Appendix A1
Overview of Research Methods for the IBL Mathematics Centers Study

A1.1 Introduction

We conducted a large, mixed-methods study of inquiry-based learning (IBL) in college mathematics, comprised of six linked sub-studies of inquiry-based and comparative courses that were developed and taught at four university IBL Mathematics Centers. The study was designed to examine the following questions:

1) What are the student outcomes—including learning, attitudes, beliefs, career and education plans—of IBL mathematics courses?
2) How do these outcomes vary among student groups, and how do they compare with other types of courses?
3) How do these outcomes come about? In particular, what is the role of students, instructors and teaching assistants, course materials, assignments, assessments, and other classroom practices?
4) What are the costs and benefits for instructors and departments who teach with IBL methods?

The full report described selected results from our analyses. Chapters 2-8 each focus on the results from a particular sub-study. To communicate these results efficiently, we do not provide technical details of our research methods in the chapters. Rather, we summarize our approach to each in Chapter 1 of the report, highlighting the strengths and limitations of each method. We strongly encourage readers to begin with this chapter to understand the study as a whole, as well as the purpose and scope of this report. We also briefly recap the sub-study conceptual design at the start of each findings chapter.

A1.2 Organization of the Appendices

Our methods of data-gathering and analysis, and the samples for each sub-study, are described in an appendix corresponding to each chapter, as listed in Table A1.1. Some of the tools that we developed may be useful to other evaluators or researchers, such as observation and interview protocols and survey instruments. These are included as exhibits, labeled E2.1, E2.2, and so on, and they follow the appendix to which they are relevant. These are also listed in Table A1.1. All the appendices are available online at:
http://www.colorado.edu/eer/research/steminquiry.html#Reports

Cite as: Assessment & Evaluation Center for Inquiry-Based Learning in Mathematics (2011). (Report to the IBL Mathematics Project) Boulder, CO: University of Colorado, Ethnography & Evaluation Research.

Table A1.1: List of Appendices and Exhibits of Research Methods

Chapter	Appendix	Topic
1	A1	Introduction & background of the study; overview of methods
2	A2	Classroom observations E2.1 - summary sheet E2.2 - observer survey E2.3 - classroom log
3	A3	Student surveys E3.1 - attitudinal pre-survey E3.2 - learning gains post-survey (SALG-M)
4	A4	Linkages between observations and surveys
5	A5	Mathematics tests E5.1 - proof test questions E5.2 - instructor ratings, questions
6	A6	Student academic records
7 8	A7	Student interviews Instructor interviews E7.1 - instructor interview protocol E7.2 - student interview protocol

A1.3 Broad Design of the Study

As discussed in Chapter 1, the overall project addresses the broad question of whether, how, and for whom IBL teaching and learning approaches are effective in college mathematics. We sought to examine a range of student learning and affective outcomes as well as longer-term impacts on students' education and career paths. We studied the classroom context and the teaching and learning processes that took place in and out of class, and the contextual factors that influenced instructors' choices and the success and sustainability of the IBL programs on each campus.

With these goals, the study was designed as a checkerboard of sub-studies (Figure A1.1) that combine to build a detailed picture of where and how IBL methods do and do not "work" for students and their instructors. Classroom observations provide a foundation enabling us to describe the teaching methods in use and link student outcomes to particular teaching approaches. Surveys, tests, academic records, and interviews allow us to probe both student outcomes and learning processes using multiple methods. End-of-course measures focus on student outcomes, while pre-course measures let us assess whether students are selectively choosing (or being advised) into and out of IBL courses. Interviews with faculty and TAs provide their observations of student outcomes, crucial perspectives on teaching goals and methods, and document the costs, benefits and career impact of teaching an IBL course.

Figure A1.1: Design Matrix for Investigation of Student Outcomes: Approaches for Examining Outcomes for Key Student Audiences

	Math learning & thinking: Tests	*Math learning & thinking: Self-report*	*Attitudes & beliefs*	*Career & educational outcomes*	*Classroom processes*
Math, science, & engineering majors	Proof test*	Survey* Interviews	Survey* Interviews	Academic records* Interviews	Observation* Interviews
Pre-service K-12 teachers	LMT test	Survey Interviews	Survey Interviews	Interviews	Observation Interviews

* Comparative data was gathered from non-IBL sections of some courses.

Table A1.2 shows the total amount of data gathered in the study. The numbers of participating IBL Mathematics Centers, course sections (IBL and non-IBL), and individuals (students, faculty or TAs; IBL and non-IBL) are itemized. By any measure, this is a large study.

Table A1.2: Data Gathered for IBL Mathematics Centers Evaluation Study, 2008-2010

INSTRUMENT	**IBL Centers**	**IBL Sections**	**Non-IBL Sections**	**IBL Individuals**	**Non-IBL Individuals**	**Individuals Total**
Attitudinal surveys						
Pre-survey	4	47	17	847	399	1246
Learning gains	3	6	1	112	88	200
Post-survey	4	55	18	840	325	1165
Mathematical Knowledge and Thinking						
LMT pre-test	2	9	-	187	-	187
LMT post-test	2	9	-	173	-	173
Proof test	2	8	8	87	35	122
Academic Records						
Transcripts	3	28	110	552	2866	3418
Interviews						
Students	4	15	-	68	-	68
Faculty	4	N/A	-	23	-	23
TAs	4	N/A	-	20	-	20
Observation	**IBL Centers**	**IBL Sections**	**Non-IBL Sections**	**IBL Class Sessions**	**Non-IBL Class Sessions**	**Sessions Total**
Courses	3	36	15	213	89	302

Appendix A2
Study Methods: Classroom Observation

A2.1 Introduction

The classroom observation study was designed to document classroom practices and interactions so that we could characterize IBL and non-IBL teaching practices, including their variation, and link these practices to student outcomes. We sought to address the following questions

- How are classrooms designated "IBL" alike or different from comparative classrooms?
- What practices and features are commonly applied in IBL courses, and what are the variations among them?
- What classroom features are seen in course sections where there are good student outcomes? How do these compare with classroom features of sections where outcomes are less positive?

We sought to address these questions with high-quality, internally consistent data yet without extensive investment of both observation and analysis time. Thus we designed a protocol that could be largely completed during class in real time, by trained but non-expert observers. The protocol documented classroom activities with simple, quantifiable indicators of student and faculty behaviors, rather than requiring subtle judgments of pedagogical effectiveness as used in instruments such as the Reformed Teaching Observation Protocol (RTOP) (Pilburn et al., 2000). To augment the quantitative data, we asked observers to provide notes and comments to help us interpret the data and to capture their own evidence-based judgments of less tangible variables such as classroom atmosphere. Moreover, we did not videotape or transcribe course sessions, because of the cost involved of collecting and, especially, analyzing such data. These choices sacrifice some detail but enabled us to gather a large volume of data across many sections.

A2.2 Study sample

Three campuses participated in the observation study. Altogether, 52 course sections were observed for multiple class sessions by trained observers, 36 sections in Year 1 and 16 in Year 2. Seven of the Year 2 sets were omitted from the analysis because there was a high incidence of missing data or too few hours of class time were observed. Another set of observations was procedurally valid but was omitted from the data analysis as representing an "experimental" hybrid course that was not easily classified as IBL or non-IBL.

For two large lecture courses that also included a recitation session, we sought to ensure that the observation data to represent a student's overall experience, not just the lecture. The lecture session was observed for multiple periods, and separately, a sample of recitation periods was observed (e.g. 6 recitations taught by different TAs). Observation data from lecture and recitation were combined in a 3:1 weighted average to reflect the three hours of lecture and one hour of recitation that any student would experience in a week.

With these adjustments, the total observation sample included 42 course sections: 31 IBL and 11 non-IBL sections of 18 different courses on 3 campuses. All the non-IBL sections observed were chosen from courses that also offered IBL sections, but fewer non-IBL sections were observed because (1) a comparable non-IBL section was not available for every IBL course we wished to observe; (2) some non-IBL instructors declined to participate; and (3) the non-IBL courses emphasized lecture. Since these were more homogeneous in style, a smaller sample seemed to be representative. This last assumption is borne out by the data: non-IBL courses exhibit much narrower distributions around the mean for nearly every practice-oriented variable.

A2.3 Observation protocol

The observation protocol was based on a study by Gutwill-Wise (2001; see also ModularCHEM Consortium, 1998, 1999) comparing active-learning and more traditional versions of a reformed undergraduate chemistry courses at two institutions. The content of the protocol was adapted using information from a preliminary study of five IBL Mathematics Centers, which included sites visits and observation of two to six IBL class sessions at each campus. Additional information was drawn from focus groups with students and graduate teaching assistants, and interviews with campus leaders and faculty. The protocol included three main components:

A. A summary sheet, where observers recorded basic data about the time and date of the class, the instructor, and the number, gender and apparent ethnicity of students attending. Observers also provided, in writing, their overall impressions of features such as classroom interactions, mood, morale, and any special context features (e.g., class held the day before an exam). This sheet is included as Exhibit E2.1.

B. An observer survey, where observers estimated the proportion of students who participated in class and rated the frequency of 14 student and instructor behaviors on a 5-point Likert-like scale (1=never to 5=very often). Behaviors included offering ideas, asking questions, working with others, listening to others, setting the pace or direction, giving feedback, and were chosen as indicators of the classroom atmosphere and interactions of class members. The observer survey is included as Exhibit E2.2.

C. A classroom log, where observers tracked class activities, leadership roles, and question-asking behaviors, as detailed below. The classroom log is included as Exhibit E2.3.

The classroom log included three different classification schemes, each based on a set of simple letter codes to record categories of behavior: class activities, leadership roles, and question-asking. The scheme for class activities (Scheme 1) was adapted from Gutwill-Wise (2001) to incorporate all activity types observed or reported in the preliminary site visits. Coding classroom activities enables us to determine differences in practice between IBL and non-IBL courses, and to gather a rough measure of the level of inquiry actually implemented in an IBL course. Scheme 2 was added to document the active roles of instructors, TAs, and students. Scheme 3 addressed the nature of questions asked by instructors and students.

Starting at the beginning of class, observers recorded the clock time and categorized the main activity and the lead role, using Schemes 1 and 2 respectively.

Scheme 1: Main activity

B	Addressing class business, procedural activity (e.g. returning papers)
L	Professor lecturing—presenting *previously prepared* material. This may include response to student questions during a lecture, but the L code is retained if the question does not turn into a multi-student discussion.
E	Extended explaining, *in response to* question or difficulty (instructor or student). This is an extended discourse or mini-lecture, different from L because it is not pre-planned, but responsive to an issue that arises on the spot.
G	Working on a problem or an example in groups, in seats or at the board (informal—class working in groups on problems, instructors circulate among groups)
P	Students presenting a solution or proof (individuals or groups).
D	Class discussing or critiquing a solution that has been presented. Usually whole-class.
C	Students working at computers—on a problem, modeling task, visualization, etc.
O	Other (describe)

Scheme 2: Lead role (these codes also used to identify anyone asking a question)

f	Faculty or lead instructor
t	Teaching assistant—any graduate or undergraduate student assisting the lead instructor
s	Single student
ns	New student (first time to participate today)
rs	Repeat student (has already participated today)
g	Group of students
c	Whole class/multiple roles

Observers were asked to make a real-time judgment of when the activity or lead role changed, and to record the time and the new letter code when either changed. They could also write comments to clarify activities or transitions. In practice, this required some judgment, since many activities changed spontaneously and observers had to decide, for example, at what point a dialogue between one student and a professor became a whole-class discussion involving multiple students. Leadership by a "new" or "repeat" student was recorded so that we could distinguish participation by a few, very active students, or a broader group.

In addition to the activity and role codes, observers recorded and coded each question that was asked by an instructor or student. We focused on question-asking because questions are central to "inquiry." Summaries of the literature indicates that questions are related to students' cognitive activity (Gall, 1970; Wilen, 1982; Edwards & Bowman, 1996). Teachers tend to ask mostly low-level questions requiring recall rather than higher-order reflection, but the cognitive level of teachers' questions correlates positively to the cognitive level of students' reply. Thus the type of questions asked by instructors should relate to students' thinking and learning. Student question-asking has been studied less than instructor questioning, but Edwards and Bowman (1996) suggest that a shift in the incidence of student-asked questions may reflect a shift in the view of the teacher from sole classroom authority to one who guides student-teacher interaction. Kawanaka and Stigler (1999) argue that not only the number but the type of student questions is related to learning. Thus we documented the frequency and type of student questions, as a further indicator of the cognitive demand and inquiry nature of a course.

Observers used Scheme 2 to indicate the question-asker and Scheme 3 to indicate the question type. Scheme 3 is a simplified version of Bloom's taxonomy (1956) adapted from Gall (1970) to include questions with functions other than cognitive ones.

Scheme 3: Question type

R	Recall or factual. Closed-ended; there is a right answer. Lowest in cognitive demand.
E	Explanatory/descriptive—seeks to draw out or build toward an explanation of how or why something is done. More cognitively demanding than recall questions.
C	Critiquing—asking for evaluation or judgment of an idea. High in cognitive demand.
S	Stretching/linking—asking for expansion, connection, creativity. These are high on cognitive demand.
M	Monitoring/involving. Includes instructor monitoring (Did you understand this? Are we ready to move on?) and student self-monitoring, checking, clarifying (Am I sure I understood this?). Also questions intended to generate metacognition or to involve others (Sally, what do you think?). These are process-oriented more than cognitive.
B	Business/procedural, accomplishing course business (Did everyone get their paper back? Who wants to present next? When is our homework due? Will that be on the test?)
X	Other, unknown, or unclear in type

Question types were linked to the activity episode during which each question occurred, based on findings by Edwards and Bowman (1996) that student questions are influenced by the instructional format in which the question occurred. Thus, for example, we can analyze the number and types of questions that occurred during lecture or during discussion. Observers listened but did not track questions during small group work, because they could not attend to more than one group at a time and we did not have a way to choose a group randomly. In practice, observers found categorizing question types (Scheme 3) to be more difficult than categorizing the activity or lead role (Schemes 1 and 2).

A2.4 Data Collection

Because the research team members could not be present on all three campuses for the length of time needed to observe multiple class sessions, we recruited classroom observers on each campus. We provided a job description to the campus leader or internal evaluator, who then recruited or circulated the description. Different pools of people were available on different campuses, but all the observers had degrees in mathematics and interest in teaching, including graduate students in mathematics and mathematics education and a doctorally trained mathematics education researcher.

We trained the initial group of observers in a 2-3 hour session conducted in person at their own campus. Observers read advance materials about the study design and observation protocols, and in the training session we reviewed the materials and practiced applying the protocol to video-recordings of two different IBL classrooms. Every observer signed a confidentiality agreement to keep the data and their own opinions confidential. On two campuses, we retained our observers from Year 1 to Year 2. On the third campus, the campus project evaluator, who had previously participated in the training, trained a new graduate student, and we held a conference call to discuss the protocol and answer his questions. At the beginning of Year 2, we also sent new copies of the protocol (with a few minor revisions called out) and a reminder of responsibilities to the returning observers.

To ensure a representative sampling of actual classroom processes, each class was visited multiple times at least two different points in the academic term. The aim was to capture 8-12 hours of class time for each course section. In practice, observers' adherence to this schedule varied, but we were nonetheless able to document several class sessions for every section observed. On average, 6.9 hours of class time were observed for every section included in the analysis, for a total of 298 hours of observation. During this time, over 2200 distinct episodes of instructional activity and over 10,300 questions were logged.

A2.5 Data analysis

Raw data were entered into a pre-formatted Excel spreadsheet by undergraduate assistants, who also assisted with compiling and tallying data for individual class sections, using Excel's conditional counting and summing functions. Text data (such as comments and notes) were transcribed for qualitative analysis.

For the observer survey, we computed mean ratings for each survey item across all observed sessions. For the classroom log data, most variables were analyzed on a cumulative or average basis over all class hours observed, rather than by class session, because class sessions varied in length from 45 to 90 minutes. Frequencies of specific behaviors or events are normalized to an hourly basis (e.g. episodes of discussion per hour of class time).However, variables that depend on student attendance (e.g. % of all students who ask a question) are necessarily based on class sessions, because attendance varies by session. For question-asking variables, we normalize to hours of non-group work, because questions were not tracked during small group work (e.g. questions asked by students per hour of non-group work).

Variables constructed for analysis included:

- Percentage of observed class time spent on each of the activity types in Scheme 1
- Mean frequency of each activity in Scheme 1, per hour of class observed

- Percentage of observed class time under leadership by each participant type in Scheme 2
- Mean frequency of each leadership role in Scheme 2, per hour of class observed
- Mean number of questions asked by each participant type in Scheme 2, per hour of non-group work
- Percentage of all questions asked by each participant type in Scheme 2
- Mean number of students, and percentage of all students in attendance, who asked a question, for a given class session
- Percentage of all questions of each question type in Scheme 3
- Mean number of questions per hour of each question type in Scheme 3
- Mean frequency of Scheme 3 question types for each Scheme 1 activity type (e.g. number of recall questions per hour during discussion or during lecture)

We computed basic descriptive statistics for IBL and non-IBL courses, and t-tests for statistical significance, using Microsoft Excel. Analyses of correlations among observation variables and relationships between observation and student survey data were conducted using SPSS.

A2.6 Data reliability and validity

We performed several types of checks on the data to ensure that they were as precise and accurate as possible. Observer coding omissions represent one type of uncertainty in the data. Observers omitted very few codes for class time by instructional activities and leadership role (Schemes 1 & 2). Counting and coding questions was more difficult, in part because questions sometimes came rapidly. Questions that were uncategorized by asker (Scheme 2) accounted for 1.2% of all questions, and those uncategorized by type (Scheme 3) accounted for 4.5% of all questions. Because Scheme 2 is straightforward to apply, we use that omission rate to estimate the rate of simple error (forgetting to circle a choice), i.e. 1.2%. Because Scheme 3 is more difficult to apply, the higher omission rate likely reflects real difficulty in categorizing questions.

Early in the study, some pairs of observers visited the same class session so that they could compare notes and discuss issues. We used these observations as a check on inter-rater reliability, though we did not repeat this test after observers were experienced and confident. Comparison of independent observations show that observers agreed to a very high extent in categorizing the nature of the activity (Scheme 1) and lead role (Scheme 2) but were less well aligned in recording question-asking behaviors. The total counts of questions were most variable, depending on how observers counted a single utterance with multiple embedded questions (e.g. recording “What do you think? Did everybody follow?” as 1 or 2 questions). Observers categorized questions with reasonable consistency, however: While their raw numbers of questions varied, the percentages of questions categorized both by asker (Scheme 2) and type (Scheme 3) were the same within 5 to 20%.

On the observer survey, observers did not differ by more than one point on any single item in their rating of the frequency of various student and instructor behaviors. Typically, they agreed exactly on a majority of the items, and the sum of their scores on all 14 ratings differed by less than 5% of the total possible rating points.

Given the high volume of data, data entry errors are another potential source of uncertainty. Because two codes were recorded for classroom time (Schemes 1 and 2) and two codes for question (Schemes 2 and 3), we could catch most data-entry errors by comparing the number of minutes or questions logged under each scheme, which should be identical. Where the totals did not match, we checked the spreadsheet data against the raw data and were able to identify and correct nearly all such errors. We estimate the proportion of undetected data entry errors as rather less than 0.5%.

A2.7 References cited

Bloom, B. S., ed. (1956). *Taxonomy of educational objectives: Handbook 1: Cognitive domain*. New York: David McKay.

Edwards, S., & Bowman, M. A. (1996). Promoting student learning through questioning: A study of classroom questions. *Journal on Excellence in College Teaching*, **7**(2), 3-24.

Gall, M. D. (1970). The use of questions in teaching. *Review of Educational Research,* **40**(5), 707-721.

Gutwill-Wise, J. (2001). The impact of active and context-based learning in introductory chemistry courses: An early evaluation of the modular approach. *Journal of Chemical Education,* **78**, 684-690.

Kawanaka, T., & Stigler, J. W. (1999). Teachers' use of questions in eighth-grade mathematics classrooms in Germany, Japan, and the United States. *Mathematical Thinking and Learning,* **1**(4), 255-278.

ModularCHEM Consortium (1998, January 23-24). MC^2 evaluation progress report: January 1998, in *Fourth Meeting of the ModularCHEM Consortium National Visiting Committee*. (Report to the National Visiting Committee) Berkeley, CA: ModularCHEM Consortium.

ModularCHEM Consortium (1999, January 29-30). ModularChem Consortium evaluation report to the National Visiting Committee, in *Fifth Meeting of the ModularCHEM Consortium National Visiting Committee*. (Report to the National Visiting Committee) Berkeley, CA: ModularCHEM Consortium.

Pilburn, M., Sawada, D., Falconer, K., Turley, J., Benford, R., & Bloom, I. (2000). *Reformed Teaching Observation Protocol (RTOP).* Tempe, AZ: Arizona Collaborative for Excellence in the Preparation of Teachers.

Wilen, W. W. (1982). *Questioning skills, for teachers. What research says to the teacher.* Washington, DC: National Education Association.

Exhibit E2.1: COVER SHEET and SUMMARY
Classroom Observation Protocol for Mathematics

Date: Class start time: Observer:

Class end time:

Course name/number: Instructor:

Description of student population counted or estimated (circle one)

# of students	White women	White men	Women of color	Men of color	TOTAL
Present at start					
Entering later					

Notes – class context, interactions, mood, morale. What was interesting about this class? Please record what you observed as well as how you interpret it.

Exhibit E2.2: Classroom Observation Post-Survey

Fill out this survey right after observing a class. It aims to capture some general features to complement the detailed log. Use your own assessment based on what you saw and experienced, using the scale (1 = Never to 5 = Very often) to indicate the extent to which the activity or feature occurred in the session you saw. Please add any explanatory comments.

A. What percentage of students (approximately) participated in class? (circle one)

0-25% 25-50% 50-75% 75-100%

Was participation generally representative by gender and ethnicity? Yes No
Please comment.

B. To what extent did (single or groups of) students …	**Never**				**Very often**
1. Offer their own ideas during class?	1	2	3	4	5
2. Ask instructor/TAs questions?	1	2	3	4	5
3. Review or challenge other students' work?	1	2	3	4	5
4. Work together with other students?	1	2	3	4	5
5. Set the pace or direction of class time?	1	2	3	4	5
6. Get help from other students?	1	2	3	4	5
7. Receive personal feedback on their work?	1	2	3	4	5

C. To what extent did the instructor and TAs …	**Never**				**Very often**
1. Listen to students' ideas/explanations?	1	2	3	4	5
2. Express their own ideas or solutions to problems?	1	2	3	4	5
3. Set the pace or direction of class time?	1	2	3	4	5
4. Give concrete feedback on students' work?	1	2	3	4	5
5. Offer help to students?	1	2	3	4	5
6. Establish an overall positive atmosphere?	1	2	3	4	5
7. Summarize or place class work in a broader context?	1	2	3	4	5

Exhibit E2.3: Classroom Observation Log – Classroom Activities & Questioning Behaviors

Date: Course: Observer: Page _____ of _______

I. Classroom Activities. Mark start time & general activity at each change.			**II. Question-Asking Behaviors.** Mark asker & categorize.	
Time	Main activity B L E G P D C O	Lead role f t s g c	Asked by	Question type (with notes)
			f t ns rs	R E C S M B X
			f t ns rs	R E C S M B X
			f t ns rs	R E C S M B X
			f t ns rs	R E C S M B X
			f t ns rs	R E C S M B X
			f t ns rs	R E C S M B X
			f t ns rs	R E C S M B X
			f t ns rs	R E C S M B X
			f t ns rs	R E C S M B X
			f t ns rs	R E C S M B X
			f t ns rs	R E C S M B X
			f t ns rs	R E C S M B X
			f t ns rs	R E C S M B X

Appendix A3: Study Methods for Student Surveys

A3.1 Introduction

We used two survey instruments to measure student outcomes from inquiry-based learning in undergraduate mathematics and to compare these outcomes between various student groups, in particular, between IBL and non-IBL students. The attitudinal survey was designed to detect the quality of and changes in students' mathematical beliefs, affect, learning goals, and mathematical problem-solving strategies. The learning gains survey (SALG-M) measured students' experiences of class activities and their cognitive, affective and social gains from a college mathematics class. The surveys addressed the following questions

- What learning gains do students report from an IBL mathematics class?
- How do students experience IBL class activities? How do students' class experiences account for their gains?
- What kind of beliefs, affect, goals and strategies do IBL students report at the start of a mathematics course?
- How do these approaches change during a college mathematics course? How do these changes relate to or explain students' learning gains?
- For each of these outcomes—learning gains, experiences, attitudinal measures, and changes—how do the outcomes for IBL students differ from those of non-IBL students, and among IBL student sub-groups?

The survey instruments provided us with large student data sets from four campuses, gathered during the two academic years 2008-2010. They offered us a comprehensive picture of students' approaches to learning college mathematics as well as of their experiences and gains from IBL classes. Moreover, the survey data could be used to analyze differences in reported learning approaches, classroom experiences and learning outcomes among various student groups. In addition to structured questions, students also could write about their experiences and gains in the open-ended survey questions. Both the open-ended survey answers and student interview data were used to validate, confirm, and fill in the picture of student outcomes obtained from the structured survey responses.

A3.2 Study sample

The data were gathered on all four campuses in a variety of undergraduate courses. These included courses entitled:

- (Honors) Analysis 1-3,
- (Honors) Calculus 1-3,
- Cryptology
- Discrete mathematics,

- Explorations in mathematics,
- Exploratory calculus,
- Group theory,
- Introduction to proofs,
- Introduction to real analysis,
- Multivariate calculus 1-2,
- Number theory,
- Probability,
- Real analysis 1.

They covered the full range of introductory to advanced mathematics courses.

Mathematics courses specifically developed for elementary and middle school or secondary school pre-service teachers represented another type of course in the sample. This kind of survey data was obtained from two campuses. Additional smaller data sets came from a geometry course designed (but not required) for prospective high school mathematics teachers at one campus.

In all, we collected surveys from 82 college mathematics sections, of which 65 were IBL sections and 17 non-IBL sections. Data obtained with our surveys consisted of an attitudinal pre-survey, a learning gains post-survey, and a combined post-survey including both the attitudinal and the learning gains questions. We received pre-surveys from 1245 students, learning gains post-surveys from 200 students, and combined post-surveys from1165 students. Combining the pre-survey data with the post-survey data produced us information from 800 individually matched surveys. These surveys included responses from 412 IBL math track students (i.e., students who studied mathematics as their major or minor subject), 156 non-IBL math track students, 208 IBL pre-service teachers, and 25 non-IBL pre-service teachers.

Tables A3.1-A3.4 display features of our sample based on the personal information from the pre-survey responses.

