Measuring the Impact of an Individual Course on Students' Success

VILMA MESA, OZAN JAQUETTE, and CYNTHIA J. FINELLI
University of Michigan

BACKGROUND
At the University of Michigan, qualified first-year students who place out of the first-semester calculus course may enroll in either the regular second-semester calculus course or Applied Honors Calculus II. Students who enroll in Applied Honors Calculus II show higher academic performance than students enrolling in the Regular Calculus II.

PURPOSE (HYPOTHESIS)
The study addressed the question: does enrollment in Applied Honors Calculus II have a positive causal impact on subsequent academic performance for engineering students at the University of Michigan?

DESIGN/METHOD
We acquired seven years of institutional data for engineering students who entered the University of Michigan from 1996 through 2003 and who qualified to enroll in Applied Honors Calculus II. Using regression analyses, we tested a causal model of impact of Applied Honors Calculus II on four measures of subsequent academic performance: grade in Physics II and average grade in all subsequent physics, mathematics, and engineering courses.

RESULTS
After controlling for students' personal characteristics and prior academic achievement, the impact of Applied Honors Calculus II on students' academic performance was not statistically significant. In particular Advanced Placement scores accounted for the higher performance observed in Applied Honors Calculus II students.

CONCLUSIONS
We recommend including Advanced Placement scores in models that predict academic performance. Future research should also include measures of socioeconomic status (SES) and explore interactions between SES and academic background. Finally, in evaluations of specific curricula, the treatment effect—measured as treatment group mean minus control group mean, after controlling for covariates—is unlikely to be large if the control group receives high quality instruction.

KEYWORDS
academic performance, advanced placement, course impact

I. INTRODUCTION

What is the impact of the individual courses that comprise a college curriculum on students' subsequent academic performance? The engineering education literature has partially addressed the question with studies of the relationship between enrollment in engineering courses in which teaching has been modified in some way and performance variables (such as course grades, grade point average (GPA), persistence, and retention) while accounting for non-cognitive variables (such as learning styles, motivation, or study skills). It is common for studies of this type to present both longitudinal and cross-sectional student data from three sources: (1) course grades and students' GPA (Budny, LeBold, and Bjedov, 1998; Felder, Felder, and Dietz, 1998; French, Immekus, and Oakes, 2003; French, Immekus, and Oakes, 2005; Hoit and Ohland, 1998) (2) institutional data (e.g., students' prior academic characteristics, retention, graduation status), and (3) surveys that collect information on learning styles (e.g., Kolb, 1986) or study skills (e.g., the Learning and Study Strategies Inventory, LASSI). The analyses typically presented in these studies include reporting frequencies and correlations, as well as linear or logistic regressions (Bernold, Spurlin, and Anson, 2007; Budny, LeBold, and Bjedov, 1998; Felder, Felder, and Dietz, 1998). Studies that use measures of learning (usually with diagnostic tools such as concept inventories) that are not tied to course grades are less common in engineering (e.g., Martin, Mitchell, and Newell, 2004; Wage et al., 2005), though they are more widespread in physics and mathematics education (Epstein, 2005; Hestenes, Wells, and Swackhamer, 1992).

The current literature suggests that changes in instruction that engage students with the course material (thus matching a wider range of learning styles) are positively related with both higher students' grades (course and GPA) and increased retention (usually after three years in the program) (Burtner, 2005; Hoit and Ohland, 1998). However, the robustness of these findings is not obvious for several reasons. Most studies do not randomly assign students to different programs or do not include controls to isolate effects of the intervention (Bernold et al., 2000; Hoit and Ohland, 1998), the samples are small relative to the length of the surveys administered (Burtner, 2005), or key information—such as the size of the sample or the specific intervention that is used—is missing (Budny, LeBold, and Bjedov, 1998; French, Immekus, and Oakes, 2005).

Felder and colleagues (Felder, 1995; Felder, Felder, and Dietz, 1998; Felder et al., 1995, 1993, 1994) have reported the most prominent series of correlational studies in the engineering education literature. The studies show that when compared to students who take a series of traditionally taught courses (control), students who take the same courses using a teaching approach that includes active and cooperative learning have more positive perceptions about their level of preparation and about the quality of their education, better retention, and higher graduation rates than the control group (Felder, Felder, and Dietz, 1998). These studies have concentrated on establishing the significance of the association between the changes in instruction and student performance, and they differ from studies attempting to measure a causal effect. Correlational designs describe the strength of the relationship between an outcome variable and a treatment variable, whereas causal designs describe the strength of the relationship between an outcome variable and a treatment variable after controlling for all other variables (both observable and unobservable) that are potentially correlated with both the outcome variable and the treatment variable.

In contrast, our study attempts to establish a causal relationship by comparing the subsequent academic performance of students who enrolled in Applied Honors Calculus II, with those who were eligible for the course but did not enroll. Our study is different from previous work in two ways. First, the data we analyze comprise seven years of data and include information about all courses students took while enrolled at the university. Most other reported studies employ shorter time spans, ranging from one term to up to three years of data. Second we propose a theoretical model to explain the way in which taking the course might affect students' achievement, and we test the model to determine causality, while previous work has typically not used a model and/or has established the relationships without exploring causality.

A. Background

The typical engineering curriculum at the University of Michigan (U-M) includes a series of four regular calculus classes (Calculus I, Calculus II, Calculus III, and Introduction to Differential Equations), taken in order beginning with the student's first semester. Many engineering students who enter U-M place out of Calculus I, and they have two options for their first-semester math course. In general, they enroll in Math 116: Regular Calculus II, but if they earned a 4 or 5 on any Advanced Placement (AP) calculus exam in high school, they may enroll in Math 156: Applied Honors Calculus II (AHCII).

