# 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 prior math courses—count of math courses taken prior to the target course.
- Number of subsequent 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 required 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 elective 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 IBL 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 prior 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.
- Next 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 required 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 elective 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 IBL 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/>.