A3.2.1 Survey Sample by Gender

Students reported their gender both in the pre- and post-survey. Even though these were not always the same students, the percentages of women and men were rather consistent in the two surveys. About 60% of all the students were men. This varied along with student groups. Typically, nearly 70% of the math-track students were men, whereas most of the IBL pre-service teachers (84% pre; 86% post) were women.

Table A3.1: Survey Respondents by Gender and Course Type

Gender	IBL math-track		Non-IBL math-track		IBL pre-service		Non-IBL pre-service		Total	
	Count	%	Count	%	Count	%	Count	%	Count	%
Pre-Survey										
Women	194	*33.6*	104	*30.4*	190	*83.7*	12	*48.0*	500	***42.7***
Men	383	*66.4*	194	*69.6*	37	*16.3*	13	*52.0*	671	***57.3***
TOTAL	577	*100%*	342	*100%*	227	*100%*	25	*100%*	1171	***100%***
Learning Gains Survey										
Women	169	*32.3*	92	*28.5*	190	*85.6*	17	*53.1*	468	***42.5***
Men	354	*67.7*	231	*71.5*	32	*14.4*	15	*46.9*	632	***57.5***
TOTAL	523	*100%*	323	*100%*	222	*100%*	32	*100%*	1100	***100%***

A3.2.2 Survey Sample by Academic Major

We classified students by their reported main major, prioritizing their most mathematically oriented major. Accordingly, all students with a major in mathematics or applied mathematics were classified into one category, even if they had a second, non-mathematics major. Science majors included students with a major in physics, chemistry or another science, but not in mathematics or applied mathematics. Engineering and computer science majors formed another category, as did students with a major in economics. All students who reported any a non-science major were classified into one group.

Table A3.2: Survey Respondents by Academic Major

Main academic major	IBL math-track		Non-IBL math-track		IBL pre-service		Non-IBL pre-service		Total	
	Count	%	Count	%	Count	%	Count	%	Count	%
Math or applied math	337	*60.3*	163	*49.4*	86	*38.7*	19	*76.0*	605	***53.3***
Science	82	*14.7*	53	*16.1*	18	*8.1*	1	*4.0*	154	***13.6***
Engineering or computer science	63	*11.3*	57	*17.3*	4	*1.8*	2	*8.0*	126	***11.1***
Economics	28	*5.0*	47	*14.2*	3	*1.4*	0	*0.0*	78	***6.9***
Other non-science	49	*8.8*	10	*3.0*	111	*50.0*	3	*12.0*	173	***15.2***
TOTAL	559	*100%*	330	*100%*	222	*100%*	25	*100%*	1136	***100%***

More than half the students reported a mathematics or applied mathematics major. Students who had a non-science major mostly represented IBL pre-service teachers. Science majors formed the next biggest student group. Engineering or computer science majors (11.1%) and economics majors were the two other majors represented in the sample. In addition to mathematics or applied mathematics students, students from other STEM fields were also well represented in the sample.

IBL math-track students were pursuing a math major slightly more often (60.3%) than non-IBL math-track students (49.4%). The proportions of science majors was similar, but more of the non-IBL math-track students were economics majors or engineers. Fully half of the IBL pre-service teachers were non-science (e.g., education) majors. These represented mostly elementary or middle school pre-service teachers. But the sample also included many secondary pre-service teachers with a math major.

A3.2.3 Ethnicity and Race

We classified students by race into three different categories. All the students who considered themselves white and not a representative of any other race were denoted White. The category Asian consists of all the students who considered themselves only Asian, or Asian and some other race. If the students did report some other race besides White or Asian, they were classified as multiracial students. Ethnicity was a separate item; here students could choose between Hispanic or Latino, or Not Hispanic or Latino. The distributions of respondents by ethnicity and race are shown in Table A3.3.

Table A3.3: Survey Respondents by Ethnicity and Race

	IBL Math Track		**Non-IBL Math Track**		**IBL Pre-Service**		**Non-IBL Pre-Service**		**Total**	
	Count	%	Count	%	Count	%	Count	%	Count	%
Ethnicity										
Hispanic or Latino	47	*8.3*	37	*11.1*	27	*12.2*	10	*41.7*	121	***10.6***
Not Hispanic or Latino	520	*91.7*	296	*88.9*	195	*87.8*	14	*58.3*	1025	***89.4***
	567	*100%*	333	*100%*	222	*100%*	24	*100%*	1146	***100%***
Race										
Asian	138	*25.6*	118	*37.9*	22	*11.0*	6	*31.6*	284	***26.5***
Multiracial	24	*4.4*	12	*3.9*	19	*9.5*	1	*5.3*	56	***5.2***
White	378	*70.0*	181	*58.2*	159	*79.5*	12	*63.2*	730	***68.2***
TOTAL	540	*100%*	311	*100%*	200	*100%*	19	*100%*	1070	***100%***

Less variety appeared in students' ethnicity and race. Most of the students were white and not Hispanic or Latino. About a quarter of the students were Asian (26.5%), but the sample included only a few students from other races (5.2%). The sample represents a distribution that is typical for mathematics students in the large research universities that our study targeted.

A3.2.4 Academic Status

The pre-survey provided us with information about students' academic status at the beginning of their mathematics course. Table A3.4 shows the distribution of respondents' academic background by course type.

Table A3.4: Survey Respondents by Academic Status

	First year		Sophomore or Junior		Senior or more		*Total*	
Student Group	Count	*%*	Count	*%*	Count	*%*	Count	***%***
IBL math-track	206	*35.8*	183	*31.8*	187	*32.5*	576	*100*
Non-IBL math-track	147	*43.2*	117	*34.4*	76	*22.4*	340	*100*
IBL pre-service teachers	3	*1.3*	99	*43.8*	124	*54.9*	226	*100*
Non-IBL pre-service teachers	0	*0.0*	10	*40.0*	15	*60.0*	25	*100*
TOTAL	356	*30.5*	409	*35.0*	402	*34.4*	1167	*100%*

Our sample included students across all stages of their college studies. However, nearly one-third (30.5%) of all the students were first-year students. This applied especially to IBL (35.8%) and even more to non-IBL (43.2%) math-track students. The pre-service teachers in the sample were further along in their studies. More than half (54.9%) were seniors or even more advanced students but only three of them were first-year students. The same trend applied to the small group of non-IBL pre-service teachers.

A3.3 Survey instruments

The final survey instruments consisted of an attitudinal pre-survey, a learning gains post-survey, and a combined post-survey including both the attitudinal questions and the learning gains questions. Both the pre- and post surveys gathered personal information about students' gender, race and ethnicity, class year, academic majors, grade-point average, and plans to pursue teaching certification. We asked students to set themselves an identifier at the end of each survey. These identifiers were used to match the pre-survey responses with the post-survey responses individually.

In order to check the survey items and the structures, both the attitudinal and the learning gains survey were tested with two small samples of college mathematics students. Descriptive statistics and principal component analysis were used with these preliminary data sets to check the reliability of the questions and theoretical constructs in the surveys. Based on these analysis, we left out ill-behaving questions and shortened the attitudinal survey. In order to shorten the combined post-survey, we also left out some overlapping questions from the initial learning gains survey. The final surveys are presented as Exhibit E3.1 and E3.2.

A3.3.1 Attitudinal Survey

We wanted to study the nature of students' mathematical beliefs, affect, learning goals, and mathematical problem-solving strategies, and changes in these during a college mathematics course. We designed a structured survey to measure undergraduate students' mathematical beliefs, affect, learning goals and strategies of problem solving, to be administered at the beginning and end of a college mathematics course. The seven sections measured students' interest in and enjoyment of mathematics, preferred goals in studying mathematics, and their frequency of use of various problem-solving actions when doing mathematics, and their beliefs about learning mathematics, problem solving, and proofs.

A3.3.1.1 Theoretical basis of the attitudinal survey

The sub-sections and items were constructed on the basis of theory and previous research on mathematical beliefs, affect, learning goals and strategies of learning and problem solving. Mathematics education research on beliefs has introduced concepts such as beliefs about the nature of mathematics, about learning mathematics, about problem-solving, and beliefs about the self as a mathematics learner (Malmivuori, 2001; McLeod, 1992). All these categories of beliefs appear to have important implications for how students approach the study of mathematics and act in mathematics learning situations at various age and schooling levels. They may either significantly hinder or help student learning, performance and problem solving (Leder, Pehkonen, & Torner, 2002; Schoenfeld, 1992). Moreover, they influence the development of negative or positive attitudes toward mathematics that have longer-term impacts on students' choices of studying mathematics.

We wanted to check the quality of and changes in these types of important mathematical beliefs. In addition, we chose to study certain types of beliefs that were particularly important for studying college mathematics and that might display possible differences between students in traditional and IBL mathematics courses. For example, mathematical proving represents an important area of beliefs in college mathematics. How students see the nature of proofs significantly affects their success in college mathematics (Knuth, 2002; Selden & Selden, 2007; Sowder & Harel, 2003). Moreover, previous research has identified some differences in these beliefs between students who took traditional or student-centered IBL mathematics classes students (Ju & Kwon, 2007; Yoo & Smith, 2007).

Based on these criteria, testing and revisions of the attitudinal survey, we measured students' beliefs about:

- learning of mathematics (instructor-driven, group work, exchange of ideas)
- mathematical problem-solving (practice vs. reasoning)
- mathematical proving (proving as a constructive activity or as confirming truths; Yoo & Smith, 2007),
- beliefs about the self (confidence in their own math ability, in teaching mathematics)

Studies on affect have a long tradition in mathematics education research. Attitudes about mathematics, confidence, motivation and anxiety are the most-studied factors, found to essentially strengthen or diminish students' willingness and ability to learn mathematics (Frost, Hyde, & Fennema, 1994; Goldin, 2000; Malmivuori, 2001, 2007; McLeod, 1992). Interest in and enjoyment of mathematics learning represent central features of affect and students' motivation. Interest is suggested to facilitate deep rather than surface-level processing, and the use of more efficient learning strategies (Entwistle, 1988; Schiefele, 1991). In turn, students who enjoy learning tend to exert more effort and persist longer when they are challenged (Stipek, 2002). Both interest and enjoyment indicate students' strong positive relationship to mathematics and willingness to spend time and effort in studying mathematics. This relationship is also importantly weakened or strengthened by students' confidence in their own ability to do and learn mathematics. This applies to female students in particular (Fennema, Seegers & Boekaerts, 1996). But confidence as related to enhanced self-efficacy is found to essentially promote all students' engagement and cognitive performance (Bandura, 1993; Malmivuori, 2001; Zimmerman, 2000).

Recent education psychological literatures suggest that the types of learning goals pursued by students also profoundly impact the quality of their learning. They direct students' level of achievement, self-regulation and problem-solving strategies (Pintrich, 2000). Closely related to personal interest, we studied students' learning goals, categorizing them broadly as intrinsic vs. external. For example, students who pursue high grades, seek particular degrees, and display high competence and self-concept express external or performance goals that are related to superficial learning. In contrast, intrinsic or mastery-focused goals such as a focus on one's own effort, pursuit of knowledge, and desire to understand the learned material are seen to result in independence, responsibility and deeper learning (Ames & Archer, 1988). In contrast to externally motivated students, intrinsically motivated students show higher interest, excitement, and confidence that enhance their performance, persistence, and creativity (Ryan & Deci, 2000).

IBL teaching practices often involve group work and collaboration that both require and develop communication skills (Gillies, 2007; Duch, Groh, & Allen, 2001). Indeed, recommendations for undergraduate programs in mathematics include development of analytical thinking, critical reasoning and problem-solving but also communication skills (Pollatsek et al., 2004). In the attitudinal survey, we wanted to study students' preferences for communicating about mathematics and any change in this during their IBL course. Items on students' goal of communicating about mathematics measured this preference.

In addition to mathematical confidence, our attitudinal survey studied students' affect and motivation in the form of:

- personal interest in mathematics,
- willingness to pursue a math major (or minor),
- plans to study more math in the future,

- interest in teaching,
- intrinsic and extrinsic learning goals,
- goal for communicating about mathematics,
- enjoyment of learning mathematics

Current understanding of mathematical knowledge as a constructive activity (Steffe & Thompson, 2000; Tall, 1991) focuses on skillful problem-solving (Schoenfeld, 1992). Competent mathematicians use strategies to make sense of new problem contexts or to make progress toward the solution of problems when they do not have ready access to solution methods for them (Schoenfeld, 2004). The nature of the problem-solving strategies they choose influence students' approaches to and success at challenging mathematical problems. For example, unlike novices, expert problem-solvers use high-level planning and qualitative analysis before attacking a problem. They also demonstrate facility in choosing appropriate strategies in various situations (Kroll & Miller, 1993; Schoenfeld, 1985). This contrasts with a lack of strategies, mechanical use of concepts, and rote memorization of previous similar problems.

IBL approaches provide opportunities for students to engage in knowledge creation and argumentation (Rasmussen & Kwon, 2007). Such activities are generally suggested to promote problem-solving skills, independent thinking and intellectual growth (Buch & Wolff, 2000; Duch, Gron & Allen, 2001). Competent problem solvers can communicate the results of their mathematical work effectively, both orally and in writing (Schoenfeld, 2004). Planning and self-monitoring the solving process also help to ensure skillful problem-solving (Schoenfeld, 1985). Moreover, mathematics students must develop persistence in the face of difficulties, tolerance for ambiguity, and willingness to try multiple approaches, and they must learn to apply the necessary amount of rigorous and judgmental reasoning (Hanna, 1991; Pollatsek et al., 2004). These skills require them to develop self-reflective and self-regulatory strategies (Burn, Appleby & Maher, 1998; De Corte, Verschaffel & Eynde, 2000).

We wanted to check what kind of learning and problem-solving strategies students report and how these change during IBL and traditional math courses. Our attitudinal survey studied students' use of:

- independent (or individual),
- collaborative, or
- self-regulatory strategies.

Items related to these strategies were intended to explore the extent to which students counted on their own thinking and creativity when solving math problems and proofs, shared their thinking and strategies with other students, and actively reflected on and regulated (planned or checked) their own thinking and actions while solving math problems.

A3.3.2 Example items from the pre/post attitudinal survey instrument

We divided the attitudinal survey into eight subsections, each consisting of 7-15 structured items. The answers varied on a 7-point Likert-scale between negative and positive responses. The subsections of the survey measured students' mathematical beliefs, motivation, learning goals, enjoyment, confidence, and strategies for leaning and problem-solving. All the attitudinal pre-survey questions and items are presented in Exhibit E3.1. We studied students'

- *Personal interest* in studying mathematics: How likely is it that you will…
  e.g., "Bring up mathematical ideas in a non-mathematical conversation?"
- *Enjoyment* of doing and discovering mathematics: How much do you enjoy…
  e.g., "Discovering a new mathematical idea?"
- *Goals in studying mathematics*: Below are some goals that students may have in studying mathematics. How important is each goal for you?
  e.g., "Memorizing the sets of facts important for doing mathematics." (extrinsic goal)
  "Learning to construct convincing mathematical arguments." (intrinsic goal)
- *Beliefs about the self*: confidence: How confident are you that you can…
  e.g., "Apply a variety of perspectives in solving problems?" (math ability)
  "Teach mathematics to high school students?" (teaching mathematics)
- *Beliefs about learning mathematics*: I learn mathematics best when…
  e.g., "The instructor lectures." (instructor-driven)
  "I work on problems in a small group." (group work)
  "I explain ideas to other students." (exchange of ideas)
- *Beliefs about problem-solving*: In order to solve a challenging math problem, I need…
  e.g., "To have lots of practice in solving similar problems." (practice)
  "To use rigorous reasoning." (reasoning)
- *Beliefs about proofs*: The following statements reflect some students' views about mathematical proof. How much do you agree or disagree with each statement?
  e.g., "Proof is a tool for understanding mathematical ideas." (constructive)
  "The main purpose of proof is to confirm the truth of a mathematical result that is already known to be true." (confirming)
- *Problem-solving strategies*: When you do math, how often do you take each action listed below?
  e.g., "Find your own ways of thinking and understanding." (independent)
  "Brainstorm with other students." (collaborative)
  "Plan a solving strategy before attacking a problem." (self-regulatory)

A3.3.3 Demographic and background information

The attitudinal survey also asked for demographic and background information about students' previous achievement, personal information, and expectation for the grade of the target course. The questions dealt with:

- achievement history: the highest level of high school mathematics taken, any AP Calculus test taken and scores received, number of college math courses taken, estimated overall GPA;
- academic background: class year, college major, pursue for a teaching certification;
- personal background: gender, ethnicity, race;
- expected grade for the course.

At the end of both the pre- and post-survey, students were asked to assign themselves an identifier. This enabled us to match between pre-survey and post-survey responses by individual student (see Exhibit E3.1).

A3.4 Learning Gains Survey

The learning gains post-survey was based on the SALG instruments (SALG, 2008) developed to enable faculty and program evaluators to gather formative and summative data on classroom practices. The questions address students' self-reported experiences of mathematics class practices and their cognitive, social and affective learning gains due to their participation in a college mathematics course. Students provide both quantitative ratings and written responses about the course focus, learning activities, content and materials. The learning gains instrument is grounded in its authors' (Seymour, Wiese, Hunter & Daffinrud, 2000) findings that:

- students can make realistic appraisals of their gains from aspects of class pedagogy and of the pedagogical approach employed, and
- this feedback allows faculty to identify course elements that support student learning and those that need improvement if specific learning needs are to be met.

The SALG instrument is easily modified to meet the needs of individual faculty in different disciplines and it has been found to be a powerful and useful tool for instructors in student feedback and course development. When first developed, data about the use of the survey showed that eighty-five percent of the instructors reported that the SALG provided qualitatively different and more useful student feedback than traditional student course evaluations. Instructors also made modifications to course design (60%) and class activities (lecture, discussion, hands-on activities) followed by student learning activities (54%) course content (43%), and the information given to students (33%) (Recommendations for using the SALG, 2008).

We adjusted the SALG items to match college mathematics situations. The final learning gains survey, which we call the SALG-M, consisted of four structured sections on course experiences and two sections on learning gains. The first four sections asked about students' experiences of instructional practices: how much particular practices helped their learning. The practices deal with overall instructional approach, classroom activities, tests and other assignments, and interactions during the course. Answers follow a five-point scale between "no help" and " great help." Two other structured sections of the questionnaire ask about students' gains in understanding, confidence, attitude, persistence, and collaboration. These answers vary on a five-point scale between "no gain" and "great gain." The final post-survey for the SALG-M is presented as Exhibit E3.2.

In addition to structured items, the learning gains post-survey included four open-ended questions. Students were provided space to write about:

- How the class changed the ways they learn mathematics
- How their understanding of mathematics changed as a result of the class
- How the way the class was taught affected their ability to remember key ideas
- What they will carry with them from the class into other classes or other aspects of their life.

Answers to these questions complemented numerical responses on students' gains from their mathematics courses, helping us to better understand these results.

The learning gains survey also gathered information on students' expected grade at the end of the course, college major, class year, gender, and whether they were pursuing teaching certification. These questions confirmed the match between pre- and post-surveys and enabled us to detect changes in students' ideas or plans. The complete post-survey consists of the structured items in the attitudinal pre-survey and all the sections of the learning gains survey.

A3.5 Data collection

Survey data were gathered from undergraduate students studying mathematics at all the four campuses during two academic years 2008-2010. We started with online survey instruments when testing the survey instruments and also gathered pre-surveys at one campus in early fall of 2008. Due to the low response rate to the online form, we gathered the rest of the survey data as a paper-and-pencil test in class, which yielded very high response rates. The paper questionnaire was administered at the beginning and end of each course. In the courses that were part of a multi-term sequence (e.g., a three-quarter calculus sequence), we administered the full post-survey only at the end of the final section, but also gave a learning gains survey at the end of each previous related section. This provided us with some longitudinal data on the evolution of students' experiences and learning gains over multiple terms of IBL or comparative instruction.

The surveys were delivered to our project collaborators at the four campuses who also mostly administered the surveys in class. In some cases, course instructors or teaching assistants administered the survey. Instructors were given instructions on how to administer the surveys and return the completed surveys to us. We also reminded them about keeping the confidentiality and anonymity of students in every step. Filling out the surveys took students about 10-20 minutes; instructors were asked to offer enough class time for completing the surveys.

A3.6 Data analysis

A3.6.1 Attitudinal survey variables

Composite variables were constructed based on the attitudinal survey design and the factors of mathematical beliefs, affect, learning goals, and strategies of learning and problem-solving presented in Section A3.3.1.1. Exploratory factor analysis, principal component analyses and item analyses on the attitudinal survey data were used to create the final composite variables: five measures of motivation, affect and confidence; three measures of learning goals; seven measures of beliefs about mathematics and learning; and three measures of strategies. For each composite variable, averages of student ratings across the items represented the score for each student. This enabled us to interpret results on the same scale as that for the original attitudinal survey items. The composite variables were then used to report results on students' attitudes and for further analysis on group differences in attitudes.

Table A3.5 displays the survey questions and items for each composite variable, the titles and descriptions of the composite variables, and the reliability scores for the composite variables (for the pre-survey data and post-survey data separately).

Table A3.5: Composite Variables Measuring Student Beliefs, Affect, Goals and Problem-Solving Strategies

Variable	**Description**	**Scale**	**Items**		**Reliability** Cronbach alpha	
			Count	Numbers	Pre	Post
Motivation						
Interest	Interest in learning and discussing mathematics	7	3	Q1: 5,6,7	0.808	0.828
Math major	Desire to graduate with a math major	7	1	Q1: 2	-	-
Math future	Desire to pursue math in future work or education	7	2	Q1: 1,4	0.439	0.615
Teaching	Desire to teach math	7	1	Q1: 8	-	-
Enjoyment	Pleasure in doing and discovering mathematics	7	6	Q2: 1-6	0.914	0.928
Confidence						
Math confidence	Confidence in own mathematical ability	7	5	Q9: 1,2,4, 5,6	0.820	0.826
Teaching confidence	Confidence in teaching math	7	2	Q9: 3,8	0.696	0.645

Table 3.5, continued…

Goals for studying math						
Intrinsic	Learning new ways to think & to apply math	7	4	Q3: 7-10	0.791	0.828
Extrinsic	Meeting requirements; degree, good grades	7	4	Q3: 1,3,4,6	0.724	0.744
Communicating	Communicating mathematical ideas to others	7	2	Q3: 2,5	0.783	0.810
Beliefs about learning						
Instructor-driven	Exams, lectures, instructor activities	7	4	Q5: 1,6,7,8	0.642	0.667
Group work	Whole-class or small group work	7	3	Q5: 2,3,5	0.685	0.719
Exchange of ideas	Active verbal interaction with other students	7	3	Q5: 9,10, 11	0.731	0.745
Beliefs about problem-solving						
Practice	Repeated practice, remembering	7	2	Q6: 2,6	0.690	0.758
Reasoning	Rigorous reasoning, flexibility in solving	7	5	Q6: 1,5,7, 8,9	0.734	0.712
Beliefs about proofs (Yoo & Smith, 2007)						
Constructive	Process view; revealing mathematical ideas	7	4	Q8: 2,6,7,8	0.637	0.675
Confirming	Product view; recall and confirming conjectures	7	3	Q8: 1,3,5	0.692	0.672
Strategies						
Independent	Finding one's own way to think & solve problems	7	4	Q4: 5,9,11, 12	0.747	0.775
Collaborative	Seeking help, actively sharing with others	7	3	Q4: 2,4,14	0.774	0.813
Self-regulatory	Planning, organizing, reviewing one's own work	7	6	Q4: 1,3,6, 7,8,10	0.747	0.747

A3.6.2 Learning gains survey variables

Similar to the treatment of attitudinal variables, composite variables were constructed on the basis of the questions and structures in the SALG-M survey. Exploratory and principal component analyses and item analysis produced five measures of instructional practices and nine measures of learning gains (see Table A3.6). The five composite variables related to instructional practices were used in reporting results on students' course experiences. Results on learning gains from the nine composite variables represented students':

- Cognitive gains: mathematical concepts, mathematical thinking, application of mathematical knowledge,
- Affective gains: positive attitude, confidence, persistence,
- Social gains: collaboration, comfort in teaching mathematics,
- Independence in learning mathematics.

Table A3.6: Composite Variables Measuring Student Experiences and Learning Gains

Variable	**Description**	**Scale**	**Items**		**Reliability** (Cronbach)
			Count	Numbers	
Experience of course practices (what helped me learn)					
Overall	Teaching approach, atmosphere, pace, workload	5	7	Q1: 1-7	0.898
Active participation	Personal engagement in discussion & group work	5	5	Q2: 3-7	0.839
Individual work	Studying & problem-solving on one's own	5	4	Q2: 2,8,9	0.695
Assignments	Nature of tests, homework, other assigned tasks	5	8	Q4: 1-8	0.764
Personal interactions	Interaction with peers & instructor, in/out of class	5	6	Q5: 1-6	0.696
Learning gains: Cognitive gains					
Math concepts	Understanding concepts	5	2	Q6: 1,2	0.921
Math thinking	Understanding how mathematicians think	5	2	Q6: 3,4	0.819
Application	Applying ideas elsewhere, understanding others' ideas	5	3	Q6: 5,6,7	0.629
Learning gains: Affective gains					
Positive attitude	Appreciation of math	5	2	Q8: 3,6	0.821
Confidence	Confidence to do math	5	4	Q8: 1,2,7,8	0.905
Persistence	Persistence, stretching	5	2	Q8: 9,14	0.852
Learning gains: Social gains					
Collaboration	Working with others, seeking help	5	3	Q8: 10,12,13	0.841
Teaching	Comfort in teaching math	5	1	Q8: 11	-
Learning gains: Independence	Work/organize on own	5	2	Q8: 4,5	0.806

We also report results on students' cognitive, affective and social gains as three main areas of learning gains. Gain in independence in learning mathematics represented a measure that is distinct from the other three main areas.

Table A3.6 displays the titles and descriptions of the learning-related composite variables, the survey items that comprise each variable, and the reliability scores for each composite variable.

A3.6.3 Analysis methods for structured survey questions

All survey data was entered by student technicians and analyzed using the SPSS computer software package. Statistical analyses included descriptive statistics of each composite variable and background variable. Correlation analysis was used to study relationships between composite variables and their relation to background information on students' overall college GPA and expected grade at the beginning and end of their course. Parametric (independent and pair-wise T-tests, ANOVA) or non-parametric (Chi-square, Mann-Whitney, Kruskas-Wallis) tests were used to explore group differences in students' attitudes, experiences, and learning gains, sorted by demographic information on students' gender, ethnicity, race, academic status, and college major. The most important of these analyses focused on differences between IBL and non-IBL students, and between math-track students and pre-service teachers. Analysis of covariance (ANCOVA) was used to check intermediate effects (GPA, expected grade, gender) on students' learning gains. Stepwise regression analysis was applied to examine the variation in students' learning gains versus changes in their attitudes and self-reported class experiences.