Math 116 is part of the regular calculus sequence (115/116/215/216) at U-M which was revised in the early 1990s during the national calculus reform movement (Ganter, 1999). The reform-oriented model employs many small (20–25 students), team-based sections that involve extensive student interaction rather than the traditional model with a large lecture and several small recitation sections. The emphasis of the course is on solving applied problems from a range of disciplines using geometric, numerical, symbolic, and verbal representations for explaining the solutions. A central coordinator ensures that all sections of the course maintain the same pace, makes lesson plans available to instructors, maintains a bank of problems for quizzes, and organizes a week-long training program for new instructors that emphasizes active and cooperative learning. The subsequent course (Math 215: Regular Calculus III) is taught in a large lecture (~100 students per section), with multiple accompanying small recitation and lab sections (~25 students each).

On the other hand, Math 156 is part of the applied honors calculus sequence (156/255/256). It was created in 1994 in response to a request from the College of Engineering for an alternative to Math 116 for students with strong math ability and with interest in science and engineering fields. Like Math 116, the course employs small sections, it is overseen by a course coordinator to ensure uniformity, and it is focused on applications. However, unlike Math 116, Math 156 sections are lecture based, the material includes science and engineering applications only, and the presentation is theoretical (instructors provide proofs to theorems, and students are expected to understand the proofs intuitively but not reproduce them). After completing Math 156, a student may enroll in either Math 215: Regular Calculus III or Math 255: Applied Honors Calculus III. Table 1 summarizes the main characteristics of Math 116 and Math 156.

Opinion surveys conducted in 2005 and 2006 with two cohorts of freshmen who scored a 4 or 5 in the AP calculus exam (and thus were eligible to enroll in Math 156 in their first semester), and who enrolled either in Math 116 or in Math 156 highlighted differences the students perceived between the two courses offered (Krasny and Mesa, 2005, 2006). Students who enrolled in Math 116 appreciated the small class size and the team-based problem solving, whereas students in Math 156 appreciated the complexity of the material and praised the high quality of instruction. Because the survey data were collected anonymously, it was not possible to conduct a more fine-tuned analysis that would allow the researchers to establish the real impact of the course on students' performance.

However, for the present study, the authors were able to obtain course data for all College of Engineering students who entered the university since 1997 (one year after the inception of Math 156), and we conjectured that such longitudinal data could give us more information regarding the impact of the course on students' subsequent performance in the university. The specific question that we investigated was: does enrollment in Math 156: Applied Honors Calculus II have a positive causal impact on subsequent academic performance for College of Engineering students at U-M?

B. The Model

The theoretical model that we used to guide our inquiry is presented in Figure 1. The model includes three types of variables: (1) independent variables, (2) the treatment variable, and (3) dependent variables. *Independent variables* include students' personal characteristics (gender and ethnicity), pre-college academic characteristics (SAT/ACT and AP scores), and college characteristics (cohort or year enrolled as first-term students). The *treatment variable* corresponds to enrollment in Math 156: Applied Honors Calculus II—enrollment in Math 116: Regular Calculus II is considered the control. *Dependent variables* include measures of students' subsequent academic performance in physics, mathematics, and engineering courses taken *after* the treatment.

We propose three ways in which the variables can affect the outcomes. First, because some student variables (e.g., AP score) determine the type of course students will be eligible to take, there is clearly an interaction between the independent and treatment variables (Arrow A). Our sample consists only of students who were eligible to take Math 156, that is, students who scored a 4 or 5 on any AP calculus exam. Second, we assume that student variables have an impact on the dependent variables (Arrow B). The purpose of this research is to determine the impact of the treatment on the dependent variables, thus exploring the third effect (Arrow C). We hypothesize that the treatment (taking Math 156) would have some effect on students' subsequent academic performance in three areas (physics, mathematics, and engineering courses), and we designed our analysis to measure that effect.

	Math 116: Regular Calculus II	Math 156: Applied Honors Calculus II
Eligibility	Placement score of 4 in university designed math test	Placement score of 4 in university designed math test and score of 4 or 5 on any AP calculus test
Textbook[a]	*Calculus Single Variable* (Hughes-Hallett et al., 2005)	*Calculus Single Variable* (Stewart, 2003)
Number of sections per term	25 to 30	4 to 6
Class size	25 to 30	15 to 30
Audience	All majors	STEM majors
Class length	80 minutes, three times a week	50 minutes, four times a week
Lead instructor	Graduate students and post-docs	Faculty and post-docs
Mode of instruction	Team-based: In-class group-work, with short instructor-led lectures and student presentations on the board; instructor creates quizzes	Lecture-based: Instructor leads lecture and discussion, students pose questions to be answered by instructor; instructor creates quizzes
Type of applications	Science, business, and social sciences	Science and engineering
Content emphasis	Challenging problems that address fundamental ideas (e.g., change); connections between geometric, numerical, symbolic, and verbal representations; applications to science, business, and social sciences	Theorems and proofs emphasizing key ideas and de-emphasizing technical details; includes some extra topics (e.g., trigonometric substitutions, Bessel & hyperbolic functions)
Homework	Individual and group-based	Individual, but students can work in groups
Exams	Same for all sections, graded using the same rubric; groups of instructors grade all questions except those of their own students in a session immediately following the exam	Same for all sections, each instructor grades the exams of his or her own sections
Follow-on course	Math 215 (regular Calculus III), taught in large lecture hall (100 students) with small recitation sections	Either Math 215 or Math 255 (Applied Honors Calculus III) taught in small sections

Note: [a]Different editions of these textbooks have been used since 1996.

Table 1. Main characteristics of the Regular and Applied Honors Calculus II courses for first-semester students.

