Instructional Quality and Student Learning in Higher Education: Evidence from Developmental Algebra Courses

Matthew M. Chingos

Brown Center on Education Policy
Brookings Institution
1775 Massachusetts Ave NW
Washington, DC 20036

mchingos@brookings.edu

Forthcoming in *The Journal of Higher Education*

Abstract

Little is known about the importance of instructional quality in American higher education because few recent studies have had access to direct measures of student learning that are comparable across sections of the same course. Using data from two developmental algebra courses at a large community college, I find that student learning varies systematically across instructors and is correlated with observed instructor characteristics including education, full-time status, and experience. Instructors appear to have effects on student learning beyond their impact on course completion rates. A variety of robustness checks suggest that these results do not appear to be driven by non-random matching of students and instructors based on unobserved characteristics or censoring of the dependent variable due to students who drop the course before the final exam.

Introduction

It is well-documented that student learning varies substantially across classrooms in elementary and secondary schools (Hanushek & Rivkin, 2010). Yet very little is known about the importance of instructional quality in America's colleges and universities. If instructional quality in higher education varies significantly across classrooms in the same course at the same campus, then reforming instructor recruitment, professional development, and retention policies could have significant potential to improve student outcomes. Unlike K-12 schools, many postsecondary institutions—especially community colleges and less-selective four-year colleges—have significant staffing flexibility because they employ a large (and growing) percentage of part-time and untenured faculty, and thus are particularly well-positioned to make good use of such evidence.

This paper empirically examines how much student learning varies across different sections of the same courses at a large community college campus in California. This paper contributes to a relatively sparse existing literature by applying modern statistical methods to student-level administrative data in order to examine the association between student performance on common assessments and instructor characteristics, both measured and unmeasured.

There is a significant related literature aimed at measuring the validity of student ratings of courses and instructors by examining the relationship between these ratings and student achievement on a common exam across multiple sections of the same course. Cohen (1981) reviews 41 such multi-section validity studies, which cover 68 multi-section courses and were mostly published in the 1970s. Most of these studies suffer from significant methodological limitations, such as the lack of random assignment or other means to control for differences in entering student ability. Of the 68 courses covered by the review, 69 percent provided no evidence of baseline equivalence of student characteristics across sections and 63 percent did not include any statistic controls for student ability.[1]

Feldman (1989) reanalyzed data from Cohen's (1981) meta-analysis, with a focus on studies that examined student ratings of specific instructional practices instead of overall ratings of instructors. Both meta-analyses find evidence of positive correlations between student ratings and student achievement, although those relationships vary significantly across studies and are subject to the methodological caveats discussed above. The production of multi-section validity studies more or less stopped after the mid-1980s (Galbraith, Merrill, & Klein, 2012). Additionally, several of the studies that have been conducted more recently raise serious questions about whether student evaluations are a good predictor of actual learning (Carrell & West, 2010; Galbraith, Merrill, and Klein, 2012; Sheets, Topping, & Hoftyzer, 1995; and Weinberg, Hashimoto, & Fleisher, 2010).[2]

The present study is related to multi-section validity studies in that it uses a consistent measure of student achievement to compare outcomes across multiple sections of the same courses. But it is not a multi-section validity study in the traditional sense because it does not seek to validate course surveys, but rather to systematically investigate the empirical importance of instructors using modern statistical methods. In this sense, the present study is most closely related to a significant body of work documenting that K-12 teachers have a significant impact on student achievement but their effectiveness is at most weakly correlated with observable

[1] Cohen (1981) finds no evidence of a statistically significant correlation between these study design features and the instructor/achievement correlation. However, standard errors are not reported and given the sample size (67) it seems unlikely that qualitatively meaningful relationships can be ruled out with a reasonable degree of confidence.

[2] There are also studies that use student evaluations as a measure of the quality of teaching. For example, Umbach and Wawrzynski (2005) examine the correlation between student engagement (based on student surveys) and faculty self-reported practices at the institutional level. However, this line of work is of limited relevance given that the validity of survey-based measures of the quality of instruction is still a contested subject.

characteristics such as education and general measures of ability. These studies attempt to measure the “value-added” that teachers have to their students’ achievement by comparing the progress that students of different teachers make on standardized tests, taking into account the demographic characteristics of the students, classrooms, and schools.

Hanushek and Rivkin (2010) summarize a number of studies in K-12 education and find that a one standard deviation increase in value-added is associated with an average increase in student achievement of 0.15 and 0.11 standard deviations in math and reading, respectively. This research has been challenged on the grounds that value-added measures are biased by non-random sorting of students into different teachers’ classrooms (Rothstein 2009). However, value-added measures have been validated by random assignment studies (Kane and Staiger 2008; Kane, McCaffrey, Miller, & Staiger 2013), a strong quasi-experimental analysis (Chetty, Friedman, & Rockoff 2014a), and long-term outcomes such as college attendance and income (Chetty, Friedman, & Rockoff 2014b).

