the use of Learning Assistants and Chemistry Concept Challenge homework during 2006. Shifts are less negative for “All Categories” in Fall 2006, although it is unknown whether these are significantly different from the previous Fall semesters.

The subscale dealing with chemistry-specific ways of thinking, such as visualizing molecular structures, is the only one that shifted in a positive direction, and this does look quite a bit different than in 2005, when LA’s and conceptual homework were not used.

T-tests comparing pre- and post- scores for the 401 students who completed both during Fall 2006 indicate that the 2006 shifts are significant, except for “Problem Solving Sophistication” and “Applied Conceptual Understanding.”

The Fall 2006 data are in process of being further analyzed in terms of LA session attendance. The shifts in All Categories suggests differences between students who attended LA sessions and those who did not, but the significance of those differences still needs to be tested:

<table>
<thead>
<tr>
<th>LA sessions attended</th>
<th>Shifts in All Categories (favorable)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (as recorded after week 4)</td>
<td>-3.83 ($n = 299$)</td>
</tr>
<tr>
<td>1 – 3 sessions</td>
<td>-1.04 ($n = 22$)</td>
</tr>
</tbody>
</table>
4 – 8 sessions  -0.47 (n = 51)
9 or more sessions  -0.99 (n = 19)
Total  -3.10 (n = 391)

These data are also represented graphically in terms of pre- and post- scores on All Categories. The dotted line represents zero shift in attitudes / beliefs. Students above the line exhibit positive shifts; those below had negative shifts during the course of the semester.
Choice to join LA-led learning group

Students were provided with the choice to join an LA-led learning group at the beginning of the semester, which was intended to provide a small group setting to discuss concepts and to work through the Chemistry Concept Challenges. Below is a summary of student participation in learning groups:

<table>
<thead>
<tr>
<th>Choice and Learning Group attendance</th>
<th>Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did not join a group</td>
<td>387 (74%)</td>
</tr>
<tr>
<td>Joined; attended fewer than 8 of 12 sessions</td>
<td>13 (2.5%)</td>
</tr>
<tr>
<td>Joined; attended at least 8 sessions</td>
<td>6 (1.1%)</td>
</tr>
<tr>
<td>Joined; attended 9 – 12 sessions</td>
<td>115 (22.0%)</td>
</tr>
<tr>
<td>Total</td>
<td>522 (100%)</td>
</tr>
</tbody>
</table>

There were 140 CHEM 1131 students that signed up for learning groups (a few students dropped the course after signing up). Grades from CHEM 1111 are available for 134 of these students:

<table>
<thead>
<tr>
<th>CHEM 1111 Grade</th>
<th>Number of CHEM 1131 Students who joined LG’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>30</td>
</tr>
<tr>
<td>B</td>
<td>51</td>
</tr>
<tr>
<td>C</td>
<td>52</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
</tr>
<tr>
<td>No grade</td>
<td>6</td>
</tr>
</tbody>
</table>

Students were asked at the beginning of the semester why they wanted to join a learning group. Here is a summary and distribution of their responses:
At the end of the semester, students were asked whether they were satisfied with their choice to either join a learning group or not. Most students were satisfied with their choice, although 20% of students who didn’t participate in a learning group later wished they had joined. (Only 3% of students in learning groups regretted their decision at the end of the semester.)

<table>
<thead>
<tr>
<th></th>
<th>Joined learning group</th>
<th>Did not join learning group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfied with decision</td>
<td>121</td>
<td>257</td>
</tr>
<tr>
<td>Not satisfied with decision</td>
<td>4</td>
<td>66</td>
</tr>
</tbody>
</table>

A few student quotes relating directly to LA’s and the learning group environment:

The LA for the learning group I attended was very knowledgeable and very good at explaining the concepts.

The LA helped our group a lot and helped us understand what the questions were looking for instead of just giving us the answers.

All of the LA’s I have encountered have been great teachers.

I learned more from the LA group than in recitation or lecture. Things came together with those worksheets and groups.

It gave me a comfortable forum where I could ask questions about topics that I did not understand, remember, or thought were cool.
The groups enable me to not only get the right answer but helped me understand the material.

I think it helped, as it was less intimidating (I'm shy) to ask questions.

It was helpful seeing how others thought about these topics and putting my reasoning into words with the help of those around.