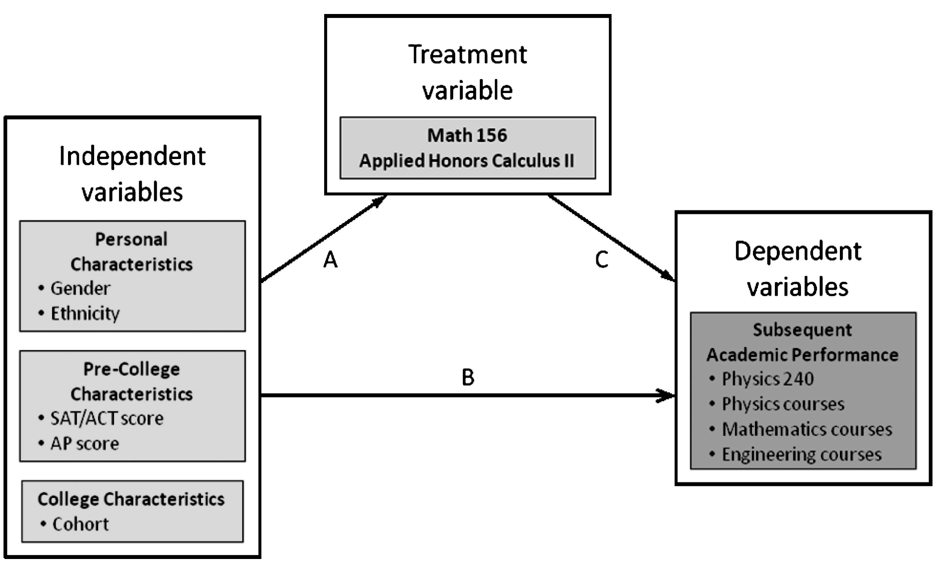

Figure 1. Model of hypothesized influences of taking a special mathematics course and other variables on engineering students' subsequent academic performance.

II. Method

To answer our question (Does enrollment in Math 156: Applied Honors Calculus II have a positive causal impact on subsequent academic performance for College of Engineering students at U-M?) we employed longitudinal student-level data. Specifically, our dataset consisted of all U-M College of Engineering students who took a Calculus II course (Math 116 or Math 156) from the Fall 1997 until Fall 2003 and who scored 4 or 5 on any AP calculus test (i.e., who were eligible to enroll in Math 156). We collected gender, ethnicity, SAT score, and AP score, as well as course-level performance data (course grades) for each student from the term of first enrollment until the Fall 2006 term. By choosing this cut-off, all cohorts have comparable numbers of credits for computing the outcome variables, except for the 2003 Cohort, which had about 13 fewer credits (three to four courses) than the other six cohorts. This difference did not affect the final results of the analysis.

A. Variables

Our model includes 16 *independent* variables describing students' gender, ethnicity, SAT score, AP score, and cohort; our *treatment* variable is student enrollment in Math 156: Applied Honors Calculus II and we study four *dependent* variables. The fundamental principle employed in constructing our dependent variables (students' subsequent academic performance) is that the treatment must occur before the outcome measured by the variable, so the dependent variables in our study include data from the second and subsequent terms of each student's academic career only (by design, all students in the study enrolled in the math class during their first semester). The variables also include only courses students took at U-M. The first dependent variable, students' grades in Physics 240 (PHYS240GRADE), represents the closest approximation to a post-test because at U-M, Physics 240 is the only class that has Calculus II as a prerequisite and that must be taken by all College of Engineering students regardless of engineering major. The other three dependent variables include GPA (calculated on a scale of 0 to 4.0) in physics courses (PHYSGPA, excluding Physics 240), GPA in math courses (MATHGPA) taken after the treatment, and GPA in engineering courses (ENGRGPA) taken after the treatment. Because Math 156 focuses on science and engineering applications of calculus, these courses were the most likely to capture the effect of a single calculus course, and therefore grades in these courses are the best variable choices. In our design we measured the effect of the treatment on students' academic performance after controlling for the effect of each independent variable, so these 16 variables are included as covariates. Our variable set is described in Table 2.

B. Sample

After acquiring approval for human subjects research from the U-M Institutional Review Board for Behavioral Sciences, we received data from the College of Engineering for all students who enrolled in an engineering program since Fall 1997 through Fall 2006. Our sample consisted of first-year College of Engineering students who took either Math 116 or Math 156, with an AP calculus score of 4 or 5, without missing SAT scores, and who first enrolled at U-M between Fall 1997 and Fall 2003. Of all the 5,756 students in the original sample, only 1,761 qualified in this sample. Of these 1,761 students, 1, 345 (77 percent) took Math 116 and 407 (23 percent) took Math 156.

Table 3 shows that we had missing dependent variables in our analysis sample. The missing observations for PHYS240GRADE represent students who never took the course at U-M. The missing observations for PHYSGPA, ENGRGPA, and MATHGPA represent students who either did not take a course in physics, engineering, or math at U-M in a semester *after* taking Calculus II or took those courses *concurrently* with the treatment. Students with missing dependent variables might have left the College of Engineering by transferring to other programs (e.g., business, science, and education), they might have discontinued the program, they might not have taken the course at U-M yet, or they might have graduated (without taking the class at U-M). Because percentages of missing data are relatively small, we can assume there is no significant bias in the samples that will be used for our analysis.

C. Data Analysis

We chose regression modeling for testing the causal relationship between the independent variables, the treatment variable, and the dependent variables; and we planned three steps for our analysis. First, we calculated descriptive statistics in order to establish potential relationships and trends among the variables. Second, we conducted simple multivariate regression analyses accounting for all of the variables we were able to measure in our dataset (i.e., the "observable" variables). Third, in case of finding a significant treatment effect on our dependent variables in the second step, we planned to analyze the "unobservable variables" by using more sophisticated modeling techniques—for example, propensity score matching (Morgan and Harding, 2006) or bivariate normal selection models (Heckman, 1979)—to account for potential non-random selection into the treatment (Math 156) group that was not accounted for by the observable variables.