The goal of the present study is to add to the small number of studies that have begun to extend the value-added literature into postsecondary education using modern methods, namely student-level administrative data, as rich a set of control variables as possible, and repeated observations of the same instructors over time. The most credible study of this topic estimated instructor effects on student performance in courses at the U.S. Air Force Academy, and found that a standard deviation in professor quality corresponds to 0.05 standard deviations in student achievement (Carrell & West, 2010). Other studies include Bettinger and Long’s (2004) analysis of Ohio data, which does not include any direct measures of student learning; Hoffman and Oreopoulos’s (2009) study of course grades at a Canadian university; and Watts and Bosshardt’s (1991) study of an economics course at Purdue University in the 1980s.

In sum, there are very few recent empirical studies of the variation in student performance across different sections of the same course, and only one recent study from the U.S. that included direct measures of student learning and dealt with potential non-random matching of students and instructors. The dearth of evidence on postsecondary student learning is in large part the result of data limitations. The kinds of standardized measures of learning outcomes common in K–12 education are rare at the college level, so the key challenge for research in this area is to gather student-level data from courses with reasonably large enrollments that have administered the same summative assessment to students in all sections of each course for several semesters. Common final exams are uncommon in practice because they present logistical challenges to the institution (such as agreeing on the content of the exam and finding space to administer it at a common time) and run against traditions of faculty independence.

This paper overcomes many of these limitations by using data from Glendale Community College in California, which has administered common final exams in two developmental algebra courses for the past decade. Developmental, or remedial, courses cover material that is below college level with the aim of preparing students to take college-level coursework. Remedial courses at community colleges form a significant slice of American higher education. Forty-four percent of U.S. undergraduates attend community colleges (American Association of Community Colleges, 2012), and 42 percent of students at two-year colleges take at least one remedial course (National Center for Education Statistics, 2012).

I use these data to assess both student learning and instructional quality. “Student learning” refers to student mastery of algebra (a subject most students should have been exposed to in high school), which in this paper is measured using scores on common final exams, as described below. “Instructional quality” refers not to any measure of actions taken in the classroom (such as observations of class sessions), but rather to the full set of classroom interactions that affect student learning, including the ability of the instructor, the quality of instruction delivered by that instructor (including curriculum, teaching methods, etc.), and other classroom-level factors such as peer effects. I measure the quality of instruction as how well students in a given section perform on the common final exam.

My analysis of data from eight semesters that cover 281 sections of algebra taught by 76 unique instructors indicates that student learning varies systematically across instructors and is correlated with observed instructor characteristics including education, full-time status, and experience. Importantly, instructors appear to have effects on student learning beyond their impact on course completion rates. These results do not appear to be driven by non-random matching of students and instructors based on unobserved characteristics, but should not be regarded as definitive given the limited scope of the dataset.

Institutional Background and Data

This study takes advantage of a sophisticated system of common final exams that are used in two developmental math courses, elementary and intermediate algebra, at Glendale Community College (GCC). GCC is a large, diverse campus with a college-credit enrollment of about 25,000 students.[3] New students are placed into a math course based on their score on a math placement exam unless they have taken a math course at GCC or another accredited college or have a qualifying score on an AP math exam. The first course in the main GCC math sequence is arithmetic and pre-algebra. This paper uses data from the second and third courses, elementary algebra and intermediate algebra. These courses are both offered in one- and two-

[3] About GCC, http://glendale.edu/index.aspx?page=2.

semester versions, and the same common final exams are used at the end of the one-semester version and the second semester of the two-semester version. Students must pass elementary algebra with a C or better in order to take intermediate algebra, and must achieve a C or better in intermediate algebra in order to begin taking college-level math classes.[4]

The algebra common final system has existed in its current form for about five years. The exams are developed for each course (each semester) by a coordinator, who receives suggestions from instructors. However, instructors do not see the exam until shortly before it is administered. In order to mitigate cheating, two forms of the same exam are used and instructors do not proctor the exams of their own students. Instructors are responsible for grading a randomly selected set of exams using right/wrong grading of the questions, which are all open-ended (i.e. not multiple-choice). The questions on the common final largely fall into one of the following categories of tasks: graphing equations, simplifying expressions, and solving an equation or set of equations for one or more unknowns.[5] Instructors do maintain some control over the evaluation of their students in that they can re-grade their own students' final exams using whatever method they see fit (such as awarding partial credit), but only the grade based on right/wrong grading (and thus consistent across instructors) is available in my data extract.[6]