A3.6.4 Analysis of open-ended survey questions

The open-ended survey questions asked about students' gains or changes in their understanding of mathematics, remembering key ideas, ways to learn mathematics, and other things they carry with themselves from a math course. Most of students' written comments addressed reports of learning gains from a course or possible difficulties or negative experiences from a course. To analyze the written responses, we applied the same categories that were constructed for analyzing student interview data. The preliminary categories were fleshed out with more detailed descriptions and subdivided into several subcategories of learning gains and processes using inductive content analysis (Miles & Huberman, 1994; Strauss & Corbin, 1990). As each statement was examined, the detected gains were classified into one of the preliminary categories or a new category creating during reading and analysis. Table A3.7 summarizes the final coding scheme for learning gains reported in the open-ended answers, and the frequency with which each was reported.

Table A3.7. Counts for Reported Learning Gains from Open-Ended Survey Comments

Main category	**Subcategory**	**Number of students reporting each gain**		
	Description	once	2-3 times	≥4 times
Cognitive gains	**Subtotal**	**455**	**115**	**5**
	Better recall	82	2	-
	Better knowledge, deeper understanding of mathematical concepts and ideas	120	54	-
	Thinking and problem solving skills	119	42	4
	Transferable mathematical knowledge	23	3	-
	Transferable thinking skills	20	1	-
	Did not gain cognitively	91	13	1
Affective gains	**Subtotal**	**158**	**24**	**1**
	Positive attitude towards mathematics	16	1	-
	Confidence to do math, solve math problems, and be a mathematician	38	8	1
	Less confidence, no gain in confidence	5	-	-
	Enjoyment, liking math	28	2	-
	Negative experience, liking less	62	13	-
	Interest and motivation	8	-	-
	Less interested	1	-	-
Changes in learning	**Subtotal**	**428**	**63**	**4**
	Beliefs about learning math, deeper learning, problem solving, creativity and discovery, finding own style of learning math	133	33	3
	Independence in mathematical thinking, learning or problem solving	93	14	1
	Persistence	17	-	-
	Work ethic, learned to study hard	21	-	-
	Metacognition, self-reflection	17	1	-
	Appreciation others' thinking, learning from others	66	14	-
	No change in learning math	81	1	-

Table A3.7, continued…

Gains in communication skills	**Subtotal**	**116**	**9**	**-**
	Speaking or presenting	14	-	-
	Writing mathematics	23	4	-
	Collaboration, group work	29	3	-
	Teaching others, explaining to others	46	2	-
	Giving or receiving critique	4	-	-
Changes in understanding the nature of mathematics	**Subtotal**	**78**	**20**	**-**
	How knowledge is built; how research math is done	14	-	-
	Change in conceptualization of math	59	20	-
	No change in concepts of the nature of math	5	-	-
Total		**1235**	**231**	**10**

In all, 544 students wrote in at least one gain in cognition, affect, communication, ways of learning mathematics, and/or understanding the nature of mathematics. They reported one to as many as nine gains each. In all, 197 students reported 1-6 times each that they did not make gains or undergo changes in cognition, affect, communication, ways of learning mathematics, or understanding the nature of mathematics. The rest of survey respondents wrote no comments.

A3.7 Reliability and validity

Most of the pre- and post-surveys were administered and gathered in mathematics classes by the project coordinators of each campus, or by instructors. The surveys were completed in class, which strengthened the response rate. The coordinators were given written instructions for administering the surveys that were intended to ensure that students had enough time to answer the surveys and that their anonymity was preserved. Completed surveys were delivered to the research team by the campus coordinators, entered by trained project assistants into separate SPSS files, and checked and analyzed by the researchers. We will describe features of our survey data from the attitudinal survey and the learning gains survey separately.

A3.7.1 Attitudinal survey data

After testing the structures and items in the attitudinal survey with a small group of students, we revised the instrument accordingly. In the full data collection, we gathered a large number of completed pre-and post-surveys, which ensured high statistical power in our results. Missing answers both in the structured attitudinal and gain survey items were rare. Even though the post-survey was rather long, most students responded to all the structured items. The number of missing answers on the items varied between 0-49 for the main pre-survey items and 0-37 for the main attitudinal post-survey items. These low numbers indicate that students understood the questions and statements in the items. All these features strengthen the reliability and validity of our attitudinal survey results.

Among the combined survey responses, some students did not report specific demographic information on their:

- gender (49),
- race (148) or ethnicity (74),
- academic major (79),
- academic status (46),
- number of prior college mathematics courses (53)
- AP test score (430; many students did not take the AP test),
- prior GPA (291; many first year students had no prior GPA),
- expected grade at the beginning of a course (86).

The most common type of missing information was self-reported AP test scores and prior GPA. Thus the results reported on group differences by prior GPA largely excluded first-year students. We used AP test scores in our study of group differences only in analyzing LMT scores for pre-service teachers.

The number of responses about students' beliefs on mathematical proofs (N=877) and about confidence (N=672) were somewhat smaller than those on other topics. Only students with prior experience on proofs were asked to answer proof-related survey questions. A subsection on confidence was added to the attitudinal questionnaire after some students had completed it. For both sections, the lower number of answers implies slightly lower statistical power in comparison to the results on other survey questions.

Descriptive statistics for the composite attitudinal variables showed variation among students. Responses for the items varied between the minimum 1 and maximum 7 for most items. The minimum for only one composite variable was above 2 (Reasoning 2.2 on the post-survey). However, standard deviations for the pre-survey variables varied between 0.88 and 2.56, and for the post-survey variables between 0.86 and 2.63, indicating low or moderate variation among students in their attitudes.

Cronbach's alphas were used to study the reliabilities of the subsections and composite variables for both attitudinal and SALG-M survey instruments. The final reliability scores for attitudinal composite variables are presented in Table A3.5, for pre-survey and post-survey data separately. Reliability scores for the pre-survey varied between 0.439 and 0.914 and between 0.615 and 0.928 for the post-survey. Only five composite variables had a low reliability score (below 0.7) in the pre-/post-survey data: math future interest, teaching confidence, instructor-driven beliefs, and beliefs about proofs.

Correlational analysis showed that the composite variables had good construct validity, meaning they produced real results on students' motivation, beliefs and strategies. For example, on the post-survey data, both composite variables related to motivation (personal interest, math future) correlated highly positively with each other (r= 0.417) but also with the variables on intrinsic learning goals (r=0.552, r=0.354) and enjoyment of learning mathematics (r=0.718, r=0.413). Confidence in mathematics (r=0.507) and teaching (r=0.309) correlated highly positively with enjoyment of mathematics. Math confidence also correlated clearly with the use of independent strategies (r=0.445) and beliefs about math problem-solving as reasoning (r=0.531).

Beliefs that mathematics learning is an instructor-driven activity correlated highly positively with extrinsic learning goals (r=0.454), beliefs about problem-solving as practice (r=0.499), and beliefs about proving as confirming the truth (r=0.359). In contrast, beliefs about math problem-solving as reasoning correlated highly positively with enjoyment of mathematics (r=0.512) and intrinsic learning goal (r=0.539). Moreover, use of independent strategies correlated highly positively with use of self-regulatory strategies (r=0.606), whereas use of collaboration correlated positively with communicating goal (r=0.289) and beliefs about mathematics learning as group work (r=0.459) or exchange of ideas with other students (r=0.422).

A3.7.2 Learning gains survey data

We collected even larger numbers of responses on the SALG-M items (N=1074-1127) than on the attitudinal pre-/post-survey measures. This ensured high statistical power for results on students' self-reported class experiences and learning gains. The numbers of analyzed responses on the SALG-M items was lower than in the attitudinal data because students could choose the response "Not applicable." Open-ended survey questions on the SALG-M were optional, and students commonly chose not to respond to these. Overall, 13 to 38% of IBL math-track students and 12 to 20% of non-IBL math-track students chose to write in a comment about their learning gains (depending on the question). These low percentages are typical for open-ended survey questions.

The whole scale between 1 and 5 was used by students in their answers to questions about both course experiences and learning gains. However, descriptive statistics showed a rather strong "halo" effect. The median for all the course experience variables (except Assignments other than tests) was above 3.8, and for the learning gains variables above 3.5 (except Application). The standard deviations ranged between 0.82 and 0.95 for the course experience variables and between 0.93 and 1.16 for the learning gains variables. These scores indicate rather low variation in students' answers to the survey items.

The reliability scores for the composite variables from the SALG-M instrument are presented in Table A3.6. The scores indicated even higher reliability and internal consistency than for the attitudinal variables. For the course experience composite variables, the reliability scores varied between 0.696 and 0.898. Reliability scores for the composite learning gains variables similarly varied between 0.629 and 0.921. Only one composite variable (gains in application) had a reliability score less than 0.7.

The SALG survey instruments are based on extensive research that support the validity of self-report in situations where students have the ability to provide accurate information (Wentland & Smith, 1993) and where they have few or no obvious reasons (such as adverse consequences or social embarrassment) for providing inaccurate information (Aaker, Kumar & Day, 1998). The SALG survey instruments meet standards of good validation. The developers of the instrument confirmed that most survey items functioned adequately and that item composites formed reliable subscales (Weston, Seymour & Thiry, 2006; Weston, Seymour, Lottridge & Thiry, 2006). Moreover, latent factors underlying items conformed to the hypothesized structures of the survey.

We adjusted the original SALG survey instrument to fit college mathematics learning situations and refer to the result as the SALG-M. The constructed composite variables represented the underlying structures of the SALG instrument and factors specific for the SALG-M survey. These factors were checked for construct validity by correlational analysis. For example, Active Participation, a variable denoting participation in class discussions and group work, was most strongly (positively) related (r=0.634) to the measure of Personal Interaction (see Table A3.6). In turn, Individual Work, measuring students' studying and problem solving on their own, was the most strongly positively related to Assignment (r=0.557), in particular to Assignments other than regular tests (r=0.529). Moreover, the Overall measure of course experience correlated highly positively (0.505-0.680) with all the other composite variables on course experiences.

All students reported rather positive course experiences and high learning gains as result of a course. Correlational analysis further showed strong positive linkages between the learning gains variables (0.403-0.740) and between course experiences and learning gains (0.314-0.680). These results confirm construct validity of the results and the fact that positive class experiences were related to higher reported learning gains. On the other hand, students who reported positive experience in one area also did so on the other measured class experiences. Moreover, students who reported high learning gains in one area also did so on other measures of learning gains. This again reflects the halo effect that makes it more difficult to find real differences in students' class experiences and learning gains.

A3.7.3 Connections between the surveys and other measures of learning gains

We checked connections between the composite variables based on the structured SALG-M survey questions and numerical variables related to the open-ended answers. Correlations between the answers to structured and open-ended questions showed that higher self-assessed cognitive, affective and social learning gains were all clearly positively related to a higher number of gains each student wrote in. This was valid both for the total count of gains reported in written comments (r=0.133 to 0.239) and for the separate counts of cognitive gains, affective gains and changes in students' ways of learning math. In particular, higher self-assessed gains in math concepts and thinking were clearly positively connected (r=0.209**) to a greater number of cognitive gains written in response to the open-ended questions. Moreover, higher self-assessed gains in collaboration were clearly positively connected (r=0.209**) to a greater number of gains in ways of learning mathematics as reflected in written comments. These indicate reliability and construct validity of the results on learning gains.

We also checked how changes in attitudes related to learning gains as measured by the SALG-M instrument. The correlations (see Table A3.8) show that students who reported higher cognitive, affective, and social learning gains also showed increases in many of the attitudinal variables. This applied particularly to enhanced motivation, enjoyment, math confidence, and intrinsic and communicating goals during a math course. Pre/post increases both in enjoyment and math confidence were clearly positively related to reported affective gains in confidence, positive attitude, and persistence. This displays construct validity of these affective measures.

Table A3.8: Statistically Significant Correlations between Changes in Beliefs, Motivation and Strategies and Learning Gains

Attitudinal Variable		Correlation with Learning Gains			
		Math concepts & thinking	Application	Affective	Social
Motivation	*interest*	0.186**	0.140**	0.199**	0.143**
	math major	0.149**	0.125**	0.158**	
	math future	0.115**		0.107**	
	teaching	0.118**		0.105**	
Goals	*intrinsic*	0.170**	0.104**	0.182**	0.156**
	communicating	0.121**		0.132**	
Enjoyment		0.228**	0.204**	0.266**	0.194**
Confidence	*math ability*	0.249**	0.233**	0.245**	0.166**
Beliefs about learning	*group work*	0.142**	0.121**	0.145**	0.195**
	exchange of ideas	0.154**	0.150**	0.166**	0.144**
Beliefs about problem-solving	*reasoning*	0.205**	0.155*	0.186**	0.189**
	practice	0.205**	0.155**	0.186**	0.189**
Strategies	*independent*	0.134**	0.129**	0.121**	0.164**
	collaborative				0.188**
	self-regulatory	0.204**	0.138**	0.204**	0.183**

*** $p < 0.01$*

Similarly, gains in mathematical concepts and thinking were positively related to increases in most of the attitudinal variables. The strongest positive relation was to increased belief in reasoning in solving math problems, but also to enhanced enjoyment and math confidence and to increased use of self-regulatory strategies. Positive relationships between reported gains in application and attitudinal variables showed similar but somewhat weaker correlations. All these positive relations display good construct validity for the variables involved.

Moreover, students with higher reported learning gains developed strengthened beliefs in the value of group work and exchange of ideas with other students that generally contribute to learning. Increased belief in reasoning as a way to solve math problems again enhances mathematics learning and problem solving. Moreover, students' increased use of independent and self-regulatory strategies in learning were clearly positively related to their learning gains. The use of these strategies generally enhances learning. In particular, gains in collaboration were positively related to increased use of collaborative learning strategies.

Correlational analysis (Spearman) between the learning gains composite variables and other measures of learning outcomes indicated low to moderate connections. Self-reported cognitive, affective and social learning gains from a mathematics course did not correlate with self-reported GPA level at the beginning of a course. But grades generally measure student *performance*, which is not necessarily related to *learning* in any given course.

In addition, our GPA measure excluded first-year students who could not report their prior GPA at the beginning of a course but who nonetheless reported higher gains than older students. Furthermore, our results on learning gains indicated higher learning gains among students with lower prior GPA. These features are also perhaps reflected in the low correlations between learning gains and GPA level at the beginning of a course.

We also checked the correlations (Spearman) between self-reported learning gains and self-reported AP test score, for the smaller number of students who reported an AP score. However, the correlations indicated only a weak positive relation to gains in math concepts and thinking (r=0.106*). Again, AP test score is a measure of past mathematics performance, but does not determine learning in the present course.

However, correlations between learning gains and expected grade in the course were somewhat stronger, especially to the expected grade reported at the end of the course. These correlations varied between 0.10** and 0.307**. In particular, gains in mathematical concepts and thinking clearly correlated with expected grade at the end of a course (r=0.238**). The positive correlation of expected grade was even stronger to affective learning gains (r=0.307**) but weaker to social gains in collaboration (r=0.10**). These correlations display clear but moderate connections between students' self-reported learning gains and their assessment of the quality of their learning during a college math course.

A3.8 References Cited

Aaker, D. A., Kumar, V., & Day, G. S. (1998). *Marketing research*. New York: Wiley.

Ames, G., & Archer, J. (1988). Achievement goals in the classroom: Students´ learning strategies and motivation processes. *Journal of Educational Psychology*, *80*, 260-267.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28,* 117-148.

Buch, N. J., & Wolff, T. F. (2000). Classroom teaching through inquiry. *Journal of Professional Issues in Engineering Education and Practice*, (July),105-109.

Burn, B., Appleby, J., & Maher, P. (Eds.) (1998). *Teaching undergraduate mathematics*. London: Imperial College Press.

De Corte, E., Verschaffel, L., & Eynde, P. O. (2000). Self-regulation: A characteristic and a goal of mathematics education. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 687-726). San Diego: Academic Press.

Duch, B. J., Groh, S. E., & Allen, D. E. (2001). Why problem-based learning? In B. J. Duch, S. E. Groh, & D. E. Allen (Eds.), *The power of problem-based learning* (pp. 3-11). Sterling, VA: Stylus Publishing.

Entwistle, N. (1988). Motivational factors in students' approaches to learning. In R. R. Schmeck (Ed.), *Learning strategies and learning styles* (pp. 21-51). New York: Plenum.

Frost, L. A., Hyde, J. S., and Fennema, E. (1994). Gender, mathematics performance, and mathematics-related attitudes and affect: A meta-analytic synthesis. *International Journal of Educational Research, 21*, 373-386.

Gillies, R. M. (2007). *Cooperative learning: Integrating theory and practice*. Thousand Oaks, CA: Sage Publications.

Goldin, G.A. (2000). Affective pathways and representations in mathematical problem solving. *Mathematical Thinking and Learning, 17,* 209-219.

Hanna, G. (1991). Mathematical proof. In D. Tall (Ed.), *Advanced mathematical thinking* (pp. 54-61). Norwell, MA: Kluwer.

Ju, M.-K. & Kwon, O.N. (2007). Ways of talking and ways of positioning: Students' beliefs in an inquiry-oriented differential equation class. *Journal of Mathematical Behavior*, *26*, 267-280.

Knuth, E. J. (2002). Secondary school mathematics teachers' conceptions of proof. *Journal for Research in Mathematics Education,* 33, 379-405.

Kroll, D. L., & Miller, T. (1993). Insights from research on mathematical problem solving in the middle grades. In D. T. Owens (Ed.), *Research ideas for the classroom: Middle grades mathematics* (pp. 58-77). NY: Macmillan.

Leder, G. C., Pehkonen, E., & Torner, G. (Eds.). (2002). *Beliefs: A hidden variable in mathematics education?* Dordrecht: Kluwer.

Malmivuori, M. L. (2001). *The dynamics of affect, cognition, and social environment in the regulation of personal learning processes: The case of mathematics.* Research report 172. Helsinki, Finland: Helsinki University Press.

Malmivuori, M.-L. (2007). Understanding student affect in learning mathematics. In C. L. Petroselli (Ed.), *Science Education: Issues and Developments* (pp. 125-149). New York: Nova Science.

Marton, F., Hounsell, D., & Entwistle, N. (1997). *The experience of learning.* Edinburgh: Scottish Academic Press.

McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning*. 575-596. New York: Macmillan.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis.* California: Sage.

Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientations in learning and achievement. *Journal of Educational Psychology, 92*, 544-555.

Pollatsek, H., Barker, W., Bressoud, D., Epp, S., Ganter, S., & Haver, B. (2004). *Undergraduate Programs and Courses in the Mathematical Sciences: CUPM Curriculum Guide 2004*. A report by the Committee on the Undergraduate Program in Mathematics of The Mathematical Association of America. MAA, Washington DC: The Mathematical Association of America. www.maa.org/cupm/

Prince, M., & Felder, R. (2007). The many facets of inductive teaching and learning. *Journal of* College Science Teaching, *36*(5), 14-20.

Rasmussen, C., & Kwon, O. (2007). An inquiry oriented approach to undergraduate mathematics. *Journal of Mathematical Behavior, 26*, 189-194

Recommendations for using the SALG (2008). Recommendations for using SALG. Accessed 4/25/2011 at http://www.salgsite.org/docs/SALG_Rationale_11_15_08.doc

Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist, 26*(3), 299-323.

Schoenfeld, A. (1985). *Mathematical problem solving*. New York: Academic Press.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334-370). New York: Macmillan.

Seegers, G., & Boekaerts, M. (1996). Gender-related differences in self-referenced cognitions in relation to mathematics. *Journal for Research in Mathematics Education, 27*, 187-215.

Selden, A., & Selden, J. (2007). *Overcoming students' difficulties in learning to understand and construct proofs.* Technical Report 1. Cookeville, TN: Tennessee Technological University.

Seymour, E., Wiese, D., Hunter, A., & Daffinrud, S. M. (2000). *Using real-world questions to promote active learning.* National Meeting of the American Chemical Society Symposium, March 27, 2000, San Francisco, CA.

Smith, J. C. (2006). A sense making approach to proof: Strategies of students in traditional and problem-based number theory courses. *Journal of Mathematical Behavior*, *25*, 73-90.

Snow, R. E., Corno, L., & Jackson, D. (1996). Individual differences in affective and conative functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 243-292). New York: MacMillan.

Sowder, L., & Harel, G. (2003). Case studies of mathematics majors' proof understanding, production, and appreciation. *Canadian Journal of Science, Mathematics and Technology Education*, *3*(2), 251-267.

Steffe, L. P., & Thompson, P. W. (Eds.)(2000). *Radical constructivism in action: Building on the pioneering work of Ernst von Glasersfeld.* Studies in Mathematics Education Series, 15. London: Routledge Falmer.

Stipek, D. (2002). Good instruction is motivating. In A. Wigfield & J. S. Eccles (Eds.). *Development of achievement motivation* (pp. 309-332). San Diego: Academic Press.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage Publications.

Tall, D. (1991). The psychology of advanced mathematical thinking. In D. Tall (Ed.), *Advanced mathematical thinking* (pp. 3-21). Norwell, MA: Kluwer.

Wentland, E. J., & Smith, K. W. (1993). *Survey responses: An evaluation of their validity*. New York: Academic Press.

Weston, T., Seymour, E., Lottridge, S., & Thiry, H. (2006). The validity of SENCER-SALG online survey for course revision and assessment of student outcomes. Unpublished manuscript.

Weston, T., Seymour, E., & Thiry, H. (2006). *Evaluation of Science Education for New Civic Engagements and Responsibilities (SENCER) project*. [Report to SENCER/NCSCE] Boulder, CO: ATLAS.

Yoo, S., & Smith, J. C. (2007). Differences between mathematics majors' views of mathematical proof after lecture-based and problem-based instruction. In Lamberg, T., & Wiest, L. R. (Eds.). *Proceedings of the 29th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 596-598). Stateline (Lake Tahoe), NV: University of Nevada, Reno.

Zimmerman, B. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology, 25*, 82-91.

Exhibit E3.1: Attitudinal Pre-Survey

Dear student,

Our research team is studying methods of improving teaching and learning in college mathematics courses, including the methods used in the course you are taking now. Because you are enrolled in a college math course, we would like to know about your own experiences in learning mathematics. This survey asks about your views about mathematics, your strategies for learning math, and your personal reasons for studying mathematics.

Your participation is voluntary. You may skip questions you do not wish to answer, or choose not to participate. Your answers are anonymous and will not be reported in any way that may identify you individually; they will be aggregated with responses by other students from your course and other courses. Your instructor will not know how you answered.

By completing this survey, in part or in whole, you agree that we may use this data to understand and improve the quality and effectiveness of mathematics instruction.

Please, mark clearly the best answer to each question. You do NOT need to fill in the bubble completely.

Thank you for your candid responses! Please contact us with any questions.

Sandra Laursen, study director
Marja-Liisa Hassi, research associate

Ethnography & Evaluation Research
University of Colorado at Boulder
sandra.laursen@colorado.edu
hassi@colorado.edu

Your interest in mathematics

1. HOW LIKELY is it that you will...

	Not at all likely						Extremely likely
Take additional math courses after this course?	○	○	○	○	○	○	○
Graduate with a college math major?	○	○	○	○	○	○	○
Graduate with a college math minor?	○	○	○	○	○	○	○
Study hard for a college math course?	○	○	○	○	○	○	○
Read magazine or newspaper articles related to math?	○	○	○	○	○	○	○
Bring up mathematical ideas in a non-mathematical conversation?	○	○	○	○	○	○	○
Participate in a club or organization related to math?	○	○	○	○	○	○	○
Teach math in the future?	○	○	○	○	○	○	○

Exhibit E3.1: Attitudinal Pre-Survey

Your enjoyment of mathematics

2. HOW MUCH do you ENJOY...

	No enjoyment						Extreme enjoyment
Working on a challenging mathematical problem?	○	○	○	○	○	○	○
Discovering a new mathematical idea?	○	○	○	○	○	○	○
Seeing mathematics in everyday life?	○	○	○	○	○	○	○
Perceiving beauty in mathematical ideas?	○	○	○	○	○	○	○
Using rigorous reasoning in a math problem?	○	○	○	○	○	○	○
Thinking about abstract concepts?	○	○	○	○	○	○	○
Teaching mathematics to other people?	○	○	○	○	○	○	○

Your goals in studying mathematics

3. Below are some goals that students may have in studying mathematics. HOW IMPORTANT is each goal for YOU?

	Not at all important						Extremely important
Learning specific procedures for solving math problems	○	○	○	○	○	○	○
Improving your ability to communicate mathematical ideas to others	○	○	○	○	○	○	○
Getting a good grade in college mathematics courses	○	○	○	○	○	○	○
Memorizing the sets of facts important for doing mathematics	○	○	○	○	○	○	○
Making mathematics understandable for other people	○	○	○	○	○	○	○
Meeting the requirements for your degree	○	○	○	○	○	○	○
Learning to construct convincing mathematical arguments	○	○	○	○	○	○	○
Using mathematics as a tool to study other fields	○	○	○	○	○	○	○
Learning new ways of thinking	○	○	○	○	○	○	○
Applying mathematical thinking outside the university context	○	○	○	○	○	○	○

Other goals (please specify)

Exhibit E3.1: Attitudinal Pre-Survey

Your strategies for learning mathematics

4. When you DO MATH, how often do you take each action listed below?

	Very seldom						Very often
Study on your own.	○	○	○	○	○	○	○
Brainstorm with other students.	○	○	○	○	○	○	○
Try to organize or summarize your own ideas.	○	○	○	○	○	○	○
Share problem-solving strategies with other students.	○	○	○	○	○	○	○
Find your own ways of thinking and understanding.	○	○	○	○	○	○	○
Review your work for mistakes or misconceptions.	○	○	○	○	○	○	○
Read the assigned readings.	○	○	○	○	○	○	○
Plan a solving strategy before attacking a problem.	○	○	○	○	○	○	○
Try to find your own way to solve a problem.	○	○	○	○	○	○	○
Check your understanding of what the problem is asking.	○	○	○	○	○	○	○
Use your intuition about what the answer should be.	○	○	○	○	○	○	○
Look for an alternate strategy to solve the problem.	○	○	○	○	○	○	○
Give up when you get stuck.	○	○	○	○	○	○	○
Ask another student for help.	○	○	○	○	○	○	○
Ask the instructor or TA for help.	○	○	○	○	○	○	○

Your preferences for learning mathematics

5. Indicate how much you agree or disagree:
I learn mathematics BEST when...

	Strongly disagree						Strongly agree
The instructor lectures.	○	○	○	○	○	○	○
The class critiques other students' solutions.	○	○	○	○	○	○	○
I work on problems in a small group.	○	○	○	○	○	○	○
The exams let me prove my mathematical skills.	○	○	○	○	○	○	○
Groups present their solutions in class.	○	○	○	○	○	○	○
The instructor explains the solutions to problems.	○	○	○	○	○	○	○
The homework assignments are similar to the examples considered in class.	○	○	○	○	○	○	○
I study my class notes.	○	○	○	○	○	○	○
I can compare my math knowledge with other students.	○	○	○	○	○	○	○
I explain ideas to other students.	○	○	○	○	○	○	○
I get frequent feedback on my mathematical thinking.	○	○	○	○	○	○	○

Exhibit E3.1: Attitudinal Pre-Survey

6. Indicate how much you agree or disagree:
In order to solve a challenging math problem, I NEED...

	Not at all						Very much
To carefully analyze different possible solutions.	○	○	○	○	○	○	○
To have lots of practice in solving similar problems.	○	○	○	○	○	○	○
To understand other students' mathematical thinking.	○	○	○	○	○	○	○
To have natural talent for mathematics.	○	○	○	○	○	○	○
To try multiple approaches to constructing a solution.	○	○	○	○	○	○	○
To remember a lot of examples that I might use in constructing a solution.	○	○	○	○	○	○	○
To use rigorous reasoning.	○	○	○	○	○	○	○
To have freedom to do the problem in my own way.	○	○	○	○	○	○	○
To work hard	○	○	○	○	○	○	○

Your experience and views about mathematical proof

7. Have you had math classes that included mathematical proofs?

○ yes

○ no (Go directly to question 9)

8. The following statements reflect some students' views about mathematical proof. How much do you AGREE or DISAGREE with each statement?