III. Results

We organize the presentation of results according to the steps of our analysis: Descriptive Statistics, Multivariate Regression Analysis, and Analysis of Unobservables.

A. Step 1: Descriptive Statistics

In Table 4 we present a summary of the characteristics of the sample by gender, ethnicity, and cohort. Twenty-five percent of all students in the analysis sample are female, 26 percent of students in the Math 116: Regular Calculus II group are female and 22 percent of students in the Math 156: Applied Honors Calculus II group are female. The proportion of White students is the same in each group; in general, no ethnicity group is over-represented or under-represented in the groups relative to their representation in the total analysis sample. Relative to each course, the cohorts got larger over time, but the percentage of each cohort in the Math 156 group, although larger prior to 2000, declined over time relative to the Math 116 cohorts.

The descriptive statistics for SAT and AP scores (see Table 5) suggest that the groups differ with regard to these two attributes. The SAT score for the Math 116 group was statistically significantly lower than the SAT score of the Math 156 group

Name	Description	Coding
	Independent Variables	
FEMALE	Whether student was female.	0: Male 1: Female
Ethnicity (coded as a series of six dummy variables – ASIAN, BLACK, HISPANIC, NATAM/ISLANDER (Native American/Pacific Islander), WHITE, MISSING)	Whether student reports being of the given ethnicity.	1: student reports being of the given ethnicity
SAT	Combined math and verbal SAT score. ACT scores were converted to the SAT scale.	Continuous
AP	Score obtained in any AP calculus test.	4 or 5
Cohort (coded as a series of 7 dummy variables – 1997[a],1998, 1999, 2000, 2001, 2002, 2003)	Whether student enrolled as first-year students in the Fall semester of that year.	1: student enrolled in Fall of the given year
	Treatment Variable	
TREATMENT	Whether student enrolled in Math 156, students in the control group enrolled in Math 116.	0: Math 116 (Control) 1: Math 156 (Treatment)
	Dependent Variables	
PHYS240GRADE	Grade obtained in physics course, Physics 240 taken after Calculus II.	Grade and half grade letters treated as continuous from 0.0 to 4.0[b]
PHYSGPA	GPA from all physics courses taken after Calculus II excluding Physics 240.	Continuous 0-4
MATHGPA	GPA in all math courses taken after Calculus II.	Continuous 0-4
ENGRGPA	GPA in all engineering courses taken after Calculus II.	Continuous 0-4

Note: [a]1997: 1997–1998 academic year. [b]U-M uses grades qualified with + or – that increase or decrease a unit letter grade by 0.3. An A corresponds to 4.0, A- to 3.7, B+ to 3.3, etc. (although the A+ grade is an option, for the purposes of GPA computation, this is considered equivalent to an A and it corresponds to a 4.0). We performed analyses treating this variable as ordinal and as a scale and the results were practically identical. For this reason we decided to treat this variable as continuous.

Table 2. Name and description of variables used in the study.

	Valid Data	Transferred	Discontinued	Has not taken (current student)	Has not taken (graduated)
PHYS240GRADE	1,205 (68%)	91 (5%)	57 (3%)	86 (5%)	322 (18%)
PHYSGPA	1,441 (82%)	52 (3%)	35 (2%)	39 (2%)	194 (11%)
MATHGPA	1,704 (97%)	42 (2%)	13 (1%)	1 (0%)	1 (0%)
ENGRGPA	1,722 (98%)	23 (1%)	12 (1%)	2 (0%)	2 (0%)

Table 3. Number and percent of students in the analysis sample ($N = 1,761$) with valid dependent variables.

($t = 4.62$, $p < 0.001$); and similarly, the mean AP calculus test score for the Math 116 group was statistically significantly lower than the AP calculus test score for the Math 156 group ($t = 6.11$, $p < 0.001$). These differences point to the need for using these variables as covariates in the analysis.

Table 6 shows the sample size, mean value, and standard deviation for each of the four dependent variables for our analysis sample and by sub-group. Two trends are apparent. First, mean scores are lowest for Physics 240 and highest for engineering GPA. Second, for each variable, Math 116 means are lower than Math 156 means, and the differences are statistically significant (independent t-tests for the differences of the means produced values for t that ranged from 1.69 to 3.05.).

	Total Sample $N = 1,761$	Control Group Math 116 $N = 1,354$	Treatment Group Math 156 $N = 407$
FEMALE	440 (25%)	352 (26%)	88 (22%)
ASIAN	302 (17%)	225 (17%)	77 (19%)
BLACK	64 (4%)	53 (4%)	11 (3%)
HISPANIC	55 (3%)	44 (3%)	11 (3%)
NATAM/ISLANDER	8 (0%)	6 (0%)	2 (0%)
WHITE	1239 (70%)	952 (70%)	287 (70%)
MISSING (ethnicity)	93 (5%)	73 (5%)	20 (5%)
COHORT1997	75 (4%)	37 (3%)	38 (9%)
COHORT1998	237 (13%)	122 (9%)	115 (28%)
COHORT1999	241 (14%)	176 (13%)	65 (16%)
COHORT2000	295 (17%)	239 (18%)	56 (14%)
COHORT2001	272 (15%)	239 (18%)	33 (8%)
COHORT2002	315 (18%)	268 (20%)	47 (12%)
COHORT2003	326 (19%)	273 (20%)	53 (13%)

Table 4. Frequency and proportion of students in the sample by gender, ethnicity, and cohort for the entire sample and by sub-group.

	Total Sample $N = 1,761$		Control Group Math 116 $N = 1,354$		Treatment Group Math 156 $N = 407$	
	Mean	(SD)	Mean	(SD)	Mean	(SD)
SAT	1339	105	1332	101	1361	114
AP	4.47	0.50	4.43	0.50	4.60	0.49

Table 5. Mean and standard deviation for SAT and AP Scores for the entire sample and by subgroup.