The chemistry “groups” created an atmosphere of working together to understand, which allowed all of us to think about the ideas in ways we may not have thought about otherwise.

These student quotes suggest that LAs were able to put into practice what they were learning in their weekly education seminar and departmental meetings. They were able to create a safe place for students to explore ideas and to develop their understanding. Although some students’ reasons for being satisfied with the group were based on earning points and improving their grade, it seems that the LA’s were able to provide an atmosphere that focused more on reasoning and understanding and less on identifying right answers.

**Conceptual Learning**
The CHEM 1131 concept survey data have not yet been processed. There is only a small set of questions from pre to post that is the same; this survey instrument is in the beginning stages of development. Results will be reported to the LA-TEST group within a few weeks.

**Learning group membership and course grades / exam scores**

Course grades take into account more than student learning; they also reflect student motivation, ability to turn homework in on time, test-taking skills, etc. Although use of a good concept survey is more desirable in terms of measuring student learning, course grades also provide useful comparisons.

Chi square analyses were carried out to examine the relationship between learning group membership and final CHEM 1131 course grade. Specifically, we wanted to know what happened to students who came into CHEM 1131 with an “A” or a “B” or a “C” from CHEM 1111. The analyses showed that “A” and “C” students benefited from learning groups in terms of being able to at least maintain (or in the case of “C” students, improve) their grade in CHEM 1131, compared to students who opted not to join a group. (The one “A” student who regularly attended a learning group but still earned an “F” in CHEM 1131 is quite an anomaly; this student didn’t actually turn in any Concept Challenges, labs, or take the final exam.)
Learning group membership did not make a difference in exam scores. Although some attempt was made to write exam questions that were more conceptual in nature and reflected the kinds of thinking students did on the Concept Challenges, the overall flavor of the exams didn’t change much from previous semesters. Thus changes made in this one aspect of the course (adding learning groups that focus on conceptual thinking) may not show up here.
**Attitudes and Beliefs: CLASS data**

Learning Assistant data: These data were not collected for LA’s.

All CHEM 1131 students: Shifts were quite small (although generally in the undesirable direction). Although the shifts are slightly larger in magnitude than Spring 2006, the absolute scores are higher—often by 10%.

There were no statistical differences between students who joined LA-led learning groups and those who did not.
Parting Thoughts

The Department of Chemistry has only been involved in the LA program for one year. In that time, our model for using LA’s has evolved, and it continues to do so. Although dramatic results were not evident from this first year (such as large learning gains in CHEM 1111, or even significant differences in students’ attitudes and beliefs), it may be unrealistic to expect that at this point. Even though the newly developed Chemistry Concept Challenges and use of Learning Assistants were designed to impact students’ conceptual understanding and attitudes and beliefs about chemistry, it is one small part of a course that has multiple components (lecture, lab, computerized homework, recitation). Students may need to encounter conceptual problems in multiple course components and receive explicit messages related to chemistry learning and the applicability of chemistry to the real world before we see real shifts in their attitudes and conceptual understanding. We continue to integrate course components as we transform General Chemistry.

That said, students who joined LA-led learning groups clearly valued their experiences, and they did seem to make a difference in terms of being able to at least maintain the grades they earned in CHEM 1111. The learning assistants learned a lot in our weekly meetings and by leading these groups. It’s difficult to support this claim without evidence from a concept survey; more analysis needs to be completed in this regard. In terms of data collection for Year 2, it is much more evident what type of data needs to be collected and analyzed as we move forward with the use of LA’s in Chemistry courses.

Perhaps the biggest indicator that students valued their experiences as a Chemistry LA is that many of them are returning either as paid LA’s for a second semester, or in a volunteer capacity. Two of the ten LA’s from Spring are returning in the Fall. Four LA’s are volunteering in some way next semester so that they can remain involved in the project. (As of now, these students do
not intend to become teachers, or need to explore other options, and they don’t want to tie up a paid LA position.) Two of the newly-hired LA’s for Fall participated in learning groups in the Spring. We are excited to integrate LA’s with TA’s in recitation in the Fall (CHEM 1131) as well as expand into Organic Chemistry I. That course will be very interesting to observe, as those students have experienced LA’s in Chemistry this past year.
References


---


ii Supported by the National Science Foundation Grant DUE-0434144.