	Strongly disagree						Strongly agree
The main purpose of proof is to confirm the truth of a mathematical result that is already known to be true.	○	○	○	○	○	○	○
Proof is a tool for understanding mathematical ideas.	○	○	○	○	○	○	○
Doing proofs well requires good recall of previous proofs of similar statements.	○	○	○	○	○	○	○
The main purpose of proof is to explain why a certain statement is true.	○	○	○	○	○	○	○
In math class, doing proofs means confirming conjectures that have been previously proven by an expert.	○	○	○	○	○	○	○
There are several different ways to prove a mathematical statement.	○	○	○	○	○	○	○
When evaluating a proof, the most important thing to look at is its logical structure.	○	○	○	○	○	○	○
A proof is something you have to construct based on your own understanding.	○	○	○	○	○	○	○

Exhibit E3.1: Attitudinal Pre-Survey

Your confidence in doing math

9. HOW CONFIDENT are you that you can...

	Not at all confident						Extremely confident
Get a high grade in this course?	○	○	○	○	○	○	○
Successfully work with complex mathematical ideas?	○	○	○	○	○	○	○
Teach mathematics to high school students?	○	○	○	○	○	○	○
Develop new mathematical ideas?	○	○	○	○	○	○	○
Apply a variety of perspectives in solving problems?	○	○	○	○	○	○	○
Present your work at the board in a math class?	○	○	○	○	○	○	○
Work on math problems with other students?	○	○	○	○	○	○	○
Teach math to children?	○	○	○	○	○	○	○

Your math background

10. What was the highest level of math that you took in HIGH SCHOOL?

- ○ Algebra, one year
- ○ Algebra, two years
- ○ Geometry with an algebra prerequisite
- ○ Pre-calculus or trigonometry
- ○ Calculus
- ○ Other (please specify)

11. Did you take the AP Calculus test?

- ○ Yes
- ○ No (go directly to question 14)

12. Which of the AP Calculus tests did your take?

- ○ A/B
- ○ B/C

13. What was your score in the AP Calculus test?

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5

Exhibit E3.1: Attitudinal Pre-Survey

14. How many COLLEGE math courses have you taken prior to this course? Please count the total number of semesters or quarters.

- ◯ 0
- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5
- ◯ 6
- ◯ 7 or more

15. What grade do you expect to receive in this course?

- ◯ A
- ◯ A-
- ◯ B+
- ◯ B
- ◯ B-
- ◯ C+
- ◯ C
- ◯ C-
- ◯ D
- ◯ F

Your academic background

16. What is your overall UNDERGRADUATE GPA? (estimated)

- ◯ 3.8 or higher
- ◯ 3.5 - 3.79
- ◯ 3.0 - 3.49
- ◯ 2.5 - 2.99
- ◯ 2.0 - 2.49
- ◯ below 2.0
- ◯ Not applicable

17. What is your class year?

- ◯ First-year
- ◯ Sophomore
- ◯ Junior
- ◯ Senior
- ◯ Graduate student
- ◯ Other (please specify)

Exhibit E3.1: Attitudinal Pre-Survey

Your academic interests

18. What is your college major? (Check ALL that apply)

- [ ] Math or Applied Math
- [ ] Physics
- [ ] Chemistry
- [ ] Engineering
- [ ] Computer science
- [ ] Other science or technical field
- [ ] Economics
- [ ] Other non-science field

19. Are you pursuing a teaching certification?

- ○ no
- ○ yes, elementary (grades K-6 or K-8)
- ○ yes, secondary math (grades 6-12, 8-12, or 9-12)
- ○ yes, secondary in a field other than math

Other (please specify)

Your personal background

Our funding agency requires us to gather data on the gender, race and ethnicity of study participants. Please choose the answers that best apply.

20. What is your gender?

- ○ male
- ○ female

21. What is your ethnicity?

- ○ Hispanic or Latino
- ○ Not Hispanic or Latino

22. What is your race? (please check ALL that apply)

- [ ] American Indian or Alaskan Native
- [ ] Asian
- [ ] Black or African American
- [ ] Native Hawaiian or other Pacific Islander
- [ ] White

Exhibit E3.1: Attitudinal Pre-Survey

Assign yourself an identifier

On this page, we ask for some information that will enable us to match your survey responses with those in other surveys. The information will be unique to you but will not identify you individually.

* 23. Enter the following data. Please, print neatly.

FIRST two letters of your FIRST NAME

Two-digit DAY of your BIRTHDAY (01 through 31)

FIRST two letters of your MOTHER'S FIRST NAME

FIRST two letters of the TOWN where you were BORN

Course information

* 24. What is this math course?

Number of the course

Section of the course

Name of the instructor

Survey completed

Thank you for completing the survey! Your input is important to us, and will help us to help math instructors improve teaching and learning in their courses.

If you have any questions, please contact us:

Sandra Laursen, project director
sandra.laursen@colorado.edu
Marja-Liisa Hassi, research associate
hassi@colorado.edu

Exhibit E3.2: Learning Gains Post-Survey (SALG-M)

Dear student,

Our research team is studying methods of improving teaching and learning in college mathematics courses, including the methods used in the course you are taking now. Because you are enrolled in a college math course, we would like to know about your own experiences in learning mathematics. This survey asks about your experiences in this course.

Your participation is voluntary. You may skip questions you do not wish to answer, or choose not to participate. Your answers are anonymous and will not be reported in any way that may identify you individually; they will be aggregated with responses by other students from your course and other courses. Your instructor will not know how you answered.

By completing this survey, in part or in whole, you agree that we may use this data to understand and improve the quality and effectiveness of mathematics instruction. We may compare your responses with your gains from the course, assessed by your instructor. This will be done anonymously by using the identifiers. We will not know your individual grades and your instructor will not know how you answered the questions in this survey.

Please, mark clearly the best answer to each question. You do NOT need to fill in the bubble completely.

Thank you for your candid responses! Please contact us with any questions.

Sandra Laursen, study director
Marja-Liisa Hassi, research associate

Ethnography & Evaluation Research/University of Colorado at Boulder
sandra.laursen@colorado.edu
hassi@colorado.edu

The course as a whole

1. HOW MUCH did the following aspects of the class HELP YOUR LEARNING?

	No help	A little help	Moderate help	Much help	Great help	NOT APPLICABLE
The overall approach to teaching and learning in the course	○	○	○	○	○	○
How class topics, activities, & assignments fit together	○	○	○	○	○	○
The pace of the class	○	○	○	○	○	○
The workload of the class	○	○	○	○	○	○
The general atmosphere of the class	○	○	○	○	○	○
The course material	○	○	○	○	○	○
The mental stretch required of you	○	○	○	○	○	○
The information you were given about the class when it began	○	○	○	○	○	○

Other (please specify)

Exhibit E3.2: Learning Gains Post-Survey (SALG-M)

Class activities

2. HOW MUCH did the following CLASS activities HELP YOUR LEARNING?

	No help	A little help	Moderate help	Much help	Great help	DID NOT HAPPEN
Listening to lectures	○	○	○	○	○	○
Studying on your own	○	○	○	○	○	○
Participating in class discussions	○	○	○	○	○	○
Participating in group work during class	○	○	○	○	○	○
Explaining your work to other students	○	○	○	○	○	○
Hearing other students explain their work	○	○	○	○	○	○
Giving presentations in front of class	○	○	○	○	○	○
Writing solutions to problems	○	○	○	○	○	○
Checking solutions to problems	○	○	○	○	○	○
Working on a computer	○	○	○	○	○	○
Examining children's mathematical work	○	○	○	○	○	○

3. Please comment on how this class has CHANGED THE WAYS YOU LEARN mathematics?

5

6

Assignments and tests

4. HOW MUCH did the assignments and tests HELP YOUR LEARNING?

	No help	A little help	Moderate help	Much help	Great help	DID NOT HAPPEN
Taking tests	○	○	○	○	○	○
Doing other assignments	○	○	○	○	○	○
Doing homework	○	○	○	○	○	○
The fit between class content and tests	○	○	○	○	○	○
The match between the grading system and what you needed to work on	○	○	○	○	○	○
The mental stretch required on tests	○	○	○	○	○	○
Preparing class presentations	○	○	○	○	○	○
The feedback you received on your written work	○	○	○	○	○	○

Exhibit E3.2: Learning Gains Post-Survey (SALG-M)

Support for you as a learner

5. HOW MUCH did each of the following HELP YOUR LEARNING?

	No help	A little help	Moderate help	Much help	Great help	DID NOT HAPPEN
Interacting with the instructor DURING class	○	○	○	○	○	○
Interacting with the instructor OUTSIDE class	○	○	○	○	○	○
Interacting with teaching assistants DURING class	○	○	○	○	○	○
Interacting with teaching assistants OUTSIDE class	○	○	○	○	○	○
Working with peers DURING class	○	○	○	○	○	○
Working with peers OUTSIDE class	○	○	○	○	○	○

Your understanding of class content

6. As a result of your work in this class, what GAINS did you make in your UNDERSTANDING of each of the following?

	No gain	A little gain	Moderate gain	Good gain	Great gain	NOT APPLICABLE
The main concepts explored in this class	○	○	○	○	○	○
The relationships among the main concepts	○	○	○	○	○	○
Your own ways of mathematical thinking	○	○	○	○	○	○
How mathematicians think and work	○	○	○	○	○	○
How ideas from this class relate to ideas outside mathematics	○	○	○	○	○	○
How children solve mathematical problems	○	○	○	○	○	○
How to make mathematics understandable for other people	○	○	○	○	○	○

Please comment on how YOUR UNDERSTANDING OF MATHEMATICS has changed as a result of this class.

Exhibit E3.2: Learning Gains Post-Survey (SALG-M)

7. Please comment on how THE WAY THIS CLASS WAS TAUGHT affects your ability to REMEMBER key ideas.

Confidence, attitudes and abilities

8. As a result of your work in this class, what GAINS did you make in the following?

	No gain	A little gain	Moderate gain	Good gain	Great gain	NOT APPLICABLE
Confidence that you can do mathematics	○	○	○	○	○	○
Comfort in working with complex mathematical ideas	○	○	○	○	○	○
Development of a positive attitude about learning mathematics	○	○	○	○	○	○
Ability to work on your own	○	○	○	○	○	○
Ability to organize your work and time	○	○	○	○	○	○
Appreciation of mathematical thinking	○	○	○	○	○	○
Comfort in communicating about mathematics	○	○	○	○	○	○
Confidence that you will remember what you have learned in this class	○	○	○	○	○	○
Persistence in solving problems	○	○	○	○	○	○
Willingness to seek help from others	○	○	○	○	○	○
Comfort in teaching mathematics	○	○	○	○	○	○
Ability to work well with others	○	○	○	○	○	○
Appreciation of different perspectives	○	○	○	○	○	○
Ability to stretch your own mathematical capacity	○	○	○	○	○	○

9. What will you CARRY WITH YOU from this class into other classes or other aspects of your life?

Exhibit E3.2: Learning Gains Post-Survey (SALG-M)

Your expectation

10. What grade do you expect to receive in this course?

- ◯ A
- ◯ A-
- ◯ B+
- ◯ B
- ◯ B-
- ◯ C+
- ◯ C
- ◯ C-
- ◯ D
- ◯ F

Your background

11. What is your college major? (Check ALL that apply)

- ☐ Math or Applied Math
- ☐ Physics
- ☐ Chemistry
- ☐ Engineering
- ☐ Computer science
- ☐ Other science or technical field
- ☐ Economics
- ☐ Other non-science field

12. Are you pursuing a teaching certification?

- ◯ no
- ◯ yes, elementary (grades K-6 or K-8)
- ◯ yes, secondary math (grades 6-12, 8-12, or 9-12)
- ◯ yes, secondary in a field other than math

Other (please specify)

13. What is your gender?

- ◯ male
- ◯ female

Exhibit E3.2: Learning Gains Post-Survey (SALG-M)

14. What is your class year?

- ○ First-year
- ○ Sophomore
- ○ Junior
- ○ Senior
- ○ Graduate student
- ○ Other (please specify)

Assign yourself an identifier

On this page, we ask for some information that will enable us to match your survey responses at the beginning and end of your math classes. The information will be unique to you but will not identify you individually.

*** 15. Enter the following data. Please, print neatly.**

FIRST two letters of your FIRST NAME

Two-digit DAY of your BIRTHDAY (01 through 31)

FIRST two letters of your MOTHER'S FIRST NAME

FIRST two letters of TOWN where you were BORN

Course information

*** 16. What is this math course?**

Number of the course

Section of the course

Name of the instructor

Survey completed

Thank you for completing the survey! Your input is important to us, and will help us to help math instructors improve teaching and learning in their courses.

If you have any questions, please contact us:

Sandra Laursen, project director
sandra.laursen@colorado.edu
Marja-Liisa Hassi, research associate
hassi@colorado.edu

Appendix A4
Study Methods for Connecting Observation Data and Student Survey Data

A4.1 Introduction

Chapter 4 was intended to explore direct connections between the classroom observation data and the student survey outcomes, addressing the research question:

- How do student learning outcomes relate to the nature of the instruction they experience?

In other chapters, we have used the labels "IBL" and "non-IBL" to gloss the broadly different approaches used in particular classes, comparing student groups by these campus-designated labels. However, these labels hide a good bit of variation (see Section 2.3). In Chapter 4 we attempted to directly link the data from the classroom observations to student gains reported on the SALG-M post-survey. There we discussed the principles involved in linking the observation data to the student self-reported gains from the surveys. Here, we record the specific procedures used in sufficient detail that they could be reproduced in future studies.

A4.2 Study Sample

This analysis relies on two types of data—student post-course surveys and classroom observations—that must be collected from the same course sections in the same term. We collected observation data for 43 course sections at three campuses. The details of the observation study sample are available in Chapter 2 (Section 2.1) and Appendix A2. Unfortunately, we did not obtain post-surveys from seven of the sections included in this sample. The details of the survey study sample can be found in Chapter 3 (Section 3.1) and Appendix A3. Likewise, we could not gather observation data from all sections that returned student post-surveys. The final data set for the combined observation/survey analysis includes 30 IBL and six non-IBL sections, with averages representing 670 students. The Hierarchical Linear Model for full survey data set (see Section A4.5.2) includes 1239 student survey responses from 80 course sections.

A4.3 Data Collection

The data collection methods for the observation study are discussed in detail in Chapter 2 (Section 2.1) and Appendix A2. The data collection methods for the survey study are detailed in Chapter 3 (Section 3.1) and Appendix A3. The observation data and the survey gains data pertain to two different units of analysis: the classroom observation data describe course sections, and the survey data describe individual students. To address this mismatch, we computed student gain averages for each course section. In the next section, we detail how the specific observation data composite variables were constructed and labeled.

A4.4 Construction of Variables

We constructed several composite variables for the purpose of identifying student-centered approaches in classroom activities and practices.

A4.4.1 Classroom Time Variables

These are based on observers' minute-by-minute records of what activities were conducted during class and who led each activity, averaged over several hours of observation of multiple class sessions.

- Total percentage of student-centered time – indicates total percentage of class time spent on student-centered activities. This variable combines the percentage of class time spent on student presentation, percentage of time spent on discussion, percentage of class time spent on group work, and percentage of class time spent on computer-assisted learning.
- Total percentage of student-led time – indicates total percentage of class time spent with an individual student or a group of students in the leadership role. This variable combines the percentage of time with an individual student in the leadership role, percentage of class time with a group of students leading the class, and the percentage of time with the entire class in the leadership position.

A4.4.2 Observer Survey Ratings Variables

Observer ratings variables are based on the observer survey ratings of the frequency of specific student and instructor behaviors discussed in detail in Chapter 2 (Section 2.2.7) and Appendix A2. We constructed four composite variables from the observer survey items based on the four clusters of items (factors) that emerged from the Exploratory Factor Analysis. The resulting composite variables use mean observer ratings for these items as reported for several separately observed class sections; they are discussed in detail below.

- Student-instructor interaction: composite index that averages the ratings for seven items from the observer survey, indicating the frequency with which:
    - students offer ideas during class
    - students receive personal feedback on their work
    - instructors listen to students' ideas
    - instructors give concrete feedback on students' work
    - students ask questions
    - instructor offer help to students
    - inverse rating for instructors express their own ideas or solutions to problems.

For the first six items included in this composite variable, a higher rating indicates a higher frequency of various student-instructor interactions. However, the seventh item—instructors express their own ideas or solutions to problems—is just the opposite in that a higher rating for this item indicates lower interaction with students. Thus, in order to combine it with the other items in the student-instructor interaction composite variable, it was necessary to reverse the direction of this scale. Therefore, we subtracted the "instructors express their own ideas or solutions to problems" ratings from 5, the highest rating possible, arriving thus at the rating for "instructors do not express their own ideas or solutions to problems."

- Student-student interactions: composite index that averages the mean frequency ratings for three items all related to students' interaction with each other:
  - students review or challenge others' work
  - students work together with others
  - students get help from others
- Students' role in setting course pace and direction: composite index that averages the mean frequency ratings for two items about course pace and direction:
  - student set pace or direction of the class time
  - inverse rating for instructors set pace or direction of the class time

Similarly to the student-instructor interaction, one item in this index has the opposite direction to the orientation of the index. While the first item in this scale points to more student contribution in setting the direction of the course, the second item indicates more instructor and thus less student contribution in these matters. Thus, we reversed the direction of the second item, by subtracting its rating from the highest rating of 5. This provided us with rating for "instructors do not set pace or direction of the class time." The average of both items points to students' active role in setting the pace and direction of the class.

- Instructor behaviors: composite index that averages ratings for two instructor behaviors:
  - instructors establish a positive atmosphere
  - instructors summarize or place class work in a broader context

A4.5 Data Analysis

A4.5.1 Correlation Analysis

To check for a quantifiable relationship between the classroom observation variables and the section means for student gains, we used the non-parametric Spearman correlation test. As most of our data in this study was not normally distributed, according to the Shapiro-Wilk test of normality, using a non-parametric correlation test is the most appropriate choice, since it is specifically suited for non-normally distributed data. We used SPSS (version 18) to compute the section means for student gains responses and then performed statistical analyses on the resulting course-level data.

We used scatter plot function in SPSS to obtain visual representations of the relationships between various classroom observation variables and section means for different student gains. We used the Add Fit Line option in the chart editor to experiment with linear, quadratic, and cubic approximations for the data and add them to the scatter plots. We also checked the statistical significance of linear, quadratic, and cubic approximations of the data by using the

Curve Estimation function in SPSS. Besides performing a statistical test, this function also provided the coefficient of determination (R^2) for each curve estimation, indicating the percentage of variability in the data that is explained by each curve.

A4.5.2 Hierarchical Linear Model for Full Survey Data Set.

We used a Hierarchical Linear Model to assess the relative impact of participation in IBL courses on self-reported student gains. Using this type of model assesses the relative influence of student characteristics such as gender and class year, and of course characteristics (including participation in the IBL program) on self-reported gains in the course.

The Hierarchical Linear Model follows the standard intercept model with student and course level variables. The model appropriate for analysis is described below. The general form of the equation is:

$$Y = \beta_{01} + \beta_1 X1 + \beta_2 X2 + \ldots + \beta_6 X5 + \varepsilon_i$$

$$\beta_{01} = \gamma_{00} + \gamma_{01} W_1 + \ldots + \gamma_{0k} W_k + \mu_{0i}$$

$$\beta_{11} = \gamma_{11} W_{1i}$$

Where β_{01} is the level 1 intercept, β_i are level 1 regression weights, X_i are level 1 variable scores, γ_{00} is the level 2 intercept, γ_l are level 2 regression weights, and W_k are level 2 variables scores. ε_i and μ_{0i} are error terms. Y is the outcome variable that is being predicted; in this case learning gains. Level 1 refers to student-level variables, and Level 2 to course-level variables.

We formed factor variables from items loading greater than 0.4 on Varimax rotated factors using a Principal Components extraction. Factors scores were then assessed for independence from each other: only relatively independent factors with less than 0.3 correlation between factors were used in the analysis. The composite variables with low internal reliability were also excluded from the model.

The outcome variable for the model is a composite variable of weighted factor scores for five learning gains items from the SALG-M post-survey, all related to cognitive gains in understanding mathematical concepts, thinking, and relationships: the main concepts explored in class, the relationships among the main concepts, students' own ways of mathematical thinking, how mathematicians think and work, and how ideas from this class relate to ideas outside mathematics. The composite variable showed high internal reliability at $\alpha = 0.88$. The outcome variable is defined such that it has a mean of 0 and Standard Deviation of 1.

Independent variables included student demographics variables of gender and college class level. Course level variables included participation in the IBL program and participation in pre-service teaching courses. The coding of the variables is detailed in Table A4.1.

Table A4.1: Student-Level and Course-Level Variables

Student Level	**Coded**
Class standing: First Year Sophomore Junior Senior Graduate	No=0 Yes=1
Gender	Male = 1 Female = 2
Ethnicity	Hispanic = 1 Not Hispanic = 2

Course Level	**Coded**
IBL course	IBL = 1 Non-IBL = 0
Pre-service teacher course	Pre-service course = 1 Not a pre-service course = 0

The means and standard deviations for the outcome factor variable at different levels of the independent variables are presented in Table A4.2. Table A4.2 also includes sample sizes for different levels of the independent variables.

Table A4.2: Student Gains Factor Variable by Levels of Independent Variables

Independent Variables		Outcome Student Gains Factor Variable		
		Mean	Standard deviation	N
IBL	Non-IBL	-0.11	0.90	366
	IBL	0.05	1.03	873
	(Difference)	0.16		
Pre-service teacher course	No	0.10	0.91	953
	Yes	-0.35	1.18	286
	(Difference)	0.45		
Gender	Female	-0.13	1.05	647
	Male	0.14	0.92	575
	(Difference)	0.27		
Class year	First-year	0.11	0.89	438
	Sophomore	0.17	0.94	172
	Junior	-0.18	1.08	277
	Senior	-0.09	1.08	299
	Graduate	0.45	0.63	15

A4.6 *Hierarchical Linear Model for Combined Observation-Survey Data Set*

We constructed a second Hierarchical Linear Model with the intention to include the classroom observation variables. This model examines the effect of three course-level variables on student outcomes: total percentage of student-centered time, plus the previously tested indicators of IBL and pre-service courses. This model does not use any student-level variables (such as gender and class year) as predictors of outcomes.

For the second model, the means and standard deviations for the outcome factor variable at different levels of categorical independent variables are presented in Table A4.3. The table also includes the sample sizes for various levels of the independent variables.

Table A4.3: Student Gains Factor Variable by Levels of Categorical Independent Variables

Student Level Independent Variables		Student Gains Outcome Factor Variable		
		Mean	Standard Deviation	N
IBL	Non-IBL	-0.25	0.96	177
	IBL	-0.08	1.02	493
	(Difference)	0.17		
Pre-service teacher course	Not Pre-service	-0.03	0.91	428
	Pre-service	-0.31	1.13	242
	(Difference)	0.29		

On the other hand, Table A4.4 includes descriptive statistics for the continuous independent variable in this model: total percentage of student-centered time.

Table A4.4: Descriptive Statistics for Percentage of Student-Centered Time

Course-Level Independent Variable	Mean	Standard Deviation	N
Percentage of class time spent on student-centered activities	51.42	28.43	39

A4.7 Conclusions and Limitations

This type of analysis requires a very large volume of data, collecting which is a very labor-intensive task. Even with over 1200 student survey responses and 300 hours of observation, the size of the data set is on the very edge of what is needed to extract good correlations or to construct a complete, two-level hierarchical model. The cost and effort required to document these linkages are very high and should be seriously considered when undertaking such type of analysis.

Appendix A5: Study Methods for Mathematics Tests

A5.1 Introduction

In addition to students' own reports of their learning gains from a mathematics course (Ch. 3, Appendix A3), we gathered information from tests about possible changes in students' content knowledge and mathematical thinking during a college mathematics course. The first test, called *Learning Mathematics for Teaching* (LMT), offered us well-validated instruments for measuring pre-service teachers' cognitive gains from an IBL mathematics course. The other test, nicknamed the *Proof Test*, measured students' ability to evaluate a mathematical argument and determine its validity (see Exhibit E5.1). Using these two tests we gathered two mid-sized data sets on the development of students' mathematical knowledge and thinking. These studies addressed our research questions:

- How do students' mathematical knowledge and thinking change during an IBL college mathematics course?
- How do the changes differ by student groups, especially between IBL and non-IBL students?
- How do the changes align with results from other measures of students' learning gains?

We also used one more method to get comparative data on students' learning gains during a mathematics course. We refer to this as the *Instructor Ratings* instrument (see Exhibit E5.2). We designed a rubric that asked instructors to assess their students' learning from their course by rating the students on both their initial expertise in mathematics (as a check on instructors' perceptions) and overall learning gain in mathematics from the course. Because these data were mainly used to check the validity of our other measures of student mathematics learning and gains, results from this data set are reported in this Appendix.