Comparisons within the Math 116 group and within the Math 156 group reveal that mean scores for the dependent variable vary substantially by AP calculus score, showing higher scores for students with the higher AP score (Table 7). For example, within the Math 116 group, students with an AP score of 4 have a mean PHYS240GRADE of 2.69 and students with an AP score of 5 have a mean PHYS240GRADE of 3.04 (a statistically significant difference of 0.35, $t(886) = 6.23$, $p < 0.001$). Within the Math 156 group, students with an AP score of 4 have a mean PHYS240GRADE of 2.78 and students with an AP score of 5 have a mean PHYS240GRADE of 3.09 (a statistically significant difference of 0.31, $t(234) = 2.98$, $p < 0.01$). Thus an important question is whether the differences between the Math 156 group and the Math 116 group are maintained once the effects of the covariates are taken into account. This is the purpose of the regression analysis.

Table 8 shows the mean number of post-treatment credits attempted at U-M in physics, engineering, and math, respectively, which includes all courses for which students earned a grade. Because three of the dependent variables, PHYSGPA, ENGRGPA, and MATHGPA, are only calculated for courses taken in semesters after the student has taken Calculus II, we include this table to be clear about how many credits are being counted towards these average grades.

B. Step 2: Multivariate Regression Results

Table 9 shows multivariate ordinary least-squared (OLS) regression results for each of our dependent variables PHYS240GRADE, PHYSGPA, ENGRGPA, and MATHGPA. As seen from the table, the effect of the treatment variable on the dependent variables is negligible; that is, relative to taking Math 116: Regular Calculus II course, taking Math 156: Applied Honors Calculus II results in increases in the dependent variables that are *not significant* (β ranges from 0.008 to 0.029, with p ranging from 0.209 to 0.780). Thus when controlling for other aspects such as SAT or AP score, the differences observed in the raw data are not significant.

The table also shows that holding other variables constant, being female compared to being male is associated with a statistically significant reduction of 0.059 in the grade in Physics 240 but is associated with a significant increase of 0.074 in the Math GPA. Independent of the treatment, Black and Hispanic students have significantly lower GPAs than their White peers in physics and engineering courses (the decreases range from 0.049 to 0.146); Black students also have significantly lower GPAs in math than white students (a decrease of 0.103). In practical terms, however, these differences may not be significant.

The SAT and AP scores have a significant conditional correlation with all of the dependent variables, independent of the treatment. A change of one standard deviation in the SAT score (105 points) will result in grade increases that range from 0.057 (MATHGPA) to 0.145 (PHYSGPA). The conditional correlation with AP scores is more dramatic. An increase of one standard deviation in the AP score (0.5 points) increases the grade in Physics 240 and the GPAs in physics, engineering, and mathematics by at least 0.11; such increases have a recognizable impact on students' grades.

Finally, Table 9 shows that, independent of the treatment, students in the 2003 Cohort had statistically significantly lower

	Total Sample N = 1,761		Control Group Math 116 N = 1,354		Treatment Group Math 156 N = 407	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
PHYS240GRADE	1,205	2.88 (0.87)	927	2.85(0.87)	278	2.97(0.86)*
PHYSGPA	1,441	2.93(0.76)	1,116	2.90(0.74)	325	3.05(0.79)**
ENGRGPA	1,704	3.13(0.71)	1,305	3.11(0.70)	399	3.18(0.73)†
MATHGPA	1,722	2.99(0.81)	1,319	2.97(0.82)	403	3.06(0.80)**

Notes: The difference of means between the control group and the treatment group is statistically significant at the following levels: †α = 0.10; *α = 0.05; and **α = 0.01.

Table 6. Mean and standard deviation for dependent variables for entire sample and by subgroup.

	Control Group Math 116 N = 1,354				Treatment Group Math 156 N = 407			
	AP = 4		AP = 5**		AP = 4		AP = 5**	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
PHYS240GRADE	517	2.69(0.86)	410	3.04(0.84)	109	2.78(0.84)	169	3.09(0.86)
PHYSGPA	636	2.76(0.73)	480	3.07(0.72)	129	2.79(0.85)	196	3.22(0.70)
ENGRGPA	743	3.04(0.70)	562	3.20(0.69)	157	2.93(0.73)	242	3.34(0.69)
MATHGPA	750	2.86(0.81)	569	3.12(0.81)	159	2.74(0.83)	244	3.28(0.71)

Note: ** All the differences of means between students with AP of 5 and AP of 4 are statistically significant at α < 0.01.

Table 7. Mean and standard deviation for dependent variables by AP score and by subgroup.

	Total Sample N = 1,761		Control Group Math 116 N = 1,354		Treatment Group Math 156 N = 407	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
Physics	1,441	7.78 (3.59)	1,116	7.60 (3.24)	325	8.37 (4.54)
Engineering	1,704	51.14 (23.22)	1,305	50.92 (22.90)	399	51.87 (24.24)
Mathematics	1,722	10.71(4.38)	1,319	10.45 (4.13)	403	11.56 (5.01)**

Notes: The difference between the number of credits attempted by the control group and the treatment group is statistically significant at the following levels: *α = 0.05 and **α = 0.01. The differences suggest that students in the treatment group took slightly more credits than students in the control group.

Table 8. Mean number and standard deviation of post-treatment credits taken in physics, engineering, and math for the entire sample and by subgroup.

grades in Physics 240 and in physics courses relative to the students in the 1997 Cohort (decreases of 0.147 and 0.104, respectively). Overall, the variables included in the models, explained between 7 percent and 11 percent of the variance observed in the dependent variables.

C. Step 3: Analysis of Unobservables

Although raw descriptive statistics show that students who enroll in Math 156 have higher levels of academic performance than students who enroll in Math 116, the effect disappears in all four measures of academic performance once we use regression modeling to control for the effect of other variables. Thus, we did not investigate whether there were unobservable variables that could explain the effect observed due to the treatment.

IV. Discussion

Before discussing our results, we highlight two limitations of the study. First, the goal of both Math 116: Regular Calculus II and Math 116: Applied Honors Calculus II is to provide students with the calculus they will need for subsequent work in their discipline. Math 116 does this by using applications from a wide range of disciplines and presenting the concepts of calculus from four points of view: geometric, numerical, symbolic, and verbal. Math 156 presents calculus in the context of engineering and science applications from a theoretical point of view that includes understanding some proofs intuitively. In an ideal design to test causality, all students enrolled in either course would respond to a pre- and a post-test measuring their ability to solve real-life engineering and science problems that would employ the competencies

	PHYS240GRADE N = 1,205	PHYSGPA N = 1,441	ENGRGPA N = 1,704	MATHGPA N = 1,722
TREATMENT	0.008	0.033	0.025	0.029
FEMALE	–0.059*	–0.046	0.042	0.074**
ASIAN	–0.028	–0.045	–0.091**	0.005
BLACK	–0.051	–0.117**	–0.146**	–0.103**
HISPANIC	–0.053	–0.069**	–0.049*	–0.029
NATAM/ISLANDER	–0.017	0.002	0.000	–0.013
MISSING (ethnicity)	–0.054	–0.036	–0.026	–0.013
SAT	0.129**	0.145**	0.080**	0.057*
AP	0.170**	0.182**	0.112**	0.174**
COHORT1998	0.027	–0.022	0.038	0.013
COHORT1999	–0.036	–0.045	0.042	0.034
COHORT2000	0.022	0.000	0.057	0.045
COHORT2001	–0.031	–0.048	0.068	0.064
COHORT2002	0.011	–0.010	0.092	0.074
COHORT2003	–0.147**	–0.104*	0.042	0.021
% variance explained by model (R^2)	9%	11%	7%	7%

Note: Total proportion of variance of the dependent variable explained by the model. The coefficient is statistically different from 0 at the following levels: $^*\alpha = 0.05$ and $^{**}\alpha = 0.01$.

Table 9. Standardized regression coefficients for the four dependent variables for students whose AP Score is 4 or 5.

learned in Calculus II with tests that would not be biased in favor or against any particular course. Furthermore, it would be prudent to have multiple post-tests over time because the effect of the treatment would be expected to diminish as time passes (Shadish, Cook, and Campbell, 2002). These designs would provide us with measures of students' competencies both as they take the courses and after they finished taking the courses, allowing us to tease out effects from courses taken and from students' maturation. Unfortunately, these ideal dependent variables were not available. Another ideal dependent variable would measure competency in workplace activities that require the calculus learned in the class. Because of the time span between when students take the class and when they have the opportunity to apply their learning, these types of tests are impractical.

A second limitation regards the assignment of students to conditions. Regression allows us to establish the relationship between a particular independent variable and the dependent variable controlling for the effect of other variables. An ideal situation for modeling effects is that of a random assignment of individuals to conditions. With random assignment it can be assumed that if differences between the control group and the treatment group are observed, the differences result exclusively from the treatment. In addition, it is assumed that when the randomization is properly done, individuals with particular attributes are equally likely to be in either condition, thus reducing the likelihood of having biased samples, and therefore increasing the robustness of the effects found. Our assignment was not randomized because individuals had the option of selecting the type of course in which they wanted to enroll. Thus, our samples are susceptible to having individuals with particular characteristics associated with the dependent variable that are distributed unequally between the two groups. We attempted to control these possible biases in the sample by including as many measures of individual characteristics as possible. In addition, some characteristics associated with the treatment and the outcome could not be measured (i.e., unobservables such as individuals' preference for group work, who may be more likely to enroll in Math 116). Given that we did not find a significant treatment effect after controlling for the set of observable characteristics, we did not attempt to control for unobservable variables.

In light of the results of the analysis, we answer our research question negatively; we did not detect a (statistically significant) positive, causal impact of enrollment in Applied Honors Calculus II (Math 156) on subsequent academic performance of U-M College of Engineering students, here measured with students' grades in different courses. The large grade differences that we observe for students in each group disappear once we account for student characteristics such as SAT and AP score. There are (at least) two possibilities for this—either there is no effect resulting from taking Math 156: Applied Honors Calculus II and it really does not matter which calculus course students take as freshmen, or there is an effect but we have not been able to detect it.

The first possibility is plausible given the sample of students we have considered. Students who enroll in a Calculus II course (whether the Regular course or Applied Honors Calculus II) as first year students may have different characteristics from other students (e.g., they chose to take AP courses offered in their schools), and a particular teaching method experienced in one particular course may have little impact on their subsequent performance. It is possible that these students will be high achievers independently of the opportunities we offer them. It would be important to study this possibility even further by including students' socioeconomic status (SES) in the model, because it is known that this variable has an impact on students' college academic performance (DesJardins, Ahlburgh, and McCall, 2002; Ishitani and DesJardins, 2002). If the impact of AP scores on subsequent performance is significant, then we could say that scores on AP courses are measuring attributes of students who will perform similarly no matter what instructional intervention is given to them, and therefore it might be a good predictor of how students will do in college. If, on the other hand, the impact of AP scores disappears, then SES is a confounding variable; it would indicate that schools which offer AP courses might be attended by wealthier families who can also afford paying the AP fees: for the 2007–2008 year, the fees per exam were $86. Students with demonstrated financial need could pay $56, and some states have further assistance to encourage students' participation in the program, but AP courses are not offered uniformly across public schools (Solorzano and Ornelas, 2004). Thus, AP scores would measure the privilege the students have had, rather than their ability or mathematical readiness for college. This is likely to be an undesirable outcome if the university is to serve a purpose of advancing the education of all in the society, and it needs empirical testing. The variables that we chose explained less than 10 percent of the variance in the dependent variables, which suggests that there might be other variables that need to be considered as well. Another possibility lies in the progressive pedagogy—group work, conceptual work, and small class sizes—employed by the control group. Perhaps we would have found a significant treatment effect if the control group employed the more traditional lectures and large class sizes. We are unable to test this possibility given the conditions under which the calculus programs are run at U-M; but there might be other settings in which it would be possible to test it.