A5.2 Learning Mathematics for Teaching (LMT) Tests

A5.2.1 LMT Tests

We used well-validated instruments, called *Learning Mathematics for Teaching* (LMT) tests, in studying pre-service teachers' gains in mathematical knowledge from an IBL mathematics course. The LMT instruments have been developed and validated by a team at the School of Education, University of Michigan, for assessing professional development courses for K-12 mathematics teachers (Hill, Schilling & Ball, 2004). Their project investigates the mathematical knowledge needed for teaching, and how such knowledge develops as a result of experience and professional learning. The LMT tests reflect both the mathematical content that teachers teach and the special knowledge they need to teach that content to students. The LMT measures are not designed to make statements about <u>individuals'</u> mathematical knowledge but rather to compare the mathematical knowledge of <u>groups of teachers</u> (such as those participating in particular courses) and how their knowledge develops over time.

The test items are designed to measure the development of mathematical knowledge needed for teaching: solving problems, using definitions, and identifying adequate explanations (Hill, Schilling & Ball, 2004). Each item and form has been piloted with over 600 elementary teachers, yielding information about scale reliability and item characteristics. Some examples of released LMT test items can be found at the project's website (Learning Mathematics for Teaching, n.d.). After participating in a training session provided by the LMT developers, we signed a terms of use contract for using the instruments, which helps to protect the utility and validity of the items by keeping them confidential except for research and evaluation purposes (for example, the items may not be used for teaching).

For this study, we chose a pre-test and related post-test on elementary *Number Concepts and Operations* in order to study changes in pre-service teachers' mathematical knowledge during an IBL course. The pre-test consisted of 24 and the post-test of 23 items. Standardized IRT (item response theory) scores provided by the developers were applied to match the results between pre-test and post-test. We use these scores in analyzing our data and reporting results.

We added a separate section at the end of a pre- and post-test with demographic questions about students':

- gender,
- class year,
- prior teaching experience (if any), in total number of years,
- grade level of prior mathematics teaching experience (if any),
- intended grade level for future mathematics teaching.

Items on teaching experience were included because some graduate students in the study might be completing teacher certification programs after some prior teaching experience.

We verified the course, section and instructor for each student. In addition, we asked students to establish an identifier that helped us to match students' pre- and post-test answers and also their survey responses.

A5.2.2 Study sample for the LMT Test

Data from the LMT measures were gathered from students taking targeted IBL courses for pre-service teachers at two campuses. Altogether, we got pre- and post-test data from pre-service teachers in several distinct two-course sequences preparing teachers for elementary and middle school, elementary, or secondary teaching. In all, six data sets from three course sequences were received during the two academic years 2008-2010. No comparative (non-IBL) sections of these courses were offered at any of the campuses. The results reported in Chapter 5 are based on data from the 109 pre-service teachers who took both the pre-test and post-test at two campuses. The sample is described in Table A5.1, showing demographic characteristics of students in the three main groups, divided by the target audience (elementary, elementary/middle, or secondary) of the courses in which the students participated.

Table A5.1: LMT Test Sample Teachers by Gender, Ethnicity, and Race

Indicator	Group 1 Elementary		Group 2 Elementary/middle		Group 3 Secondary		Total	
	Count	%	Count	%	Count	%	Count	%
Gender								
Women	27	*100*	62	*97*	9	*50*	98	***90***
Men	0	*0*	2	*3*	9	*50*	11	***10***
	27	*100*	64	*100*	18	*100*	109	***100***
Ethnicity								
Hispanic or Latino	6	*23*	4	*7*	1	*6*	11	***11***
Not Hispanic or Latino	20	*77*	57	*94*	15	*94*	92	***89***
	26	*100*	61	*100*	16	*100*	103	***100***
Race								
Asian	4	*20*	3	*5*	1	*10*	8	***9***
Multiracial	2	*10*	5	*8*	1	*10*	8	***9***
White	14	*70*	53	*87*	8	*80*	75	***82***
	20	*100*	61	*100*	10	*100*	91	***100***

Table A5.1 indicates that most of the pre-service teachers who took both the pre- and post-test are women (90%). We had only 11 men in our sample—a typical distribution of gender among teachers. In particular, all the elementary teachers were women and only two in Group 2 (elementary/middle school) teachers were men. Half of the secondary school teachers were men, but their number was still low in our sample overall.

Table A5.1 also shows that only 11% (11) of the pre-service teachers in our sample overall were Hispanic or Latino, although their proportion was a bit larger among elementary teachers (23%). Most (82%) of the students were white and the number of students reporting other races was 16 (18%). Again, this shows the sample had little variation in students' demographic characteristics.

Table A5.2: LMT Test Sample by Academic Status and Major Subject

Indicator	Group 1		Group 2		Group 3		Total	
	Count	%	Count	%	Count	%	Count	%
Academic status								
Sophomore	1	*4*	2	*3*	2	*11*	5	***5***
Junior	8	*30*	53	*83*	3	*17*	64	***59***
Senior	18	*67*	6	*9*	13	*72*	37	***34***
Graduate, other	0	*0*	3	*5*	0	*0*	3	***3***
	27	*100*	64	*100*	18	*100*	109	***100***
Major subject								
Math or applied math	0	*0*	14	*23*	14	*88*	28	***27***
Science, engineering, computer science	3	*12*	8	*13*	1	*6*	12	***12***
Non-science	23	*89*	40	*65*	1	*6*	64	***62***
	26	*100*	62	*100*	16	*100*	104	***100***

Table A5.2 displays pre-service teachers' academic status and college majors. Typically, these students were well along in their academic careers. Only five were second-year students and all the others were juniors (59%) or seniors (34%). Slight variation appeared between the three groups. Most of the elementary and secondary school teachers were seniors (67%), whereas nearly all the elementary/middle school teachers (83%) were juniors.

The distribution of the elementary and middle school teachers' college majors was typical. Most of them had a non-science (education) as their major subject (89%, 65%), whereas nearly all of the secondary school teachers (88%) reported a major in Mathematics or Applied Mathematics.

A5.2.3 Methods for Administering and Analyzing the LMT Test

Both the pre- and post-test were administered as a paper-and-pencil test in class by the IBL project coordinators at two campuses. The coordinators were provided with written instructions for administering the tests and preserving students' confidentiality. The written tests were returned to researchers, coded, and the data entered into separate SPSS data files by trained student assistants. Finally, the data on students' responses to the test items were matched with their responses on our survey instruments (Ch. 3). A smaller sample set of students' LMT test results was matched with instructor ratings (see Section A5.4).

The raw scores on the pre- and post-test were converted to standardized IRT scores according to a scoring table provided by the developers of the tests. All the analyses were performed by using these standardized scores that also enabled matching of students' pre-test scores to their post-test scores. To report results, we also apply the IRT scoring table in illustrating average score gains in mathematical knowledge from actual LMT test scores. In addition to descriptive statistics, we applied correlational analysis and parametric tests (independent and pairwise t-tests, ANOVA) to analyze the data. Stepwise regression was used to study the extent to which LMT test score gains were explained by other measures of mathematical knowledge and learning.

A5.3 Proof Test

We used another test, which we refer to as the "Proof Test," to measure students' mathematical knowledge and thinking. This test studied students' ability to evaluate a mathematical argument and determine its validity. Test data were gathered from interviews and paper-and-pencil tests.

A5.3.1 Proof Test instrument

The proof test was based on items on evaluating mathematical arguments that were designed by Weber (2009). We reformulated the original test into a paper-and-pencil test. Wording of some of the claims was clarified with a few additional words suggested by a mathematics professor. We also numbered the lines in the arguments so that students could reference specific lines in their comments. In order to obtain equal numbers of answers to each argument, we used two tests, forms A and B, which were identical except for the order of the arguments.

The one-hour test included nine of the ten original arguments from Weber (2009) on algebra, number theory and calculus (see Exhibit A5.1). Three of these arguments were valid and six arguments had some flaws for students to detect. Each argument was followed by structured questions to probe:

- Did students understand the argument?
- To what extent were students convinced by the argument?
- To what extent did students find it to have explanatory power?
- Did students consider the argument to be a mathematical proof?

Students answered the first three questions on a scale between 1 (strong disagreement) and 5 (strong agreement). On the fourth question, students assessed whether an argument was a proof (fully rigorous, not fully rigorous, not a proof, don't understand). At the end of this question, we requested students' explanations for their reasoning behind their decisions (see Exhibit A5.1).

A cover sheet for the proof test gathered demographic data on students, including ethnicity, race, gender, class year, academic major, and whether or not students planned to become a K-12 teacher. We also verified the course sections in which students were enrolled, the number of their college mathematics courses they had taken before and during or after the target course, and their expectations for their course grade (see Exhibit A5.1).

A5.3.2 Study sample for Proof Test

The first data set was gathered from one-on-one problem-solving interviews. Later, the test was revised into a paper-and-pencil form that was administered either in class to all students, or out of class to volunteers. In the interviews, students were asked to verbally explain the reasoning behind their answer about each argument, and these responses were recorded. On the paper-and-pencil test, students wrote down their reasoning about whether or not each argument was a mathematical proof. Both the interviews and the paper-and-pencil tests took an hour.

In all, we obtained tests from 42 IBL students (27 men, 15 women) and 35 non-IBL students (19 men, 16 women) at the end of a mathematics course. Of these, 24 students (14 IBL, 10 non-IBL) took an interview and 53 students (28 IBL, 25 non-IBL) a paper-and-pencil test. Most of the students were volunteers (63) who were paid a modest honorarium for participating. Only 14 students took an in-class post-test. In addition, we got pre/post-test data from one section (20 pre-, 14 post-tests). Table A5.3 displays demographics of the students, for IBL and non-IBL students separately.

Table A5.3: Proof (Post-) Test Sample by Gender, Ethnicity, Race, and Course Type.

Indicator	**IBL students**		**Non-IBL students**		**Total**	
	Count	%	Count	%	Count	%
Gender						
Women	15	*36*	16	*46*	31	***40***
Men	27	*64*	19	*54*	46	***60***
	42	*100*	35	*100*	77	***100***
Ethnicity						
Hispanic or Latino	3	*8*	2	*6*	5	*7*
Not Hispanic or Latino	37	*93*	33	*94*	70	***93***
	40	*100*	35	*100*	75	***100***
Race						
Asian	14	*34*	14	*44*	28	***38***
White	25	*61*	16	*50*	41	***56***
Other race	2	*5*	2	*6*	4	***6***
	41	*100*	32	*100*	73	***100***

The proportion of men among IBL students exceeded that among non-IBL students. Otherwise, the sample looked like our other student samples. Only five students' ethnicity was Hispanic or Latino, and only 4 students reported a race other than white or Asian. However, one third of IBL but 44% of non-IBL students were Asian.

Table A5.4: Proof (Post-) Test Sample by Course Type and Academic Status.

Indicator	IBL students		Non-IBL students		Total	
	Count	%	Count	%	Count	%
Academic status						
First-year	0	*0*	0	*0*	0	***0***
Sophomore	7	*18*	7	*21*	14	***19***
Junior	13	*33*	14	*41*	27	***37***
Senior	20	*50*	13	*38*	33	***45***
Other	0	*0*	0	*0*	0	***0***
	40	*100*	34	*100*	74	***100***
Major subject						
Mathematics	30	*83*	26	*87*	56	***85***
Natural science	1	*3*	1	*3*	2	***3***
Math/Natural Science	1	*3*	2	*7*	3	***5***
Non-science	0	*0*	0	*0*	0	***0***
Math/Non-science	4	*11*	1	*3*	5	***8***
	36	*100*	30	*100*	66	***100***

According to instructors, our target courses represented an introductory or mid-level proof-based course. But, in practice, we found that many students had substantial proving experience in prior courses. This is also reflected in Table A5.4: most of the students were seniors or juniors. Nearly all were also pursuing a math major.

A5.3.3 Methods for Proof Test

We used descriptive statistics and parametric (T-test) or non-parametric (Mann-Whitney) tests to examine differences between student groups in responses to the three first structured questions (see Exhibit A5.1). Table A5.5 displays averages of students' ratings on the three structured questions, for each argument separately.

Table A5.5: Average Student Ratings of Single Arguments

Argument	Course type					
	IBL			Non-IBL		
	Understanding	Conviction	Explanation	Understanding	Conviction	Explanation
Valid arguments						
Arg 1	4.8	4.7	4.3	4.7	4.7	4.5
Arg 3	4.6	4.4	4.3	4.5	4.3	4.2
Arg 4	4.4	4.2	4.4	4.2	4.0	4.1
Invalid arguments						
Arg 2	4.8	4.5	4.4	4.8	4.4	4.3
Arg 5	4.3	2.1	1.9	4.4	2.2	1.8
Arg 6	4.6	3.3	3.4	4.7	3.5	3.5
Arg 7	4.3	2.1	2.1	4.8	2.4	2.4
Arg 8	4.3	3.3	3.3	4.5	3.5	3.5
Arg 9	4.4	3.5	3.4	4.1	3.1	2.9

Scales (1-5):

Understanding: 1=not understand fundamental details to 5=understand completely.

Conviction: 1=not convinced at all to 5=completely convinced.

Explanation: 1=does not explain to 5=really illuminates why it is true.

Differences in frequency distributions between student groups in answers to the fourth question were compared by using a non-parametric test (Chi2).

In addition to the four structured questions, we analyzed students' written reasoning about each argument. These data came from 53 students (28 IBL, 25 non-IBL). These qualitative data were coded and analyzed qualitatively according to eight main themes related to the nature of students' criteria for assessing the arguments. The categories were derived from preliminary analysis of a subset of written comments and finalized using inductive content analysis (Miles & Huberman, 1994; Strauss & Corbin, 1990) of the complete set of written comments. Table A5.6 presents the eight main themes and the 29 sub-categories under the main themes, and the frequencies of student comments in each category.

Table A5.6: Frequency of Criteria Used in Written Comments about Reasoning

Main category	**Subcategory**	**Number of students reporting**		
	Description	once	2-3 times	≥4 times
Understanding	Some step(s) or the whole argument are/are not understandable	19	8	-
False statement	Makes false statement about an argument	18	2	-
Inadequate reasoning	Lacks justification	16	19	1
	Uses empirical/perceptual evidence	8	-	-
	Experience: have seen the argument/proof before	4	-	-
	An argument is complete	4	-	-
Use of steps as criterion	Steps are acceptable/not acceptable	13	7	-
	All steps are included/not included	15	1	-
	Proved step by step	5	1	-
Formalism	Theorems, formulas included/not included	15	5	1
	Written/not written formally	3	-	-
	External structure is correct/incorrect	16	6	1
	Mathematical rules, concepts, terms are included/not included	19	3	-
	Doesn't remember the concepts, terms, definitions, proof	1	1	-
Quality of explanation or reasoning	Requests (more) explanation, reasoning	20	16	5
	Level of explanation/reasoning is assessed	8	4	-
	Good explanation/reasoning in the argument	14	7	-
	Visual aid is used/not used	8	1	-
	The steps are clearly stated/explained	5	1	2
Rigor	Lack of/rigor of steps found in an argument	25	18	2
	Critical about a claim/presupposition	8	3	-
	Detects the flaw(s) in an argument	18	20	2
	Does not accept a picture/graph/equation as a rigorous way to prove	9	2	-
	Does not accept empirical evidence	30	2	-
	Suggests a more rigorous way to prove	10	1	-
Assessment of an argument as a whole	Beauty, appearance, ease of argument as a whole	11	6	-
	Style of (presentation or writing in) an argument	20	13	1
	Logic of argument as a whole	11	7	1
	Subject/mathematical level of an argument	11	2	-

A5.4 Instructor Ratings of Student Expertise

We conducted a small experiment using instructor ratings of students' mathematical expertise and learning, as one more method to get comparative data on IBL students' learning gains during a mathematics course. This is called the *Instructor Ratings* instrument.

A5.4.1 Instructor Ratings instrument

We designed a rubric that asked instructors to assess their students' learning from their course by rating the students on both their initial expertise in mathematics and their overall learning gain from the course. Instructors gave their answers in a prepared Excel spreadsheet (Exhibit 5A.2).

We asked instructors to give two ratings for each student: their overall level of expertise in mathematical knowledge and thinking at the *start* of the course, and overall *gain* or improvement in mathematical knowledge and thinking by the *end* of the course. That is, we asked instructors to distinguish students' incoming *ability* from their *learning* in their course. The instructors gave their ratings on a scale between from 5 to 1: very high, high, moderate, low, very poor or strongly lacking (see Exhibit E5.2).

A5.4.2 Study sample for Instructor Ratings

The sample on instructor ratings is based on data from four sections at one campus. We asked the campus coordinators to establish an identifier for each student that was later matched with the other data sets from these students. Students in two sections were math-track students and the two other sections represented courses for elementary/middle school pre-service teachers. Instructors rated students' mathematical expertise at the end of a course. In all, we matched instructor ratings from 27 math-track students and 37 pre-service teachers to the other data sets.

A5.4.3 Methods for Instructor Ratings

We compared data from the instructor ratings to data on the same students' learning gains from a mathematics course as measured by the SALG-M, using correlational analysis. Descriptive statistics and parametric (independent and paired T-test, ANOVA) or non-parametric tests (Chi2) were used to study subgroup differences in initial mathematical expertise and gains in mathematical expertise.

A5.4.4 Results from Instructor Ratings

Table A5.7 displays frequencies for the instructor ratings, for math-track students and pre-service teachers separately. The two IBL student groups differed from each other. On average, pre-service teachers' (M=3.4) rated initial mathematical expertise exceeded that of math-track students (M=2.6, $p<0.01$). But math-track students' (M=3.4) rated gains were higher than that of pre-service teachers (M=2.3, $p<0.001$). Also, comparisons between initial expertise and gain in expertise showed a difference between these two student groups. While there was a clear improvement in math-track students' rated mathematical expertise ($p<0.001$), pre-service teachers' gains in mathematical expertise were rated clearly lower ($p<0.001$) than their initial mathematical expertise. While 48% of math-track students had high or very high rated gain in mathematical expertise during an IBL course, only 5% of pre-service teachers' gain in mathematical expertise was rated at this level.

Table A5.7: Instructor Ratings by Course Type

Instructor rating	**Math-track students**		**Pre-service teachers**		**Total**	
	Count	%	Count	%	Count	%
Initial expertise	27	*100*	37	*100*		***100***
very high	-	-	5	*4*	5	***8***
high	2	*7*	15	*41*	17	***27***
moderate	11	*41*	10	*27*	21	***33***
low	14	*52*	5	*14*	16	***25***
very poor	-	-	5	*5*	5	***8***
Gain in expertise	27	*100*	37	*100*	64	***100***
very high	3	*11*	1	*3*	4	***6***
high	10	*37*	1	*3*	11	***17***
moderate	11	*41*	11	*30*	22	***34***
low	2	*7*	18	*49*	20	***31***
very poor	1	*4*	6	*16*	7	***11***

A5.5 Reliability and Validity of Mathematics Tests

A5.5.1 LMT Tests

The *Learning Mathematics for Teaching* (LMT) instruments are carefully developed and well-validated instruments (Hill, Schilling & Ball, 2004). The test items are designed to measure the development of mathematical knowledge needed for teaching: solving problems, using definitions, and identifying adequate explanations (Hill, Schilling & Ball, 2004). Each item and form has been piloted with over 600 elementary teachers, yielding information about scale reliability and item characteristics. Our sample from two campuses was also large enough to detect real gains and differences among students. However, because IBL methods were used in all sections of the courses targeted to pre-service teachers that were available for this study, we had no opportunity to compare student learning with that in a traditionally taught course.

At the start of the course, pre-service teachers reported their score (1-5) on the AP Calculus test (if they had taken it), their current estimated undergraduate GPA, and their expected grade in the present course. They also reported their expected course grade at the end of the two-term sequence. We checked to see how well these other academic measures correlated with students' LMT scores. Table A5.8 summarizes findings on these correlations.

Table A5.8: Correlations of Students' LMT Scores with Other Achievement Indicators

LMT score component	Expected course grade *at start*	Expected course grade *at end*	AP Calculus score *(1-5)* *at start*	Estimated GPA *at start*	Cognitive gains (SALG-M)	Affective gains (SALG-M)	Instructor rating: gain in math expertise
Pre-test	+ **	+ **	+		+*	+ **	+
Post-test	+ **	+ **	+ **			+ *	+
Score gain	+	+	+				

* p< 0.05, ** p< 0.01

Students' estimated undergraduate GPA did not correlate with the LMT test scores. This is understandable since these students' undergraduate studies were not usually focused on mathematics; their GPA represents a broad range of courses and not mathematical ability. However, all the three LMT test measures—pre-test, post-test, and test score gain—did correlate positively with students' AP test scores. This result shows that the LMT tests measured mathematical knowledge that was somewhat related to that measured by AP Calculus tests, even though the LMT test content addressed number and operations, not calculus. However, this finding is limited: we had AP test data from just 36 students, who came mostly from one university, and only the correlation between AP test score and LMT post-test score was statistically significant ($p<0.01$).

Students' expected grades reported at both the start and end of their course correlated positively with both LMT pre- and post-test IRT scores ($p<0.01$). But the positive correlations between expected grades and students' test score *gains* were weaker. In general, students who expected a higher grade in the course tended to earn higher LMT scores and to make greater LMT test score gains, but this latter relation was not statistically significant. In other words, students who thought they would get a good grade did get better test scores, but their expected grade was not well linked to their LMT test score growth.

We also examined students' LMT score gains in comparison with their self-reported cognitive, affective, and social gains from the SALG-M survey instrument (Ch. 3). Students' self-reported gains correlated positively with the LMT pre-test scores, while LMT post-test scores were positively related only to gains in confidence ($p<0.05$) and positive attitude ($p<0.01$). However, students' self-reported gains were not generally related to LMT test score gains. But among those students who started with the lowest initial LMT scores, LMT test score gains were positively (but not statistically significantly) related to cognitive, affective and social learning gains. That is, students' LMT score gains were consistently reflected only in the self-reports from students with low pre-test scores.

For a subset of students, we compared LMT test score gains with instructor ratings of students' gain in mathematical expertise during an IBL course. No direct correlation appeared between instructor ratings of students' gain in mathematical expertise and LMT score gains. However, instructor ratings of students' expertise in the beginning of a course were (statistically significantly) positively related to both LMT pre- and post-test scores, and slightly to their LMT-test gain scores. That is, the instructor's assessment of initial mathematical expertise was consistent with LMT pre-test and post-test scores; instructors could identify stronger or weaker students overall. But the instructor was a less successful judge of *learning* as measured by LMT score gains. This is a similar result to the relationship between instructors' assessment of gains in mathematical expertise and students' self-assessed grade, in that students and instructors both judge relative performance with some accuracy, but do not accurately predict learning.

A5.5.2 Proof Test

The proof test was intended to measure students' ability to assess mathematical arguments on algebra, number theory and calculus. It was based on items on evaluating mathematical arguments that were designed and previously tested by Weber (2009). After gathering and analyzing proof test data from individual student interviews, we reformulated the original test into a paper-and-pencil test that was further reviewed by a mathematics professor. Students were provided with an opportunity to offer written feedback or additional comments on the arguments. Both the student interviews and written test sheets indicated that students did not have difficulty in understanding the questions or goals of the proof test. However, the proof test sample was not the same as the samples from our surveys. Thus, we were unable to compare results from proof test with other indicators of student learning or gains.

Our ability to draw strong conclusions is limited by our sample of students. The students who volunteered to take the proof test were strong mathematics students, based on their self-reported grades and high numbers of prior mathematics courses taken. We surmise that differences in the responses and reasoning of IBL vs. non-IBL students are less easily detected among this group than among lower-achieving or less experienced mathematics students. Thus, while the test itself seems to be sensitive to differences in students' understanding, our sample is not optimized to detect group differences that might result from IBL instruction focused on proof processes. Moreover, this particular test is likely to be more sensitive in "introduction to proof" courses where enrollment is controlled or sequenced in such a way as to assure that most students have relatively little prior proof experience. In this sample, many students had proof experience already, and we cannot rule out that the test measured expertise developed in earlier courses.

A5.5.3 Instructor ratings

We used instructor ratings in order to get comparative data on IBL students' learning gains during a mathematics course. We checked to see how well initial expertise and gains in expertise correlated with each other and how these instructor ratings correlated with other indicators of mathematical knowledge or academic gains. Because the patterns in instructor ratings for math-track students and pre-service teachers differed (see section A5.4.4), we studied these relations within the two student groups separately. Table A5.9 summarizes findings from the nonparametric (Spearman) correlations between mathematical expertise and other indicators of student knowledge or learning.

Table A5.9: Correlation of Mathematical Expertise with Other Performance Indicators, by Student Group

Instructor rating (1-5)	Initial expertise	Expected course grade *at start*	Expected course grade *at end*	AP Calculus score *at start*	Estimated GPA *at start*
Math-track students					
Initial expertise		+	+ **	+	+
Gain in expertise	+	+	+	-	+
Pre-service teachers					
Initial expertise		+ **	+ **	+ *	+ *
Gain in expertise	-*			-	+

* p< 0.05, ** p< 0.01

Instructor ratings of initial mathematical expertise were somewhat consistent with students' mathematical performance level as indicated by their self-reported AP Calculus test score and GPA level at the start of a course. This applied more clearly to the ratings by the pre-service teachers' instructor than to those of the instructor of the math-track students. Moreover, ratings of initial mathematical expertise were consistent with both the grade expectations of both student groups at the start and end of a course. Students who expected a higher course grade were rated higher in mathematical expertise by their instructor, and the opposite was true for students who had lower course grade expectations.

However, in courses for pre-service teachers, the instructor's ratings of students' gains in mathematical expertise did not correlate with students' grade expectations at the end of a course . Similarly, in math-track courses, the correlation between the instructor-rated gains in expertise and students' own grade expectation at the end of a course was only slightly positive. That is, students' own expectations of their success in an IBL course did not match with their instructors' ratings of their learning gains. This may be an accurate assessment of the situation, if grades are seen by both parties to reflect achievement rather than learning.

Among pre-service teachers, initial mathematical expertise (as assessed by the instructor) was positively related to all other indictors of knowledge or gains. That is, instructor ratings were consistent with students' own expectations and the external performance indicators of AP test scores and estimated GPA at start of an IBL course. Moreover, the negative correlation between pre-service teachers' initial expertise and gain in expertise indicted that instructor ratings of gain tended to be higher for students who started with lower mathematical expertise. Further analysis by student groups showed that, on average, those pre-service teachers who started with high or very high initial expertise got lower ratings in gain in expertise than other pre-service teachers ($p<0.05$). Instructors saw these students as initially strong, therefore making lower gains. Also pre-service teachers' AP test score correlated slightly negatively to their instructor-rated gain in mathematical expertise. These results are consistent with other findings on pre-service teachers (see Section 5.2.3): weaker students at start of an IBL course may gain more than students with stronger mathematical background.

In contrast, among math-track students, instructor ratings of students' gains in mathematical expertise correlated slightly positively with their ratings of initial expertise. That is, students who started with stronger mathematical expertise also tended to gain more during an IBL course. But, unlike the result for pre-service teachers, the correlation was not statistically significant for math-track students. Overall, instructors seemed less successful in assessing learning during an IBL course than they were in assessing the initial level of students' mathematical expertise.