Perhaps the second possible explanation is valid, namely that there is an effect, but that we have not been able to measure it. It is possible that our dependent variables (created by using institutional data from *student grades*) do not adequately capture the effects of the course. When we use grades as a measure of student learning, we make a gross reduction of what students have actually gained—a single course grade encompasses a complex array of knowledge and skills and may not be assigned with uniformity. Better dependent variables would be in the form of valid and reliable tests that compare the ability of students to complete real-life science and engineering applications that build on the skills for individual and group work supported by both courses. An independent measure of concept understanding, such as those provided by concept inventories (Epstein, 2005; Hestenes, Wells, and Swackhamer, 1992) could be useful in terms of overcoming some of these difficulties with course grades. The Calculus Concept Inventory (Epstein, 2005) has proven adequate for measuring students' learning of understanding of a first course of calculus (functions, limits, derivation, and growth). However, it would not provide an appropriate learning assessment for Calculus II.

Although both Math 116 and Math 156 have similar content, the style of instruction for the two is quite different. Whereas Math 116 emphasizes teacher-student and student-student interactions in every class, teacher-student interactions in Math 156 are promoted in the form of students asking questions, but student-student interactions are less frequent. Students in the Regular Calculus II are required to work in groups both inside and outside the class whereas students in Applied Honor Calculus II are encouraged to work in groups outside of the class as needed. In addition, in Math 116 there is strong emphasis on using multiple representations of concepts (geometric, numerical, symbolic, and verbal) in all solutions whereas in Math 156 instructors spend substantial time presenting proofs of theorems, emphasizing the main ideas behind the proof rather than the technical details, and discussing the relation between applications and the theorems. Yet student grades (our only current institutional measures) do not target any of these differences.

V. Implications

We suggest two practical implications of this work. First we highlight the importance of including AP scores in models that predict college students' academic performance. Our initial analyses included some models without this variable, and those models showed a statistically significant effect (stronger than other prior performance measures) of the treatment in our dependent variables. Once the AP score was included in the model, though, the effects disappeared; students who scored 5 on the AP test might indeed have a stronger preparation in calculus, and either course might be simply a refresher for them. The students, as a group, may also share other characteristics that might be captured by this variable (e.g., higher income that allows them to go to better schools, an inclination for doing challenging work, or a genuine interest in engineering). Thus we encourage other researchers to consider including this variable in their analyses as it might prove useful in better estimating the impact of students' pre-college characteristics in their college experiences, as we did not locate studies that used this variable in their prediction models.

A second implication is that the study of the impact of a single course on subsequent student performance is problematic. Instead, sustained exposure to the treatment, under very well controlled settings, might be needed if institutional data are to be used to establish a causal link between mode of instruction and student performance. In addition there is a need for parallel longitudinal studies that focus on specific aspects of instruction (e.g., instructors posing questions that require different formats of interaction) under different conditions, using both independent measures of learning and institutional data to accurately measure the learning that happens and its effects on students' performance. Such a strategy will allow researchers to make sound claims about impact of instruction on learning and student performance.

Acknowledgments

This research was supported by the Office of the Associate Dean for Undergraduate Education in the College of

Engineering and by the Office of Vice President for Research at U-M. Special thanks to the Office for Resource Planning and Management in the College of Engineering for providing the data for the study and to Robert Krasny and Karen Rhea for providing valuable feedback on earlier versions of the manuscript.

References

Bernold, L.E., W.L. Bingham, P.H. McDonald, and T.M. Attia. 2000. Influence of learning type oriented teaching on academic success of engineering students. *Journal of Engineering Education* 89 (1): 191–99.

Bernold, L.E., J.E. Spurlin, and C.M. Anson. 2007. Understanding our students: A longitudinal-study of success and failure in engineering with implications for increased retention. *Journal of Engineering Education* 96 (3): 263–76.

Budny, D., W. LeBold, and G. Bjedov. 1998. Assessment of the impact of freshman engineering courses. *Journal of Engineering Education* 87 (4): 405–11.

Burtner, J. 2005. The use of discriminant analysis to investigate the influence of non-cognitive factors on engineering school persistence. *Journal of Engineering Education* 94 (3): 335–38.

DesJardins, S., D.A. Ahlburgh, and B.P. McCall. 2002. A temporal investigation of factors related to timely degree completion. *Journal of Higher Education* 73 (5): 555–81.

Epstein, J. 2005. *Development and validation of a Calculus Concept Inventory*. New York: Polytechnic Institute, Brooklyn.

Felder, R.M. 1995. A longitudinal study of engineering student performance and retention: IV. Instructional methods and student responses to them. *Journal of Engineering Education* 84 (4): 361–67.

Felder, R.M., G.N. Felder, and E.J. Dietz. 1998. A longitudinal study of engineering student performance and retention: V. Comparison with traditionally taught students. *Journal of Engineering Education* 87 (4): 469–80.

Felder, R.M., G.N. Felder, M. Mauney, H.J. Charles, and E.J. Dietz. 1995. A longitudinal study of engineering student performance and retention: III. Gender differences in student performance and attitudes. *Journal of Engineering Education* 84 (2): 151–63.