We also examined correlations between instructor-rated mathematical expertise and students' self-reported learning gains (SALG-M, Ch 3). Table A5.10 displays the results of the nonparametric (Spearman) correlations.

Table A5.10: Correlations of Gain in Expertise with Learning Gains (SALG-M).

Instructor rating (1-5)	Mathematical concepts & thinking	Application	Affective gains	Social gains
Initial expertise				
Gain in expertise	+	-	+	

* $p< 0.05$, ** $p< 0.01$

IBL students' self-reported gains were not generally related to their initial mathematical expertise, as rated by their instructor. But students' instructor-rated gains in mathematical expertise were modestly related to their self-reported learning gains. Overall, students who had higher gains in mathematical expertise, as rated by their instructor, also tended to self-report higher gains in understanding concepts and mathematical thinking and problem-solving. They also tended to report higher affective learning gains. Students' gains in collaboration did not relate to their gain in mathematical expertise. Rather, higher instructor ratings in gain in mathematical expertise were slightly negatively related to students' self-reported gain in application of mathematical knowledge.

These results indicate that students' own assessments of learning were somewhat consistent with their instructor's ratings of their gains in mathematical expertise during an IBL course. However, the correlations were slight and mostly applied to students with initial poor to moderate mathematical expertise.

In sum, instructor ratings of gains in mathematical expertise varied between the student groups and courses. Moreover, they were mostly not consistent with other indicators of student learning. But students' self-reported learning gains were moderately in line with their instructor's ratings. Instructors may not be as successful in assessing their students' learning during as they are in assessing the initial level of their students' mathematical expertise. This applied especially to pre-service teachers and students who started with high mathematical expertise.

A5.6 References cited

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal, 105*, 11-30.

Learning Mathematics for Teaching (LMT) Project website (no date). Accessed 5/11/2011 at http://www.sitemaker.umich.edu/lmt

Miles, M. B., & Huberman, A. M. (1984). *Qualitative data analysis : a sourcebook of new methods*. Thousand Oaks, CA: Sage Publications.

Strauss, A., & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Thousand Oaks, CA: Sage Publications.

Weber, K. (2009). Proving is not convincing. *Conference on Research in Undergraduate Mathematics Education,* Raleigh, NC, February 26-March 1, 2009.

Instructions: Assessing Mathematical Arguments

Thank you for participating in our study! This problem-solving test is part of a research study on how students learn to construct and evaluate mathematical proofs.

The problem-solving session will take about 1 hour of your time and will include 9 problems. For each problem, you will be asked to examine a mathematical argument and to decide whether or not you think it is a valid mathematical proof. You will then answer a few questions about each argument and write in some comments to help us understand your reasoning. Your comments do not need to be lengthy, but please show the thinking that led you to your answer. For convenience, the lines of each argument are numbered so that you can refer to them in your answer if you wish. Please work steadily, but do not rush. If you need more work space for any answer, please use the space on the last page and note the problem number.

Your participation is voluntary. You may skip questions or tasks that you do not wish to answer, or choose not to participate. Your answers are anonymous and will not be reported in any way that can identify you individually; they will be reported in groups with answers from other students from your course and other schools.

When you have finished the problems or are nearly out of time (whichever comes first), please complete question #10.

By taking this test, in part or whole, you agree that we may use this data to understand and improve the quality and effectiveness of college mathematics education. Thanks for your help!

Start time:

End time:

Argument 1

Claim: For all real numbers *a* and *b*: $(a + b)^2 = a^2 + 2ab + b^2$

Line:

1 $(a + b)^2 = (a + b)(a + b)$

2 $(a + b)(a + b) = a(a + b) + b(a + b)$

3 $a(a + b) = a^2 + ab$

4 $b(a + b) = ba + b^2$

5 So $(a + b)(a + b) = a^2 + ab + ba + b^2 = a^2 + 2ab + b^2$

For each question, circle the answer (1 to 5) that best fits your thinking about the argument.

A. Do you feel that you understood the argument that was presented?

1	2	3	4	5
There are fundamental details of the argument that I don't understand				I understand the argument completely

B. Are you convinced by this argument?

1	2	3	4	5
Not convinced at all				Completely convinced

C. Does this argument explain why the assertion is true?

1	2	3	4	5
No, the argument does not explain why the assertion is true				Yes, it really illuminates why the assertion is true

D. Would you consider this argument to be a mathematical proof?

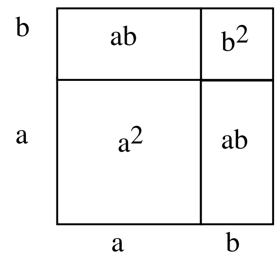
1. ______ Yes, I consider this argument to be a fully rigorous mathematical proof.
2. ______ Yes, I consider this argument to be a proof, although not fully rigorous.
3. ______ No, I think this argument does not meet the standards of a proof.
4. ______ Not sure, because I don't fully understand the argument.

Please explain your reasoning about **why** you think **this is** or **is not** a mathematical proof.

Argument 2

Claim: For all real numbers *a* and *b*: $(a + b)^2 = a^2 + 2ab + b^2$

Consider the diagram at right:



Line:

The length and width of the square are each (a+b), so the area of the diagram is $(a+b)(a+b) = (a+b)^2$. The area can also be found by adding the areas of the four cells of the square whose areas are a^2, ab, ab, and b^2, which is $a^2 + 2ab + b^2$.

So $(a+b)^2 = a^2 + 2ab + b^2$.

For each question, circle the answer (1 to 5) that best fits your thinking about the argument.

A. Do you feel that you understood the argument that was presented?

1	2	3	4	5
There are fundamental details of the argument that I don't understand				I understand the argument completely

B. Are you convinced by this argument?

1	2	3	4	5
Not convinced at all				Completely convinced

C. Does this argument explain why the assertion is true?

1	2	3	4	5
No, the argument does not explain why the assertion is true				Yes, it really illuminates why the assertion is true

D. Would you consider this argument to be a mathematical proof?

1. _______ Yes, I consider this argument to be a fully rigorous mathematical proof.
2. _______ Yes, I consider this argument to be a proof, although not fully rigorous.
3. _______ No, I think this argument does not meet the standards of a proof.
4. _______ Not sure, because I don't fully understand the argument.

Please explain your reasoning about **why** you think **this is** or **is not** a mathematical proof.

Argument 3

Claim: For all natural numbers $n > 1$, $n^3 - n$ is divisible by 6.

Line:

1 $n^3 - n = n(n^2 - 1) = n(n+1)(n-1)$.
2 Either n is even or $n+1$ is even.
3 Since both numbers are factors of $n^3 - n$, $n^3 - n$ is even.
4 Because $n-1$, n, and $n+1$ are three consecutive numbers, one of them is divisible by 3.
5 So $n(n+1)(n-1)=n^3 - n$ is divisible by 3.
6 Since $n^3 - n$ is even and divisible by 3, $n^3 - n$ is divisible by 6.

For each question, circle the answer (1 to 5) that best fits your thinking about the argument.

A. Do you feel that you understood the argument that was presented?

1	2	3	4	5
There are fundamental details of the argument that I don't understand				I understand the argument completely

B. Are you convinced by this argument?

1	2	3	4	5
Not convinced at all				Completely convinced

C. Does this argument explain why the assertion is true?

1	2	3	4	5
No, the argument does not explain why the assertion is true				Yes, it really illuminates why the assertion is true

D. Would you consider this argument to be a mathematical proof?

1. ______ Yes, I consider this argument to be a fully rigorous mathematical proof.
2. ______ Yes, I consider this argument to be a proof, although not fully rigorous.
3. ______ No, I think this argument does not meet the standards of a proof.
4. ______ Not sure, because I don't fully understand the argument.

Please explain your reasoning about **why** you think **this is** or **is not** a mathematical proof.

Argument 4

Claim. There is no real number x which solves the equation $4x^3 - x^4 = 30$.

Line:

1 Consider the function, $f(x) = 4x^3 - x^4$. Because $f(x)$ is a polynomial of degree 4
2 and the coefficient of x^4 is negative, $f(x)$ is continuous and will approach $-\infty$ as x
3 approaches ∞ or $-\infty$. Hence, $f(x)$ must have a global maximum. The global maximum
4 will be a critical point. $f'(x) = 12x^2 - 4x^3$. If $f'(x) = 0$, then $x = 0$ or $x = 3$. $f(0) = 0$. $f(3) = 27$.
5 Since $f(3)$ is the greatest y-value of f's critical points, the global maximum of $f(x) = 27$.
6 Therefore $f(x) \neq 30$ for any real number x. $4x^3 - x^4 = 30$ has no real solutions.

For each question, circle the answer (1 to 5) that best fits your thinking about the argument.

A. Do you feel that you understood the argument that was presented?

1	2	3	4	5
There are fundamental details of the argument that I don't understand				I understand the argument completely

B. Are you convinced by this argument?

1	2	3	4	5
Not convinced at all				Completely convinced

C. Does this argument explain why the assertion is true?

1	2	3	4	5
No, the argument does not explain why the assertion is true				Yes, it really illuminates why the assertion is true

D. Would you consider this argument to be a mathematical proof?

1. ______ Yes, I consider this argument to be a fully rigorous mathematical proof.
2. ______ Yes, I consider this argument to be a proof, although not fully rigorous.
3. ______ No, I think this argument does not meet the standards of a proof.
4. ______ Not sure, because I don't fully understand the argument.

Please explain your reasoning about **why** you think **this is** or **is not** a mathematical proof.

Argument 5

Claim. Any even integer greater than two can be written as the sum of two prime numbers.

Consider the following table:

Even	Sum of two primes		
4	2+2		
6	3+3		
8	3+5		
10	3+7,	5+5	
12	5+7		
14	3+11,	7+7	
16	3+13,	5+11	
18	5+13,	7+11	
20	3+17,	7+13	
22	3+19,	5+17,	11+11
24	5+19,	7+17,	11+13
26	3+23,	7+19,	13+13

Line:

1 First, note that each even number between 4 and 26 can be written as the sum of two
2 primes. Second, note that the number of pairs of primes that work appears to be
3 increasing. For 4, 6, 8, and 12, there is only one prime pair whose sum is that number.
4 For 22, 24, and 26, there are three prime pairs whose sum is that number. Every even
5 number greater than 2 will have at least one prime pair whose sum is that number.
6 For large even numbers, there will be many prime pairs that satisfy this property.

For each question, circle the answer (1 to 5) that best fits your thinking about the argument.

A. Do you feel that you understood the argument that was presented?

1	2	3	4	5
There are fundamental details of the argument that I don't understand				I understand the argument completely

B. Are you convinced by this argument?

1	2	3	4	5
Not convinced at all				Completely convinced

C. Does this argument explain why the assertion is true?

1	2	3	4	5
No, the argument does not explain why the assertion is true				Yes, it really illuminates why the assertion is true

(continued, next page)

D. Would you consider this argument to be a mathematical proof?

1. _______ Yes, I consider this argument to be a fully rigorous mathematical proof.
2. _______ Yes, I consider this argument to be a proof, although not fully rigorous.
3. _______ No, I think this argument does not meet the standards of a proof.
4. _______ Not sure, because I don't fully understand the argument.

Please explain your reasoning about **why** you think **this is** or **is not** a mathematical proof.

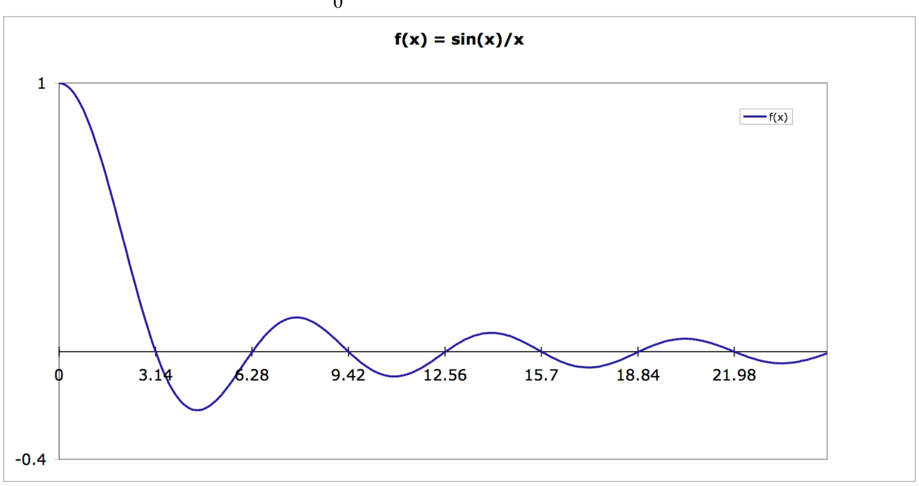
Argument 6

Claim. $\int_0^\infty \frac{1}{x}\sin x dx > 0$

The graph of $f(x) = \frac{1}{x}\sin x$ is given below.

Line:

1 $\int_0^\infty \frac{1}{x}\sin x dx > 0$ means that $f(x) = \frac{1}{x}\sin x$ has more area above the x-axis than below it.
2 To show this, note that it is clear from the graph that the first positive region—between 0
3 and π (about 3.14)—has more area than the first negative region—between π and 2π
4 (between 3.14 and 6.28). The second positive region has more area than the second
5 negative region. The third positive region has more area than the third negative region.
6 Since each positive region has a greater area than the negative region to the right of
7 it, the overall area of $\int_0^\infty \frac{1}{x}\sin x dx$ will be positive.



(continued, next page)

For each question, circle the answer (1 to 5) that best fits your thinking about the argument.

A. Do you feel that you understood the argument that was presented?

1	2	3	4	5
There are fundamental details of the argument that I don't understand				I understand the argument completely

B. Are you convinced by this argument?

1	2	3	4	5
Not convinced at all				Completely convinced

C. Does this argument explain why the assertion is true?

1	2	3	4	5
No, the argument does not explain why the assertion is true				Yes, it really illuminates why the assertion is true

D. Would you consider this argument to be a mathematical proof?

1. ______ Yes, I consider this argument to be a fully rigorous mathematical proof.
2. ______ Yes, I consider this argument to be a proof, although not fully rigorous.
3. ______ No, I think this argument does not meet the standards of a proof.
4. ______ Not sure, because I don't fully understand the argument.

Please explain your reasoning about **why** you think **this is** or **is not** a mathematical proof.

Argument 7

Claim: Let n be a natural number. If n^2 is divisible by 3, then n is divisible by 3.

Line:

1 We need to show that n is divisible by 3.
2 If n is divisible by 3, then there exists an integer k such that $n = 3k$.
3 $n^2 = (3k)^2 = 9k^2$.
4 So n^2 is divisible by 9.
5 All numbers divisible by 9 are also divisible by 3.
6 So if n^2 is divisible by 3, then n is divisible by 3.

For each question, circle the answer (1 to 5) that best fits your thinking about the argument.

A. Do you feel that you understood the argument that was presented?

1	2	3	4	5
There are fundamental details of the argument that I don't understand				I understand the argument completely

B. Are you convinced by this argument?

1	2	3	4	5
Not convinced at all				Completely convinced

C. Does this argument explain why the assertion is true?

1	2	3	4	5
No, the argument does not explain why the assertion is true				Yes, it really illuminates why the assertion is true

D. Would you consider this argument to be a mathematical proof?

1. ______ Yes, I consider this argument to be a fully rigorous mathematical proof.
2. ______ Yes, I consider this argument to be a proof, although not fully rigorous.
3. ______ No, I think this argument does not meet the standards of a proof.
4. ______ Not sure, because I don't fully understand the argument.

Please explain your reasoning about **why** you think **this is** or **is not** a mathematical proof.

Argument 8

Claim: Let $f(x)$ be a real valued function and let a and b be real numbers such that $b > a$.

Then $\int_a^b |f(x)|dx \geq \int_a^b f(x)dx$

Line: (Proof by cases).

1 Either $f(x) \geq 0$ or $f(x) < 0$.

2 Case 1: $f(x) \geq 0$.

3 If $f(x) \geq 0$, then $|f(x)| = f(x)$.

4 Thus, $\int_a^b |f(x)|dx = \int_a^b f(x)dx$.

5 Case 2: $f(x) < 0$.

6 If $f(x) < 0$, then $\int_a^b f(x)dx \leq 0$.

7 Since $|f(x)| > 0$, then $\int_a^b |f(x)|dx \geq 0$.

8 So $\int_a^b |f(x)|dx \geq 0 \geq \int_a^b f(x)dx$.

9 Thus, $\int_a^b |f(x)|dx \geq \int_a^b f(x)dx$.

For each question, circle the answer (1 to 5) that best fits your thinking about the argument.

A. Do you feel that you understood the argument that was presented?

1	2	3	4	5
There are fundamental details of the argument that I don't understand				I understand the argument completely

B. Are you convinced by this argument?

1	2	3	4	5
Not convinced at all				Completely convinced

C. Does this argument explain why the assertion is true?

1	2	3	4	5
No, the argument does not explain why the assertion is true				Yes, it really illuminates why the assertion is true

(continued, next page)

D. Would you consider this argument to be a mathematical proof?
1. ______ Yes, I consider this argument to be a fully rigorous mathematical proof.
2. ______ Yes, I consider this argument to be a proof, although not fully rigorous.
3. ______ No, I think this argument does not meet the standards of a proof.
4. ______ Not sure, because I don't fully understand the argument.

Please explain your reasoning about **why** you think **this is** or **is not** a mathematical proof.

Argument 9

Claim. Let $f(x) = \ln x$. Then $f(x) \to \infty$ as $x \to \infty$.

Line:

1 Let a and b be positive real numbers with $a > b$.
2 Dividing both sides by b gives:
3 $a/b > 1$ (since b is positive).
4 $\ln(a/b) > 0$ (since $\ln x > 0$ when $x > 1$)
5 $\ln(a) - \ln(b) > 0$ (by the rules of logarithms)
6 $\ln(a) > \ln(b)$
7 Hence, for positive reals a and b, if $a > b$, then $f(a) > f(b)$.
8 Therefore, $f(x) \to \infty$ as $x \to \infty$.

For each question, circle the answer (1 to 5) that best fits your thinking about the argument.

A. Do you feel that you understood the argument that was presented?

1	2	3	4	5
There are fundamental details of the argument that I don't understand				I understand the argument completely

B. Are you convinced by this argument?

1	2	3	4	5
Not convinced at all				Completely convinced

(continued, next page)

C. Does this argument explain why the assertion is true?

1	2	3	4	5
No, the argument does not explain why the assertion is true				Yes, it really illuminates why the assertion is true

D. Would you consider this argument to be a mathematical proof?

1. _______ Yes, I consider this argument to be a fully rigorous mathematical proof.
2. _______ Yes, I consider this argument to be a proof, although not fully rigorous.
3. _______ No, I think this argument does not meet the standards of a proof.
4. _______ Not sure, because I don’t fully understand the argument.

Please explain your reasoning about **why** you think **this is** or **is not** a mathematical proof.

10. Please provide some information about your personal and math background. These data help us check that we are gathering answers from a diverse group of students. Please check the choice that fits you best.

a) Ethnicity (check one)
_____ Hispanic or Latino/a
_____ Not Hispanic or Latino/a

b) Race (check one or more)
_____ American Indian or Alaskan Native
_____ Asian
_____ Black or African American
_____ Native Hawaiian or other Pacific Islander
_____ White
_____ Other; please specify: _________________

c) Gender (check one)
_____ female
_____ male

d) Class year (check one)
_____ First-year
_____ Sophomore
_____ Junior
_____ Senior
_____ Other, please specify: __________

e) Academic major (check one or more)
_____ mathematics
_____ natural science; please specify: __________
_____ engineering; please specify: __________
_____ non-science; please specify: __________

f) Are you preparing to become a K-12 teacher? (circle one)

Yes, elementary Yes, secondary No Maybe

g) What college math courses have you taken **before** this course ? List course names or numbers.

h) What **other** college math courses are you taking **this fall**? List course names or numbers.

i) What grade do you expect to receive in this course? (check one)

A+ __________ B+ __________ C+ __________ D __________
A __________ B __________ C __________ F __________
A- __________ B- __________ C- __________
Other (please explain):

Exhibit 5A2: Instructor Ratings

Dear IBL Instructor,
We seek your assessment of individual students' learning from your course. We will compare your views with students' self-assessment of their learning from this course, in order to understand how well students' self-judgments of their mathematical learning correlate with the assessment of experienced mathematicians familiar with their work.

We ask you to give two ratings for each student: his/her overall level of expertise in mathematical knowledge and thinking at the START of the course, and his/her overall GAIN or improvement in mathematical knowledge and thinking by the END of the term. We ask you to distinguish, as best you can, students' *incoming* ability (which may depend on their educational background, preparation, and talent) from their *learning* in your course. Both of these can be factors in students' final performance or grade, so we ask you to separate them as best you can based on your observations of their work for your course.

Please focus on students' overall mathematical knowledge and thinking, even though you may have observed other kinds of abilities, growth, or learning gains in your course. And feel free to add any comments on this sheet, if you wish.

Please rate your students in comparison with other students in this course or at this level in your program. Use this scale for both ratings:

5 = very high, 4 = high, 3 = moderate, 2 = low,
1 = very poor or strongly lacking

First two letters of first name	INITIAL expertise in mathematics	Overall learning GAIN in mathematics from your course

Appendix A6
Study Methods: Academic Records Data

A6.1 Introduction

The academic records study was designed to analyze effects of IBL instruction on longer-term student outcomes, such as academic grades, course-taking patterns, and pursuit of mathematics major. We sought to address the following research questions:

- What are the student outcomes from IBL instruction, as measured by grades, course-taking patterns, and academic major status?
- How do these student outcomes differ between IBL and non-IBL students?
- How do these student outcomes vary among student sub-groups?

Academic grades are a traditional and standard way of measuring academic achievement. While instructor standards may differ in their methods of assessing learning and assigning grades, grades are nevertheless widely seen to hold a stable meaning across institutional contexts. Moreover, grades may reflect long-term changes in students' abilities and achievement, when improved learning habits and analytical thinking carry over to later courses. Thus, in this study we use grades as a longitudinal proxy for academic achievement. We use number of subsequent math courses taken and math major status as proxies of student interest and motivation to study mathematics, under the assumption that course and major choices reflect students' academic, career, and personal interests.

In Chapter 6, we discussed the principles involved in developing appropriate and comparable samples for academic records. Here, we record the specific procedures used in sufficient detail that they could be reproduced in future studies.

A6.2 Study samples

Three campuses participated in the academic records study. Altogether, we obtained 6897 student academic records, from six courses at three campuses. We chose as our "targets" core IBL courses that had a well-established history, sizable enrollments, and available comparison groups. We focused on the sections taught far enough back in time that students who took them had since had an opportunity to graduate, hence having taken all the subsequent courses they wanted or needed to take. Thus, we traced back in time to identify "target" sections and the students who took them. This defined the study sample. We requested the academic records for those students and for students in comparison non-IBL sections of the same courses.

The opportunities for academic records study varied substantially from one IBL Center to the next. Here we document in detail our rationale and procedures for defining target courses, comparison groups, and preliminary investigations to verify the suitability of these populations for further study. While all these decisions were made using common principles, the specifics of each situation depended strongly on local curriculum and academic policies, as well as on the offerings at each IBL Center.

A6.2.1 Study Populations at University L

At University L, we selected four IBL courses as our targets: two mid-level courses and two upper-level courses. In all cases the comparison groups were composed of students from the non-IBL sections of these courses taught in the same term and also from sections taught before establishment of IBL center. In the analysis in Chapter 6 we discuss outcomes for two of these courses, one mid-level and one upper-level course, due to methodological and conceptual limitations present in the other two cases. From preliminary analyses, as well as from conversation with instructors, we found that the excluded mid-level course had strong issues of student selection, so the student populations in IBL and non-IBL sections were not very comparable. The excluded upper-level course was not analyzed because it is positioned rather far in the curriculum. Hence students do not have much need or many opportunities to take further math courses, and thus the data on their later grades is modest as well as un-revealing. The mid-level class that was included in the analysis hereafter is referred to as L1, and the upper-level course as L2.

L1 is intended as a transition to proof course, helping students shift from the problem-solving approach of calculus to the rigorous proof approach of advanced courses. It is a possible required course choice for most mathematics majors, some science and engineering students, and for students preparing to be high school mathematics teachers. Both IBL and non-IBL sections of L1 are taught in sections of 25-30 students by a mix of permanent and postdoctoral faculty. The IBL sections have TAs, supported by the IBL Center's grant, who assist in class and hold office hours. The non-IBL sections generally do not have a TA, although they may have a grader who reviews student work but does not attend class. IBL sections are not labeled as such in the university course schedule, although some students may know by word of mouth which instructors will offer an IBL course and select accordingly. We know from student interviews that some students do this, and others have no idea that they are enrolled in an IBL section until they arrive in class the first day. For these reasons, we believe the selection issues to be modest but not absent in this case.

We requested academic records for students who took L1 from Fall 2001 to Spring 2008, intending the sections to be far enough back that all the students have an opportunity to graduate. Over these semesters, we obtained 1781 student academic records. However, we found that students on this campus often repeatedly took the same course, even if they had achieved a passing grade. Campus policies allow them to repeat a course (and presumably, to pay tuition) if they are dissatisfied with their grade. Thus, 201 students in this data set had taken L1 more than once. Since different attempts at L1 could put the same student in both IBL and non-IBL categories, we decided to select for analysis only those students who took L1 for the first time and received a grade (as opposed to withdrawing). This final data set for L1 comprised 1341 distinct academic records: 1130 for students who enrolled in 60 non-IBL sections and 211 for students who enrolled in 12 IBL sections. These students mostly took L1 as sophomores (23%), juniors (28%), and the largest portion as seniors (39%). The course population was largely male (71%) and white (52%), with sizable portions of Asian (21%) and Hispanic (12%) students. About 60% of the students were math majors.

L2 is a more advanced course, intended to be taken later in University L's mathematics curriculum. This particular course is not required of all mathematics majors but does meet a topical requirement for the mathematics major. Like L1, both IBL and non-IBL sections are taught in sections of 20-30 students by permanent and postdoctoral faculty, and like L1, the IBL sections generally have TAs. Again, IBL sections are not labeled in the university course schedule. For these reasons, we judge student selection issues to be relatively modest in this course.

We obtained academic records for students who took L2 from Fall 2002 to Spring 2008. Overall, we collected 1259 academic records. Similarly to L1, many students (176) took L2 repeatedly, making it hard to distinguish between IBL and non-IBL outcomes. Thus, we selected for our analysis only those students who took L2 for the first time and received a passing or failing grade. Thus the final data set for this course consists of 909 academic records: 786 for students who enrolled in 42 non-IBL sections and 123 for those who enrolled in 9 IBL sections. These students mostly took L2 as juniors (26%) and seniors (52%). The population of this course was also predominantly male (65%) and mostly white (51%), with sizable portion of Asian (19%), Hispanic (16%), and foreign students (10%). A majority of students (71%) were math majors.