Felder, R.M., K.D. Forrest, L. Baker-Ward, E.J. Dietz, and P.H. Mohr. 1993. A longitudinal study of engineering student performance and retention: I. Success and failure in the introductory course. *Journal of Engineering Education* 82 (1): 15–21.

Felder, R.M., P.H. Mohr, E.J. Dietz, and L. Baker-Ward. 1994. A longitudinal study of engineering student performance and retention: II. Differences between students from rural and urban backgrounds. *Journal of Engineering Education* 83 (3): 209–17.

French, B.F., J.C. Immekus, and W.C. Oakes. 2003. A structural model of engineering students success and persistence. Paper presented at 33rd ASEE/IEEE Frontiers in Education. Boulder, CO.

French, B.F., J.C. Immekus, and W.C. Oakes. 2005. An examination of indicators of engineering students' success and persistence. *Journal of Engineering Education* 94 (4): 419–25.

Ganter, S. 1999. An evaluation of calculus reform: A preliminary report of a national study. In *Assessment practices in undergraduate mathematics MAA Notes # 49*, eds. B. Gold, S.Z. Keith, and W.A. Marion. Washington, DC: Mathematical Association of America.

Heckman, J.J. 1979. Sample selection bias as a specification error. *Econometrica* 47 (1): 153–62.

Hestenes, D., M. Wells, and G. Swackhamer. 1992. Force concept inventory. *The Physics Teacher* 30 (3): 141–58.

Hoit, M., and M.W. Ohland. 1998. The impact of a discipline-based introduction to engineering course on improving retention. *Journal of Engineering Education* 87 (1): 79–85.

Hughes-Hallett, D., A. Gleason, W.G. McCallum, D. Lomen, D. Lovelock, J. Tecosky-Feldman, T. Tucker, D.E. Flath, J. Thrash, K. Rhea, A. Pasquale, S. Gordon, D. Quinney, and P.F. Lock. 2005. *Calculus single variable*. 5th ed. Hoboken, NJ: John Wiley & Sons.

Ishitani, T.T., and S. DesJardins. 2002. A longitudinal investigation of dropout from college in the United States. *Journal of College Student Retention* 4 (2): 173–201.

Kolb, D.A. 1986. *User's guide for the learning styles inventory*. Boston, MA: McBer.

Krasny, R., and V. Mesa. 2005. *Survey of factors influencing student selection of freshman math courses at the University of Michigan*. Ann Arbor: University of Michigan.

Krasny, R., and V. Mesa. 2006. *Survey of factors influencing student selection of freshman math courses: Follow-up*. Ann Arbor: University of Michigan.

Martin, J.K., J. Mitchell, and T. Newell. 2004. Work in progress: Analysis of reliability of the fluid mechanics concept inventory. Paper presented at 34th ASEE/IEEE Frontiers in Education. Savannah, Georgia.

Morgan, S.L., and D.J. Harding. 2006. Matching estimators of causal effects. Prospects and pitfalls in theory and practice. *Sociological Methods and Research* 35 (1): 3–60.

Shadish, W.R., T.D. Cook, and D.T. Campbell. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Solorzano, D.G., and A. Ornelas. 2004. A critical race analysis of Latina/o and African American advanced placement enrollment in pubic high-schools. *High School Journal* 83 (3): 15–26.

Stewart, J. 2003. *Calculus single variable*. 5th ed. Belmont, CA: Brooks/Cole.

Wage, K.E., J.R. Buck, C.H. G. Wright, and T.B. Welch. 2005. The signals and systems concept inventory. *IEEE Transactions on Education* 48 (3): 448–61.

Authors' Biographies

Vilma Mesa is assistant professor in the School of Education. She has a B.S. in computer sciences and a B.S. in mathematics from the University of Los Andes in Bogotá, Colombia, and a master's and a Ph.D. in mathematics education from the University of Georgia. She studies the role that resources play in developing teaching expertise in undergraduate mathematics and she has been involved in several evaluation projects that analyze the impact of innovative teaching practices in mathematics for students in STEM fields. Currently she is studying mathematics instruction in community colleges.

Address: 3111 SEB, 601 East University, Ann Arbor, MI, 48109-1259; telephone: (+1) 734.647.0628; fax: (+1) 734.763.1368; e-mail: vmesa@ umich.edu.

Ozan Jaquete is a Ph.D. candidate in the Center for the Study of Higher and Postsecondary Education and Organizational and Behavioral Management at the University of Michigan. A graduate from George Washington University, he also holds a M. Phil.

from Wolfson College at Oxford in comparative social policy. He has studied the effect of performance funding on student achievement in English further education colleges.

Address: 2117 SEB, 601 East University, Ann Arbor, MI, 48109-1259; telephone: (+1) 734.764.9472; fax: (+1) 734.764.2510; e-mail: ozanj@umich.edu.

Cynthia J. Finelli is director of the Center for Research on Learning and Teaching (CRLT) North and associate research scientist of Engineering Education. She holds a B.S.E., M.S.E., and Ph.D. in electrical engineering from U-M. Previously, Dr. Finelli was the Richard L. Terrell Professor of Excellence in Teaching, founding director of the Center for Excellence in Teaching and Learning, and associate professor of electrical engineering at Kettering University. Her current research interests include evaluating methods to improve teaching, exploring ethical decision-making in engineering, and assessing the effect of the first year experience on retention of under-represented students. She is chair of the Educational Research and Methods Division (ERM) of ASEE and guest co-editor for a special issue of the *International Journal of Engineering Education* on applications of engineering education research.

Address: CRLT North, 4117 Engineering Research Building I, 2200 Bonisteel Blvd, Ann Arbor, MI 48109-2099; telephone: (+1) 734.764.0244; fax: (+1) 734.647.3600; e-mail: cfinelli@umich.edu.