Since we analyzed the data for all the students who enrolled in L1and L2 in the time frames of interest, as opposed to a sample of the students, the results of our analysis carry more statistical power.

A6.2.2 Study Populations at University G

At University G, we selected the first course of a three-term introductory-level sequence as our target. The comparison group was composed of students from non-IBL sections taught in the same term. We obtained academic records data for students who took the target course, hereafter referred to as G1, from Fall 2004, Fall 2005, and Fall 2006. Over these three semesters, we collected 962 academic records for this course: 913 for students who enrolled in 6 non-IBL sections and 49 for those who enrolled in 2 IBL sections.

However, the non-IBL students were not directly comparable to the IBL sample. The IBL sections of G1 are honors courses: students are invited to join based on a record of excellent academic performance in mathematics, effectively populating these sections with a select group of high-achieving and self-motivated students. Non-IBL sections, on the other hand, include students of all levels of ability and achievement; they are taught in traditional large lecture format to several hundred students at a time, with separate recitation sections of 25-30 students taught by TAs. For example, the biggest portion of IBL G1 students (53%) scored in the 701-800 bracket (on a scale of 200-800) on their college entrance math tests (SAT score or converted

ACT score[1]). In comparison, the biggest portion of non-IBL students (45%) placed in the next bracket down, 601-700 points—a good but not outstanding score. Students admitted to the IBL sections also on average had higher high school GPAs than non-IBL students, took G1 earlier (often in their first semester of college), and pursued mathematics majors in higher numbers. Thus, the IBL and non-IBL populations were not directly comparable.

A6.2.3 G1 Sampling Procedures

In order to make a valid comparison between IBL and non-IBL G1 students, we turned to sampling. We experimented with selecting a group of high-achieving students from non-IBL sections that would closely match the academic backgrounds and demographics of the IBL students. For that purpose, we requested additional data from the registrar on students' high school grades and college admissions test scores, and matched students on several criteria. Since existing literature points to both high school GPA and college admission scores as equally important predictors of college success (Hoffman & Lowitski, 2005; Noble, 1991), we created our own pre-college index combining these two quantities, thus allowing us to match the students on both without having to prioritize one over the other. The pre-college index was calculated for each student using the formula:

1075*high school GPA + 4*mathematics college admission score + 4*verbal admission score.

The new index ranged from 7,500 to 11,000 and was divided into 500-point brackets. One IBL student whose record included no college admission test scores was omitted from the study.

For each remaining IBL student, we selected two non-IBL students who fell into the same index bracket, had an academic major or intended major from the same category (math, science, liberal arts and sciences, or undeclared), had the same academic status (freshmen, sophomore, junior, senior), was of the same gender, and of the same race/ethnicity—in that order of priority. In most cases, we were able to select two non-IBL students who matched the IBL student on all the criteria, but in some cases it was not possible. When we were unable to find a match on all the levels, we relaxed the race and sometimes the gender criteria. In several cases where gender seemed especially important to match—there were women among the top achievers in IBL sections, but not in the non-IBL—we slightly widened the index brackets to find a match of the same gender and comparable (if slightly lower) achievement. Two students did not have high school GPA on record, and thus we could not compute our index; here we used math and verbal scores in college admission tests as the primary means of matching. Overall, such close, if sometimes creative matching, ensured high level of similarity between the IBL and non-IBL students in the sample and their definite comparability.

This sampling process resulted in the final G1 data set of 147 student academic records: 49 for students from IBL sections and 98 for students from non-IBL sections who closely matched the IBL group. These students mostly took G1 in their freshman (72%) and sophomore (24%) year.

---

[1] The ACT to SAT conversion procedures are discusses in Section *A6.4.4*

They were mostly male (62%) and white (64%), with sizable portions of Asian (18%) and Hispanic (10%) students. The majority of sample students were non-math majors (82%).

A6.2.4 Historical Sampling at University W

We selected an upper-level course, "W1," as our target at University W. The course is intended to serve as a transition to the rigorous proof approach of advanced math courses for sophomores and juniors. It is taught in small sections of 20 students and meets a requirement for the mathematics major, but is offered only in IBL format.

We obtained academic records data for students who took IBL-only W1 from Fall 2004 to Spring 2008. Since a contemporaneous comparison section was not available, we experimented instead with a historical comparison group, obtaining data from the same course taught prior to the establishment of the IBL Center—from Fall 2001 to Spring 2003. However, we learned that this course was in fact an ancestor to the IBL efforts at this campus, and was taught using practically the same methods as the current course. Furthermore, this was a new course in 2001, so no early historical data were available. An initial analysis of student demographics and outcomes confirmed the high similarity of the courses designated "IBL" and early versions not so designated. Thus, the historical comparison was not appropriate in this case, and we do not further discuss student outcomes from this course. However, this approach could be suitable in other studies of teaching innovation.

A6.3 Obtaining Raw Data

Once the target course and comparison group were identified, we requested de-identified student data from the university registrar or institutional records office. We requested data from all four campuses; three provided data within one to four months and one did not respond to multiple requests. Two campuses deemed our request as complying with FERPA, the federal laws governing student privacy (U.S. Department of Education). A third requested that we file a FERPA research exemption with university counsel, which was granted.

The de-identified raw data requested from campuses included the following:

- List of all mathematics courses taken by each student, by academic term and year, with grades
- Section number (or instructor name, if that is how they are designated) for all math courses
- Current class status (junior, senior etc.) if not graduated
- Graduation year and degree earned, if graduated
- Current GPA (or final GPA for students who have graduated)
- Current academic major(s) and minor(s), and record of changes in major/minor
- Gender
- Ethnicity, race, citizenship

- Overall admissions index score if the campus uses one, or high school GPA and test scores (SAT or ACT) if overall index is not used or unavailable.

In several cases we interacted with university officials to refine course selection or ascertain which institutional variables would best meet our needs. Raw data were received in spreadsheet form in formats specific to the institutional records system; they were compiled, cleaned, and converted to standardized formats by our research team.

In Section A6.4, we discuss the construction of analytical variables from these raw data. Common principles were used to construct the variables for each test case, but the exact procedures for construction differed due to characteristics of the course itself, the departmental curriculum, and institutional academic policies. In interpreting raw data and making these decisions, we consulted the mathematics department web sites and university registrar web sites for information such as course requirements and sequences, the meaning of grades issued for course incompletion and withdrawal, and course repetition policies. Campus leaders helped to identify IBL sections and instructors. Staff in the campus institutional research, advising or registrar's offices were gracious in answering our questions about notation and anomalies.

A6.4 Construction of Variables

Because institutional records varied widely in format, we converted the raw data into standardized variables to count course enrollments in particular, later math courses and to compute GPAs for these courses. We use enrollment in later courses as a proxy for student motivation to take more mathematics, and the GPAs for those courses as a proxy for subsequent achievement and learning in mathematics. Adding or dropping the math major is taken as another measure of student motivation or persistent interest in mathematics. The number of courses taken and GPA obtained *prior* to the target course are taken to serve as proxies of mathematical background and prior academic achievement, respectively. High school GPA and college admission test scores are used as other measures of prior achievement.

We looked separately at student outcomes for several time periods: the target course itself; the "next" term, i.e. the semester or quarter[2] immediately following the target course (which may be most sensitive to the impact of an IBL experience), and the cumulative record for all courses taken after the target course and up to graduation. Because there may be differences in how students select or approach learning in required vs. elective courses, and in IBL or non-IBL courses, we constructed variables to examine grades and enrollment in each of these subsets of courses. Thus there were multiple variables to analyze for both grade and course count measures.

We used Boolean logic functions within Microsoft Excel software to categorize and combine the raw variables into the analytical variables of interest, followed by extensive hand-checking to check that logic was applied correctly, and to handle anomalous cases.

[2] For simplicity, we use the term 'semester' to label variables for both semester- and quarter-based courses.

In the sections below, we detail how the analytical variables were constructed and labeled.

A6.4.1 Course Count Variables

The variables that count numbers of courses taken focus on the new courses completed, and thus exclude the repeat attempts at the same course if the passing grade was initially earned. Repeated courses are retained if the initial attempt ended in withdrawal or a failing grade. The course counting variables include:

- Number of <u>prior</u> math courses—count of math courses taken prior to the target course.
- Number of <u>subsequent</u> math courses—count of overall math courses taken after the target course. Sometimes abbreviated as num of subs math courses in Chapter 6 and below.
- Number of subsequent <u>required</u> courses—count of required courses taken after the target course. Required courses are the core courses necessary for graduation with the math degree at the particular campus. Abbreviated as num of subs req courses.
- Number of subsequent <u>elective</u> courses—count of elective courses taken after the target course. Elective courses are all courses other than the core courses established as required for the mathematics major. At University L there are two required Analysis courses to choose from; if both were taken, we counted the first as required and the second as elective. Abbreviated as num of subs elect courses.
- Number of subsequent <u>IBL</u> courses—count of IBL-method courses taken after the target course. IBL courses are designated as such by each campus (see Chapter 1). Abbreviated as num of subs IBL courses.

For all count variables, some mathematics courses were excluded from the counts. For example, we omitted the courses prior to Calculus III at University W. This was intended to ensure that weaker students who may have taken more introductory courses on their way to W1 do not appear to be more experienced or mathematically mature, simply because they have taken a larger number of courses, while students who have tested into a higher-level course appear to have less background. For all the campuses, we excluded one-credit courses because counting them as measure of experience or motivation might be misleading, especially if comparing to full mathematics courses.

We requested the data for course sections far enough back that all the students would have had the opportunity to graduate. However, on two campuses—University G and L—a sizable portion of the students had not in fact graduated by the time of data collection. Thus, to ensure a fair comparison of courses and grades for different students, we decided to level the playing field by only analyzing courses completed within a set period of time after the target for all students. For each student, we selected enrollment and grade data for the first two years after the target course and constructed our count and average grade variables based only on that period of time. For W1, where the vast majority of students graduated by the time of data collection, it was unnecessary to apply this procedure in order to fairly compare IBL and non-IBL students.

The sample sizes for all count variables are the same, because even the students who did not take particular types of courses serve as data points for the count variables (with counts of zero). The General Linear Model (GLM) procedure (discussed in Section A6.5) that was used to control for incoming differences slightly reduced the sample sizes, since it required the data for the pre-target GPA, which was missing for some students. The resulting sample sizes for the count variables are shown in Table A6.1.

Table A6.1: Sample Sizes for all Course Count Variables, by Course

Count Variables by Course	**IBL sample size**	**Non-IBL sample size**
L1	204	1077
L2	117	747
G1	47	98

A6.4.2 Average Grade Variables

The variables that compute average math grades focus on the courses for which a grade was received, whether passing or failing. Thus, these variables exclude courses taken on pass/fail bases, courses audited, and courses with any grades other than A, B, C, D, and F. Differently from the count variables, since valid grades in all courses, even repeated attempts at the same course, should factor into the measure of achievement, we include grades for the repeats in our calculation of average grade variables. The average math grade variables include:

- Average <u>prior</u> grade—average grades for math courses taken prior to the target course.
- Target course grade—simply the grade received in the target course itself, whether an IBL or non-IBL section.
- <u>Next</u> semester average grade—average of math grades for courses taken the semester immediately following the target course. Sometimes abbreviated as next sem avg grade in Chapter 6 and below.
- Average grade in subsequent <u>required</u> courses—average of the grades received in all required math courses taken after the target course. At University L there are two required Analysis courses to choose from; if both were taken, we counted the first as required and second as elective. Abbreviated as avg grade in subs req courses.
- Average grade in subsequent <u>elective</u> courses—average of the grades received in all elective math courses taken after the target course. Abbreviated as avg grade in subs elect courses.
- Average grade in subsequent <u>IBL</u> courses—average of the grades received in all IBL-method math courses taken after the target course. Abbreviated as avg grade in subs IBL courses.

All the variables discussed so far exclude from the calculation any courses that resulted in withdrawal or an 'incomplete' grade. It would not be appropriate to include those courses in either counting the number of new courses completed or averaging grades for the courses taken. It is important to mention that because count variables exclude repeated attempts at the same course while average grade variables include them in the calculation, the two types of variables are somewhat disconnected. That is, the number of courses that serves as the denominator in calculating the grade averages may be different from the corresponding count variable, because the former includes the repeated courses and latter does not.

While this general approach to course repetition was applied across all campuses analyzed, the issue of repeated *target* courses was handled differently on each campus. This issue was especially problematic for University L, because students could end up taking both an IBL and a non-IBL section in their repeated attempts at L1 and L2, and a surprising number did. Thus, we focused only on students' first attempts at these courses. Outcomes for repeated courses form a special subset of the data that we intend to analyze in the future. On the other hand, we focused on the last attempt in G1 and W1, where repeats were rare and never crossed the IBL/non-IBL line. At these campuses, where course repetition policies were less generous, we interpreted the need to repeat as 'failure' in previous instances, and thus focused on the successful attempt.

For the average grade variables, the sample sizes differed from variable to variable, since not all students took all types of courses and thus acquired grades in them. There are a lot of missing cases, especially for the IBL grades, since most non-IBL students did not take any IBL courses. Again, the GLM procedure further reduced the sample sized for the post-target average grade variables, since it required data for the pre-target courses taken and average grades obtained in order to control for incoming differences. The resulting sample sizes for the average grades in subsequent courses are shown in Table A6.2.

A6.4.3 Math Major Variables

Besides count variables, two other measures related to student motivation were available: adding and dropping the mathematics major. These variables are categorical: the 'Adding math major' variable is set to 1 if the student switched to mathematics from another major or added mathematics as a second major. The 'Dropping math major' variable is set to 1 if a student removed mathematics from their declared majors, whether to switch to another major or just to drop mathematics and retain the other. We employed the same logic in constructing these variables across all campuses analyzed, but because of differences in the data provided by each campus we implemented the logic differently at each campus. Data from University W and University L included students' academic majors recorded for each semester they were in school. Data from University G included only students' initial major and final major, either the degree major if the student had graduated, or their current major at the time when the data were extracted from the institutional records system. Thus, the level of granularity available for tracking the major changes was different in this situation.

Table A6.2: Sample Sizes for all Grade Variables, by Course and Variable

Average Grades Variables by Course	**IBL N**	**Non-IBL N**
L1		
Target course grade	204	1077
Next semester average grade	89	526
Average grade in subsequent required courses	104	499
Average grade in subsequent elective courses	130	725
Average grade in subsequent IBL courses	30	80
L2		
Target course grade	117	747
Next semester average grade	41	326
Average grade in subsequent required courses	22	130
Average grade in subsequent elective courses	71	446
Average grade in subsequent IBL courses	7	23
G1		
Target course grade	47	98
Next semester average grade	39	68
Average grade in subsequent required courses	40	70
Average grade in subsequent elective courses	16	20
Average grade in subsequent IBL courses	35	3

As discussed, a sizable portion of G1, L1, and L2 students had not graduated by the time of data collection. Thus, we could not examine the full evolution of their major choices, and could not reasonably compare the majors of students who graduated and those who did not. Thus, similarly with the count and average grade variables, we constructed our major variables based on only the first two years after the target course. Within this time, some students, for whom academic records indicated a drop of mathematics major, still graduated with the mathematics degree. Thus, we adjusted the ‘math major drop’ variable for them to indicate no drop. Some other students started new postgraduate programs, majoring in non-mathematics disciplines, which appeared as a drop in our algorithm. After ensuring that these students graduated with the undergraduate mathematics degree, we adjusted their ‘math major drop’ variables to indicate no drop as well. This student-by-student examination of the data was required to treat all student records consistently.

In order to establish if the mathematics major was added or dropped, we had to classify the academic majors on each campus as math majors or not. For purposes of student sample description, we went even further. We categorized the majors into four groups: math, science, non-STEM liberal arts and sciences, and undeclared or unknown.

A6.4.4 Test Score Measures

Besides average math grade prior to the target course, we have two other measures of students' prior achievement or 'ability'. For W1 and G1, the academic records include high school GPA. This variable did not require any transformation and was used directly or included in construction of our index variable.

The other prior achievement measure included in the data is college admission test scores. This measure is available for all three campuses analyzed. However, depending on the campus, either SAT, ACT, or both scores were recorded. In order to directly compare the prior achievement of students who took different tests, we required a conversion mechanism. After consulting the existing literature on the subject and testing out normalized scores and regression solutions, we decided on the concordance table conversion. It is indicated as the most precise conversion method in the literature (Dorans et al, 1997; Dorans, 2004), and concordance tables are available for both mathematics and verbal scores. The concordance table provides a corresponding SAT value for each value of ACT. The math concordance table provides a direct correspondence between ACT math and SAT math scores (in the table ACT scores range from 11 to 36). The conversion is a little more complex for the verbal skills tested, because the tests cover overlapping but not identical subjects: the SAT includes one verbal score while the ACT includes an English test and a reading test. Prior research shows that there is a strong correlation between SAT verbal scores and the sum of ACT English and ACT reading scores (Dorans et al, 1997, Dorans, 1999). The concordance table for these scores provides a corresponding SAT score for each value of the sum of ACT English and ACT reading (ranges from 18 to 72 in the table).

A6.4.5 Academic Status Variable

Besides the number of mathematics courses taken prior to the target course, we constructed another variable related to students' mathematical experience before the target, their academic class status. Essentially, we wanted to know if students took the target course as first-years, sophomores, juniors, or seniors. This information was included in the data set for G1, where student class status was recorded for each semester the student was in school. But for Universities L and W, such data were not included in the institutional records provided. Thus, we had to estimate it. While guided by the same principles, we took different practical steps to estimate class status on these two campuses.

For L1 and L2, we numbered each semester the student was in school. We excluded summer semesters from the count to make sure that students who took summer classes do not appear older than those who did not; students in this study took few summer mathematics classes. Based on the semester number when the student took the target course, we classified the student as first-year, sophomore, junior, or senior. For example, to be categorized as a first-year student, the student had to take the target course in their first or second semester in school; i.e., the target course semester number would be two or less. If the graduation date indicated that student graduated before taking the target course, we assigned him or her to the graduate category.

For W1, we followed the same logic but had to implement it differently. Among W1 students there were many transfer students, who artificially appeared younger in academic status than their peers if counting their entrance to University W as the beginning of their college career . Thus, to fairly estimate their overall academic status, and thus their prior mathematics experience, we had to specifically adjust their starting point. In order to do that we computed the average time it took transfer and non-transfer students to graduate. It appeared that on average transfer students took about a year less to graduate than non-transfer students (i.e., they transferred in one year of academic credit). Hence, we estimated their college entry time as one year prior to their enrollment at University W. After applying that procedure to the transfer students, we had realistic entry terms for both transfers and non-transfers. For each student, we then subtracted the entry term from the term student took W1. This gave us time elapsed between college entry and taking the target course, and yielded the academic status in a manner similar to that for L1 and L2.

A6.5 Data analysis

To compare the means of various variables for IBL and non-IBL students we used both parametric (t-test, ANOVA) and non-parametric (Mann-Whitney, Kruskal-Wallis, Chi-square) statistical tests. As most of our data in this study was not normally distributed, according to the Shapiro-Wilk test of normality, for the final determination we relied on non-parametric statistics. This was an appropriate choice, since the non-parametric statistics are specifically suited for the non-normally distributed data. We used Excel for preliminary investigations of sample comparability and sample matching, and SPSS (version 18) to perform statistical analyses on the final samples.

We also had to use statistical techniques to control for incoming differences between IBL and non-IBL student groups. We used the number of math courses and average math grade prior to the target course as two metrics of incoming differences. For L1, IBL students had taken statistically significantly fewer prior math courses and earned statistically significantly higher average math grades prior to the target course, as compared with non-IBL students. For L2, these differences were not statistically significant. For G1, even after our close-match sampling, there was still a statistically significant difference in the number of prior math courses between IBL and non-IBL students.

We used the General Linear Model (GLM) procedure in SPSS to control for these incoming differences, including as covariates in all analyses the number of prior math courses and average math grade prior to target course. For G1, since most students (as first-year college students) did not have a pre-target GPA, we used our pre-college index (Section A6.2.3) as a covariate in

GLM to control for incoming difference in achievement. All the levels of significance we report in Chapter 6 are based on the outcomes of GLM controlling for incoming differences. We also computed estimated marginal means, which are intended to offset the effect of intervening variables (covariates) and yield mean estimations that reflect that offset. All the variable means reported in Chapter 6 are estimated marginal means computed by GLM to control for incoming differences.

A6.5.1 Analysis by Gender

We performed several analyses of student sub-groups by gender. The sample sizes are the same for all count variables on each campus, since even students who did not take particular kinds of courses provide data points for these variables. For instance, student taking zero IBL courses still provides us with some information, whereas their grades must be treated as missing data. The GLM procedure slightly reduced the sample sizes, since it required the pre-target achievement data, which was missing for some students. Thus, the sample size breakdown by IBL and gender for the count variables is given in Table A6.3.

Table A6.3: Sample Sizes for all Course Count Variables, by Course, Gender, and IBL

Count Variables by Course	IBL men N	non-IBL men N	IBL women N	non-IBL women N
L1	147	755	57	322
L2	77	477	40	270
G1	28	61	19	37

The sample sizes for average grades measures differ from variable to variable, because some students did not take certain types of courses and received no grade, creating missing data for that particular average grade variable. The GLM procedure slightly further reduces the samples sizes. The resulting sample size breakdown by IBL and gender for average grade variables is given in Table A6.4.

Table A6.4: Sample Sizes for all Average Grades Variables, by Course, Gender, and IBL

Average Grades Variables by Course	IBL men	non-IBL men	IBL women	non-IBL women
L1				
Target course grade	147	755	57	322
Next sem avg math grade	66	351	23	175
Avg grade in subs req courses	67	336	37	163
Avg grade in subs elect courses	89	507	41	218
Avg grade in subs IBL courses	22	51	8	29
L2				
Target course grade	77	477	40	270
Next sem avg math grade	30	204	11	122
Avg grade in subs req courses	17	92	5	38
Avg grade in subs elect courses	48	289	23	157
Avg grade in subs IBL courses	6	15	1	8
G1				
Target course grade	28	61	19	37
Next sem avg math grade	24	47	15	21
Avg grade in subs req courses	24	46	16	24
Avg grade in subs elect courses	9	17	7	7
Avg grade in subs IBL courses	21	3	14	0

A6.5.2 Analysis by Prior GPA

We performed several analyses where students were divided into low-, medium-, and high-achieving groups based on their average math grade prior to the target course. We focused on the relationship between prior achievement and the subsequent outcome measures after noting the strong correlation between these variables. The nonparametric correlations (Spearman's rho) and their statistical significances are shown in Table A6.5.

Table A6.5: Correlations between Outcome Variables and Prior math GPA by Course

Outcome Measures	L1		L2		G1	
	Correlation	N	Correlation	N	Correlation	N
Number of subs math courses	0.114***	1281	0.116***	864	0.189	20
Number of subs required courses	0.037	1281	0.076*	864	0.230	20
Number of subs elective courses	0.118***	1281	0.110**	864	-0.029	20
Number of subs IBL courses	-0.006	1281	0.044	864	0.259	20
Target course grade	0.555***	1281	0.586***	864	0.237	20
Next sem avg math grade	0.560***	615	0.543***	367	-0.031	10
Avg grade in subs required courses	0.567***	603	0.341***	152	0.588	10
Avg grade in subs elective courses	0.555***	855	0.573***	517	0.316	4
Avg grade in subs IBL courses	0.586***	110	0.636***	30		1

For courses L1 and L2, most of the outcome measures correlate strongly with prior math GPA, and those correlations are highly statistically significant. This is especially apparent for the average grade variables. Thus, the higher student's prior math GPA have been, the more subsequent courses he or she enrolled in and the higher subsequent grades he or she received in and after the target. Hence, it is reasonable to divide L1 and L2 students into subgroups based on their prior achievement.

In the course G1, most students were rather high-achieving, and thus there are no statistically significant correlations between prior achievement and the outcome measures. The sample sizes for these correlations are also rather small, since most of these students did not have prior math GPA as they started their college math with G1. Since we see no relationship between prior math GPA and outcome variables in this high-achieving group, it is unreasonable to divide these students into prior achievement categories. It would also be practically difficult, since most of them do not have prior math GPA .

For courses L1 and L2, we empirically created the prior GPA groups corresponding to low, medium, and high prior achievement. We broke up each distribution into tertiles—three subsets, each containing one third of the sample. We then tested each subset for differences in prior math GPA between IBL and non-IBL students. If such differences remained statistically significant within the subset, we adjusted the cut points to ensure no statistically significant difference was present. Thus, our low, medium, and high groups by prior math GPA do not contain exactly one third of the distributed sampled population, but roughly estimate thirds of the sample. The cutoff points for the subsets are different between L1 and L2, since their underlying distributions were different.

The sub-group sample sizes are the same for all the count variables on each campus, since students who did not enroll in particular types of classes still provided data (zero count) for the analysis. The GLM procedure slightly further reduced the sample sizes for L1 and L2. The sample size breakdown by IBL and prior achievement for the count variables is given in Table A6.5.

Table A6.5: Sample Sizes for all Count Variables, by Course, prior GPA group, and IBL

Count Variables by Campus	IBL Low	IBL Med	IBL High	non-IBL Low	non-IBL Med	non-IBL High
L1	49	76	79	360	353	364
L2	32	38	47	241	261	245

The sub-group samples sizes differ from variable to variable for the average grade measures, as students who did not take particular kinds of courses show up as missing grade values in those courses. The sample sizes are slightly further reduced by the GLM procedure, which required the prior GPA data that was missing for some students. The sample size breakdown by IBL and prior achievement for the average grade variable is given in Table A6.6.

Table A6.6: Sizes for all Average Grade Variables, by Course, prior GPA group, and IBL

Average grades in subsequent math classes	IBL Low	IBL Med	IBL High	non-IBL Low	non-IBL Med	non-IBL High
L1						
Target course grade	49	76	79	360	353	364
Next sem avg math grade	18	36	35	186	184	156
Avg grade in subs req courses	24	39	41	180	162	157
Avg grade in subs elect courses	30	45	55	235	253	237
Avg grade in subs IBL courses	7	8	15	37	21	22
L2						
Target course grade	32	38	47	241	261	245
Next sem avg math grade	17	11	13	112	119	95
Avg grade in subs req courses	14	5	3	76	39	15
Avg grade in subs elect courses	24	22	25	145	158	143
Avg grade in subs IBL courses	3	3	1	10	9	4

A6.6 Comments on the Strengths and Limitations of the Academic Records Study

Overall, conducting academic records analysis proved to be a complex and very labor-intensive task. This is especially the case because the suitable opportunities for valid comparison are few and far in between. Most data sets we obtained for this study had some limitations with respect to comparison samples, which we had to address and overcome. For example, we addressed G1 institutional selection with an intricate sampling scheme, which essentially described academic and demographic characteristics of each IBL student and then selected two non-IBL students that matched those criteria. This was a very laborious and time-consuming process.

Another difficulty with this kind of study is the institutional differences between campuses. The institutional cultures differ in many ways: linearity of curriculum, enforcement of prerequisites, lenience towards retaking courses after earning a passing grade, and many other aspects. The granularity of the academic data collected by the institutional records offices also differ from campus to campus. Thus, while we adhered to a general, consistent logic in the analysis of different data sets, we had to implement it in very specific and circumstantial ways for each campus or class data set. This, again, added to the complexity, detail, and effort involved in this analysis.

Institutional collaboration is another difficulty with this kind of analysis. While we requested completely anonymous academic data, some campuses had concerns about student privacy. We addressed those concerns for most institutions and successfully obtained the academic records data. However, one campus did not provide institutional data for either concerns over privacy or other undisclosed reasons. This was a lost opportunity, since the data set we requested from that campus would have provided an opportunity to examine potentially well-matched honors sections of an introductory course, presumably without the need to construct a matched non-IBL sub-sample and therefore with better statistical power.

The data cleaning and construction of variables took a lot of effort, as lots of hand-checking was required due to non-linear and unconventional paths many students took through their curriculum. These unexpected academic routes often defied our expectations and assumptions. More data redundancy and background information would have been helpful in faster making sense of these various paths and outcomes. Also, software tools more sophisticated than Microsoft Excel may have sped up the data cleaning and variable construction. Thus, if embarking on this kind of analysis again, we would request more data with higher level of redundancy and would use more agile data analysis tools.

A6.7 References Cited and Other Resources Consulted

"ACT : 2001 ACT National and State Scores : Table 4." *ACT*. Web. 14 Apr. 2011. <http://www.act.org/news/data/01/t4.html>.

"A Comparison of the SAT I, SAT II, and ACT." *Education.com | An Education & Child Development Site for Parents | Parenting & Educational Resource*. Web. 14 Apr. 2011. <http://www.education.com/reference/article/Ref_Different_Tests_Same/>.

"ACT : 2001 ACT National and State Scores : Table 4." *ACT*. Web. 14 Apr. 2011. <http://www.act.org/news/data/01/t4.html>.

Dorans, N.J. (1999). Correspondence between ACT and SAT I Scores. *College Board Research Report 99-1*. New York, NY: The College Board.

Dorans, N.J., Lyu, C.F., Pommerich, M., & Houston, W.M. (1997). Concordance between ACT Assessment and recentered SAT I Sum scores. *College and University, 73* (2), 24-35.

Dorans, N.J. (2004). Equating, Concordance, and Expectation. *Applied psychological Measurement, 28* (4), 227-246.

"Family Educational Rights and Privacy Act (FERPA)." *U.S. Department of Education*. Web. 13 Apr. 2011. <http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html>.

Grove, Wayne A., Tim Wasserman, and Andrew Grodner. "With So Many Proxies (for Student Aptitude), Which Is a Scholar to Choose? Does It Matter?" *Unpublished Manuscript. Lemoyne College* (2003). Web. <http://agrod.com/sapaper103.pdf>.

"Guidelines on Uses of College Board Test Scores and Related Data." *College Board*. Web. 14 Apr. 2011. <http://professionals.collegeboard.com/profdownload/guidelines-on-uses-of-college-board-test-scores-and-data.pdf>.

Hilton, Margaret L. *Protecting Student Records and Facilitating Education Research: a Workshop Summary*. Washington, D.C.: National Academies, 2009. Print.

Hoffman, J. L., & Lowitzki, K. E. (2005). Predicting college success with high school grades and test scores: Limitations for minority students. *The Review of Higher Education* 28.4 (2005) 455-474.

Noble, J. P. (1991). Predicting college grades from ACT assessment scores and high school course work and grade information. Retrieved from http://act.org/research/researchers/reports/pdf/ACT_RR91-03.pdf

"SPSS Statistics 18." *IBM*. Web. 14 Apr. 2011. <http://www-01.ibm.com/software/data/spss/products/statistics/statistics-desktop/>.

Appendix A7
Study Methods for Student and Instructor Interviews

A7.1 Introduction

We conducted semi-structured, in-depth interviews with samples of faculty members, graduate students, and undergraduates at all four institutions. Unlike surveys, academic records analysis, or other quantitative research methods, interviews allow participants to offer detailed, spontaneous accounts of their experiences, views and attitudes, as well as explanations and evidence supporting their observations. As well, the conversational style of interviews allows the researcher to probe participants' comments more deeply for clarification or to better understand their basis. Participants can raise new issues, or emphasize points important to them. Thus, the instructor and student interviews complement the types of data collected in the other sub-studies, offering data that can help to explain, confirm, or validate findings from the other sub-studies.

A7.2 Instructor interviews

The instructor interview study was designed to gather instructors' perspectives on student outcomes, classroom learning and teaching processes, and their own experiences. We sought to address the following questions:

- What student gains (or failures to make gains) do instructors observe as a result of IBL instruction?
- How do instructors describe their classroom practices (including planning, course materials, assignments, assessments, teaching methods, etc.), with what rationale(s) behind them? What teaching issues do IBL instructors face, especially issues that are different in nature or scale from those in non-IBL classrooms, and how do they resolve them? What advice do they offer to other IBL instructors?
- What are the costs and benefits to instructors of teaching with IBL methods?
- What are the personal, professional, departmental and institutional conditions under which IBL courses are effective and sustainable, or not? What professional resources are needed to begin using IBL methods, and how (if at all) do instructors obtain them?

Graduate students played an important role in many of the IBL courses we studied, sometimes as teaching assistants and others as lead instructors. Here we use the generic term "instructor" to include all those in instructional roles. We specify "faculty" (including anyone in the lead instructor role, regardless of their appointment type) or "TA" when it is important to distinguish specific classroom roles.

A7.2.1 Instructor interview sample

All four campuses participated in the instructor interview study. Based on information from the campuses, we prepared a list of all active or previous instructors of IBL courses in the past three years, and all graduate student TAs who had participated in an IBL course in the same period. At some campuses, this included graduate students or postdocs who had left campus but who were teaching elsewhere and could be located. This yielded a list of 10 to 19 individuals per campus. We excluded two individuals for logistical reasons (e.g., on sabbatical overseas) and invited all the rest, by e-mail, to interview with us in person during a scheduled visit to their campus. For those who had left campus, we invited them to a telephone interview. Telephone interviews were also conducted with a few instructors who were unavailable during our campus visit.

This approach yielded five to 15 interviews from each campus, for an overall response rate of 77% (varying from 50% to 88% by campus). Altogether, 43 interviews were conducted with 44 individuals (one pair of TAs for the same course was interviewed as a focus group). Three digital recordings exhibited severe interference. After digital editing to reduce noise using Audacity software, two were able to be transcribed, but one was unsalvageable and no transcript could be prepared. Thus the final sample for analysis included 42 interviews with 43 individuals.

Of the 43 instructors interviewed, 31 were men and 12 women. Nearly all were white. Approximately 15% were born outside the US. Most of the instructors taught courses for math or STEM majors (labeled as “math-track” courses elsewhere); seven of the 42 primarily taught IBL courses for pre-service teachers.

Those in faculty positions included 23 holding tenurable (pre- or post-tenure) and non-tenure-track appointments, including both long-term lectureships and short-term postdoctoral positions. About half of these were new to IBL methods at the time we spoke, and the remainder had one year to decades of prior IBL experience. The faculty sub-sample was predominantly male.

Twenty interviewees were graduate students at the time of their IBL involvement. They ranged from second- to seventh-year students, and a few had graduated and moved on to postdoctoral or tenure-track teaching positions. Their IBL teaching experience ranged from one term to several years. The graduate student sub-sample was nearly gender-balanced.

A7.2.2 Instructor interview protocol

The interview protocol addressed participants’:

- Career stage, teaching experience, and (for graduate students and postdocs) future career plans
- Role in the IBL course, and how they became involved
- Teaching philosophy, style, and strategies
- Observations of student gains, including gains in knowledge, understanding, skills, approach to problem-solving or learning, and later impacts (such as later course choices or career selection)
- Observations of any differences in who achieves these gains—by race/ethnicity and gender, and by characteristics such as work ethic, temperament, intellect

- Costs and benefits to the instructor him/herself of teaching with IBL methods
- Influences of their IBL teaching experiences on their teaching philosophy, classroom style, or beliefs
- Experiences of learning to teach with IBL methods, including difficulties, professional development, and collegial support (or lack of it), and advice they would give to students preparing to take an IBL class, and for colleagues preparing to teach one.

With early-career interviewees, we asked about the impact of IBL teaching on their career interests and prospects. With senior faculty and department chairs, we asked about their departmental colleagues' perspective on IBL and their views of the IBL program's sustainability in their department.

A7.3 Undergraduate interviews

Interviews with undergraduates investigated students' perspectives on their experiences in IBL mathematics courses, outcomes of IBL instruction, and classroom learning and teaching processes. Thus, we sought to address the following questions:

- What do students gains (or not) as a result of IBL instruction?
- How do students describe their instructor's classroom practices (including course materials, assignments, assessments, teaching methods, etc.)? How do these differ from more traditionally-taught mathematics courses? How does this affect their learning? What advice do they offer to other students taking an IBL course?
- What difficulties do students encounter when learning with IBL methods? What resources are needed or available to help them overcome these difficulties?
- Do some students benefit more than others from IBL methods?

A7.3.1 Undergraduate interview samples

At three campuses, undergraduate interview samples were constructed from comprehensive lists provided by each department of all students currently enrolled in an IBL mathematics courses. Student samples were selected so as to balance the representation of demographic characteristics such as gender, race/ethnicity, and major. Interviews with students were solicited and scheduled by e-mail. This approach yielded about 20 interviews at each campus. At the fourth campus, students currently taking an IBL course were contacted by an academic advisor and invited to contact us to volunteer for an interview. This approach yielded a rather smaller interview sample from this campus, which is not necessarily represenitive.

The final student interview sample was composed of 68 students who had participated in an IBL mathematics course in the current or prior semester at one of the four IBL Centers during the 2008-10 academic years. Nineteen students were interviewed individually; the remaining 49 students participated in one of 22 focus group interviews, with 2-4 students each. Students who interviewed together in a focus group had taken the same IBL mathematics course and so were able to offer comments about the same class.

Demographic data were reported by students themselves on a standardized data sheet, except for course type, which was categorized by the research team. Table A7.1 provides an overview.

Table A7.1: Distribution of undergraduate interviewees by institution, gender, race/ethnicity, discipline, type of mathematics course, and interest in teaching

Demographic group	Number	Percentage
By gender	**68**	**100%**
Female	38	56%
Male	29	43%
Did not respond	1	1%
By race and ethnicity	**68**	**100%**
White, not of Hispanic origin	48	71%
White, of Hispanic origin	6	9%
Black, not of Hispanic origin	1	1%
Asian	9	13%
Multiple races	2	3%
Did not respond	2	3%
By major	**68**	**100%**
Mathematics	35	51%
Science	14	21%
Engineering	2	3%
Non-science	14	21%
Did not respond	3	4%
By type of IBL course	**68**	**100%**
Advanced	35	51%
First-year	19	28%
Pre-service teacher	14	21%
By teaching interest	**68**	**100%**
No interest	41	60%
Secondary	13	19%
Elementary	11	16%
May go into teaching	2	3%
Did not respond	1	1%

Over half of all students interviewed were female, and nearly three-quarters were white. Among non-white students, most were of Asian or Hispanic heritage; only a few students were African-

American. Overall, the race and ethnicity distribution of the student interview sample was typical of the departments participating in the study.

A range of majors was represented in the student sample. In all, just over half of undergraduate participants were mathematics majors, and about one fifth were in a science field. Another fifth of the student sample was in a non-science major. A few students were pursuing engineering.

Just over half of the student sample was enrolled in an advanced mathematics class for upper-class students, while 28% were in a course for first-year students, and 21% were taking an IBL course designated for pre-service K-12 teachers. Separately from their course enrollment, students separately reported their interest in teaching. In general, students pursuing elementary teaching were those taking designated pre-service courses, but, as the data show, other students pursuing high school teaching were enrolled in advanced mathematics courses.

In all, a total of 41 interviews were conducted with 68 students. From the demographic data, we may infer a strong interest in mathematics among many students in the sample. As well, a good number were thinking of teaching mathematics as a career.

A7.3.2 Undergraduate interview protocol

Interviews with undergraduate students probed their experiences in IBL mathematics courses, as well as their attitudes and opinions about learning by the IBL method. To better understand student outcomes arising from IBL methods, particular attention was given to exploring students' reports of how their gains had been made. Students were asked:

- About their background and academic goals (major, year in school, plans following college (i.e., career, graduate school), what math courses they had/were taking and why
- To describe how the current/most recent IBL course was taught; how it compared to non-IBL mathematics courses they had taken
- How the IBL method affected their learning, and whether they felt they learned well this way (why or why not)
- Whether they felt they were covering the material they needed for future mathematics or other coursework
- The best and worst things about how the course was taught
- What did they gain, or learn, from the class—or not—and what contributed to or detracted from their learning
- Whether they would take more mathematics courses and if they would choose an IBL over a non-IBL course
- Whether the IBL course had changed ideas about career or graduate school plans
- To offer advice to other students, their instructors, the department or institution.

All interview protocols were submitted for review and approved by the University of Colorado's Institutional Review Board to ensure that this study met high ethical, professional and legal standards for research involving human subjects. Interviewees read and signed an informed consent agreement that described the study and their rights as research participants to anonymity, confidentiality and other protections of the information they provided. They could decline to answer any questions, stop the interview if desired, or decline to be recorded.

A7.4 Methods of qualitative analysis

For both the instructor and undergraduate interview data, we followed a method of formal content analysis. Our methods of data collection and analysis are ethnographic, rooted in theoretical work and methodological traditions from sociology, anthropology, and social psychology (Berger & Luckman, 1967; Blumer, 1969; Garfinkel, 1967; Mead, 1934; Schutz & Luckman, 1974). Qualitative research, such as these interview studies, is particularly useful where existing knowledge is limited, because these methods can uncover and explore issues that shape informants' thinking and actions. Qualitative computer software allows for the multiple, overlapping, and nested coding of a large volume of text data to a high degree of complexity, thus enabling ethnographers to disentangle patterns in large data sets and to report findings using descriptive statistics. Although conditions for statistical significance are rarely met, the results from analysis of text data gathered by careful sampling and consistent coding can be powerful.

Digitally recorded interviews and focus groups are transcribed verbatim into a word-processing program and submitted to *NVivo 8.0*, a computer software program used for qualitative data analysis (QSR International, 2009). The analyst reads through all of the documents—the text data—searching for information relevant to the research questions. Text segments referencing distinct ideas are tagged by code names. Codes are not preconceived, but empirical: each new code marks a discrete idea not previously raised. All of the code names that are developed are collected in a codebook. When the analyst reads a text passage that relates an idea previously encountered, the same code name is reused to mark the relevant passage. Thus codes and their associated text passages are linked, amassing a data set of code names and their frequency of use across the data set. Once all of the text data is coded in this manner, codes similar in nature are grouped together to define analytical themes. For instance, the themes we identified for student learning gains from participating in an IBL mathematics course sorted into five categories: "cognitive or intellectual gains," "understanding the nature of mathematics," "changes in learning," "affective gains," and "communication gains."

The clustered themes or categories describe the nature and range of issues in participants' collective report, and the frequencies with which the themes appear characterize the relative weighting of these issues. That is, frequencies show us what participants have commented on "the most," "some," and "the least." The number of observations is generally much larger than the number of speakers, and thus is a measure of the depth of commentary on broad topics. The number of speakers raising an issue, however, is often a better measure of the distribution of views on a topic. We use both types of counts in reporting results of the qualitative analyses of the student and instructor interview data.

In order to discover whether or not there were meaningful variations in what was reported by different student groups, we also analyzed the student interview data by gender and by course type (first-year, advanced, or for pre-service teachers). These analyses allowed us to explore important questions, such as whether or not women offered more observations than did men concerning their learning gains, or whether students who had taken mathematics courses designed for pre-service teachers noted unique or particular learning gains compared with students who had taken an advanced or first-year mathematics course. For instructors, subgroup analyses compare faculty with graduate students, and experienced with new IBL instructors.

When the subgroups are different in size, it is easiest to compare code frequencies on a per-capita basis, considering the average number of observations per person interviewed, rather than simply the total number of observations. This takes into account the fact that a larger number of interviewees will generally offer a larger number of observations on any topic. However, because interviews are conducted in a conversational manner that allows for topics to arise in a natural order or new topics to be introduced spontaneously, these frequencies are not appropriate for statistical analysis.

A7.4.1 Content of the instructor interview codebook

The instructor interviews codebook included a total of 2414 coded passages, coded into five main themes with 8-14 sub-themes under each. About 40% of instructor comments addressed processes of teaching and another 17% addressed processes of learning. Instructors' observations of student gains (16%), impacts on instructors' own personal and professional lives (16%), and comments on their departmental, disciplinary, and institutional context (10%) comprised the remainder of the codebook.

A7.4.2 Content of the undergraduate interview codebook

The undergraduate interviews codebook included a total of 3390 coded passages, coded into 13 main themes: four main themes included 1-8 sub-themes, which were sometimes further subdivided. Thirty-seven percent of students' observations discussed how gains from an IBL mathematics course had been made; that is, processes that supported their learning. A further 24% of student observations noted specific learning gains from IBL, while smaller bodies addressed gains not made (3%) or "mixed" gains made in partial or qualified degree (1%).

Nearly 10% of student comments were comparisons between IBL and non-IBL classrooms. Another 10% of students' comments were in the form of advice, either to other students or to instructors teaching future IBL mathematics classes. Smaller numbers of observations were offered concerning: the choice to take (or not) another IBL mathematics class (5%), career and graduate school plans (3%), the longer-term impacts of taking an IBL mathematics class (3%), and whether or not women or students of color benefitted more from IBL methods (2%). The remaining comments were miscellaneous in nature.

A7.5 Limitations of the interview studies

While the interview samples are generally large for qualitative work, and appear to broadly represent the student and instructor populations at the IBL Centers, there are some limitations to the interview findings. This particular sub-study did not include a comparison group of students who had taken a non-IBL mathematics class. Data from a comparison group could help to highlight areas in which IBL students' gains were either particularly strong, or unremarkable.

As in any interview study, participation was voluntary, and our samples of both students and instructors may over-represent either those who were especially enthusiastic about the method, or those who had an axe to grind. The sample size was small at one institution where we secured only a handful of student interviews. Because each campus had a distinct style of IBL, descriptions of the classroom activities and characteristics experienced by these students may be underrepresented in the codebook. Though every attempt was made to include students of color in the interview sample, the low fraction of non-white students enrolled in these courses precluded any data analyses that might provide insight into the benefits (or lack of them) of IBL methods for students from groups traditionally underrepresented in mathematics. Likewise, we do not have large enough sub-samples for any single course to analyze patterns for specific courses or campuses.

The instructor sample is robust in size and a good match to the apparent demographics of the instructor population at these campuses. We note that the instructor sample is composed of people who chose or were invited to teach IBL courses. While our observations and interviews do not indicate that they are unusually talented as teachers, they may well be more interested in student-centered teaching methods (and in students themselves) than the average mathematics instructor. We make no claims about the applicability of our findings to all mathematics instructors, but (as discussed in Chapter 8) we do believe that their observations indicate experiences and issues that instructors in other settings are likely to encounter .

A7.6 References Cited

Berger, P. L., & Luckman, T. (1967). *The social construction of reality: A treatise in the sociology of knowledge*. London: Penguin Press.

Blumer, H. (1969). *Symbolic interactionism: Perspective and method*. Englewood Cliffs, NJ: Prentice Hall.

Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.

Mead, G. H. (1934). *Mind, self, and society: From the standpoint of a social behaviorist*. Chicago: University of Chicago Press.

QSR International (2009). *NVivo* qualitative data analysis software; QSR International Pty., Ltd. Version 8.0. See: http://www.qsrinternational.com/#tab_you

Schutz, A., & Luckman, T. (1974). *The structures of the life world*. London: Heinemann.

Exhibit E7.1: Interview Questions for IBL Math instructors individual or focus groups (pairs)

Review consent form, confidentiality and anonymity. Clarify audio recording procedures. **Tell them to keep other names confidential, per HRC:**

> Please do not disclose the name of other students, faculty or staff members when being interviewed. When participating in a focus group please keep all information that is shared in the session confidential.

Turn on mic!

A) <u>Background and teaching experience</u>

Tell me a bit about yourself – what you do here how you are connected to the IBL project here.

- Teaching experience (years, where and what)
- Career status
- For grads/postdocs: future career interests, timing (where now in education/professional process, when will they be job hunting)

B) <u>Teaching style and strategies</u>

Tell me about your course: who takes it, what are your goals for these students
What is your overall teaching approach or philosophy?
What particular strategies are you using to achieve this? Why those (and not others)?
How is this working? (What are you happy or not happy with, in how the course is going?)
Will you teach this way again? The same way or differently? (What will you change?)

C) <u>Role in IBL project (if instructor in IBL course)</u>

How did you get involved in the IBL project at your campus? Did someone ask you or did you volunteer? (why do you think they picked you?)
How do you define IBL, for yourself?

- What goes on in this course that you think is different from the way most math courses are taught?
- How does IBL fit in with your own beliefs about teaching and learning?

D) <u>Observations about student gains</u>

What do you think students get out of your courses? (Open-ended, then probe for specific categories:

- understanding of mathematics content
- understanding of the nature of mathematics, how mathematicians work
- communication skills – writing and speaking about math
- critical thinking skills or habits of mind
- attitude changes: confidence, enjoyment, interest in mathematics
- changes in your problem-solving style, ways of learning math (independence, persistence, reflection)
- Any other gains we have not mentioned yet?

Do you see any impacts on their course-taking afterwards? Career or grad school plans?

- Probe particular benefits for the grad-school bound
- For life in general

What do they not get? (OR what do they miss out on, that they get in other places?)
For what type of student does your teaching approach work best? Why do you think so?
Do you notice any patterns in which students respond well or not – e.g. with respect to gender, ethnicity, background?
For non-IBL instructors: Are you familiar with the IBL project on your campus? (to set up next questions for them – but ask All
What kind of students learn best from IBL?
What makes IBL difficult for students?
What kinds of students find this particularly difficult or daunting?
Do you think every student have an IBL experience as part of a college math degree? Why or why not?

E) Personal and professional gains/costs

What do you, as a teacher, gain from teaching this way?
What are the costs to you of teaching this way?
What do you like/not like about this approach?
Have your beliefs about teaching and learning in any way? (why or why not?)
(Ask about linkage to IBL course for IBL instructors)

F) Additional questions for **grad students** helping or TAing IBL courses

What do you find yourself doing – what is your role in supporting students in the course?
What do you think works well about that?
What is difficult about that?
What do you get, personally, out of doing this work?
How do you think your work with this course has influenced your own ideas about teaching and learning, so far?
Do you think you would choose to teach this way if you were in charge of a course yourself?
(Given career goals) How do you think working witih this course will prepare you for that career?
Probe career expectations re teaching, type of teaching, in particular.
Ask if we can keep in touch to follow up on their career progress; get permanent contact info if possible.

G) Additional questions for instructors also in a leadership role in the project

Tell me about how this department, as a group, and you as an individual, got involved in IBL.
How have you recruited or persuaded others to participate?
Has that been hard or easy?
What kind of people seem most interested in teaching this way?
What status issues do you observe around participating in IBL?
What are the barriers or difficulties that people raise who don't want to teach this way, or who are reluctant? (why do some people not want to do it?)
What in your view have been the successes ?
Why, or what are your criteria?
What have been the challenges?
Why? What makes it hard?

Any outright failures?

Tell me a bit about your interaction (or the project's) with the other campuses.

How often do you meet, how well do you know the other people or projects?

What do you get out of this interaction?

What is your perception of how IBL is done at this campus as compared with how it is done on other campuses?

H) <u>Advice</u>

Any advice you'd want to give the faculty or department about this course or this approach?

Any advice you'd want to give another instructor (TA) for this course?

Anything we should have asked that we didn't? Anything you want to add or emphasize?

Thanks, follow-up, goodbye.

Exhibit E7.2. Interview Questions for IBL Math students - individual or focus groups

Revised January 28, 2009

Review consent form, confidentiality and anonymity. Clarify audio recording procedures. **Tell them to keep other names confidential, per HRC:**

> Please do not disclose the name of other students, faculty or staff members when being interviewed. We understand that in some cases the name of an instructor may be understood if you are taking a particular class that is being discussed. When participating in a focus group please keep all information that is shared in the session confidential.

Turn on mic!

<u>Background and academic goals</u>

Tell me about yourself —what's your major, what year are you?

- What do you plan to do after college? (career, grad school plans)
- What math course are you taking (did you take recently) and why are you taking this course?

<u>Learning experiences</u>

Tell me about how this course is taught, compared to other math courses you've taken here.

- Did you know it would be like that, when you signed up?
- How does it compare to what you expected?

How do you think this way of teaching math affects your learning?

- Do you feel you learn well this way? Why or why not?
- Do you enjoy it ? Why or why not?
- Why do you think the instructor chooses to teach it this way?

How do you feel right now about how it's going ? (Ask in retrospective form, for past students—How did it go? Did you learn the material, get a good grade?)

- Do you think you'll learn the material you need to learn to go on in math or in other courses in your major?
- Do you think you'll get a good grade?
- Do those things worry you ?

What's the best thing about how this course is taught?

What's the worst thing about how this course is taught?

<u>Learning gains</u>

Overall, what do you think you are getting (got) from this course? (Open-ended first, then prompt for "some of the things faculty think students gain from this kind of course") :

- understanding of mathematics content
- understanding of the nature of mathematics, how mathematicians work
- communication skills – writing and speaking about math
- critical thinking skills or habits of mind
- attitude changes: confidence, enjoyment, interest in mathematics
- changes in your problem-solving style, ways of learning math (independence, persistence, reflection)

Any other gains we have not mentioned yet?

Are you likely to take more math courses ? (would you have done so anyway?)

Any changes in which courses you think you'll take next?

Any changes in your ideas about your career or grad school plans, after taking this course?

Advice

Based on your experience, what kind of student do you think is most likely to succeed with this type of course?

What kind of student do you think will find it difficult or not to their liking?

What would you tell a friend who was thinking of taking this course?

Any advice that you'd like to give the faculty and department about this course or this way of teaching?

Anything else that we should have asked you about?

Anything else you want to add or emphasize?

Thanks, followup, goodbye.