My data extract includes the number of items correct (usually out of 25 questions) for each student that took the final exam in the eight semesters from spring 2008 through fall 2011. The common final exam data are linked to administrative data on students and instructors obtained from GCC's institutional research office. The administrative data contain 14,220

[4] "Glendale Community College Math Sequence Chart," April 2012, http://www.glendale.edu/Modules/ShowDocument.aspx?documentid=16187.
[5] Sample exams are available at http://www.glendale.edu/index.aspx?page=699.
[6] The grade assigned by the student's instructor is the one that is factored into students' course grades, but anecdotal evidence indicates that Glendale administrators use the results of the common final exam to discourage grade inflation by instructors. Conversations with administrators also indicate that the common final results are not used as part of any formal evaluation system, but are sometimes used to spot potential problems (other than grade inflation) with individual instructors, especially part-time instructors.

observations of 8,654 unique students. Background data on students include their math placement level (upon entry to the college), race/ethnicity, gender, receipt status of a Board of Governors (BOG) fee waiver (a proxy for financial need), birth year and month, units (credits) completed, units attempted, and cumulative GPA (with the latter three variables measured as of the beginning of the semester in which the student is taking the math course). The administrative records also indicate the student's grade in the algebra course and the days and times the student's section met.

The student records are linked to data on instructors using an anonymous instructor identifier. The instructor data, which cover 76 unique instructors of 281 sections over eight semesters, include education level (master's, doctorate, or unknown), full-time status, birth year and month, gender, ethnicity, years of experience teaching at GCC, and years of experience teaching the indicated course (with both experience variables top-coded at 12 years).

Table 1 shows summary statistics for students and instructors by algebra course (statistics for instructors are weighted by student enrollment). Each record in the administrative data is effectively a course attempt, and 23 percent of the records are for students who dropped the course in the first two weeks of the semester. Of these students, about one fifth enrolled in a different section of the same course in the same semester. Given the significant fall-off in enrollment early in the semester, Table 1 shows summary statistics for both the original population of students enrolled in the course and the subgroup remaining enrolled after the early-drop deadline. However, excluding these students does not qualitatively alter the pattern of summary statistics, so I focus my discussion on the statistics based on all students.

Glendale students are a diverse group. About one-third are Armenian, and roughly the same share are Latino, with the remaining students a mix of other groups including no more than

10 percent white (non-Armenian) students. Close to 60 percent are female, about half are enrolled full-time (at least 12 units), two-thirds received a BOG waiver of their enrollment fees, and the average student is 24 years old. The typical student had completed 27 units as of the beginning of the semester, and the 90 percent who had previously completed at least one grade-bearing course at Glendale had an average cumulative GPA of approximately a C+. Student characteristics are fairly similar in elementary and intermediate algebra, except that intermediate students are less likely to be Latino, have modestly higher grades and more credits completed, and (unsurprisingly) higher math placement levels.

The typical instructor is a part-time employee with a master's degree who teaches a section of 52-55 students that drops to 41-42 students by two weeks into the semester. Only 10-14 percent have doctoral degrees, and terminal degree is unknown for 20 percent. Full-time instructors teach 16-19 percent of students, and the average instructor has 6-7 years of experience teaching at Glendale Community College, with 4-5 of those years teaching the algebra course.[7]

Student success rates, in terms of the traditional metrics of course pass rates, are similar in the two algebra courses, as shown in Table 2. Just under 80 percent make it past the two-week early-drop deadline, 58 percent complete the course (i.e. don't drop early or withdraw after the early-drop deadline), just over half take the final exam, just under half earn a passing grade, and 36-38 percent earn a grade of C or better (needed to be eligible to take the next course in the sequence or receive transfer credit from another institution). Among students who do not drop early in the semester, close to two-thirds take the final, most of whom pass the course (although a significant number do not earn a C or better).

[7] Full-time instructors are more likely to have a doctoral degree than part-time instructors, but only by a margin of 25 vs. 10 percent (not shown in Table 1). In other words, the majority of full-time instructors do not have doctoral degrees.

The typical student who takes the final exam answers 38 percent of the questions correctly in elementary algebra and 32 percent in the intermediate course. The distribution of scores (number correct out of 25) is shown in Figure 1 for the semesters in which a 25-question exam was used. Students achieve a wide range of scores, but few receive very high scores. In order to facilitate comparison of scores across both courses and semesters, I standardize percent correct by test (elementary or intermediate) and semester to have a mean of zero and standard deviation of one. I associate a student's final exam score only with the records corresponding to their successful attempt at completing the course; I do not associate it with records corresponding to sections that they switched out of.

The common final exams used at GCC are developed locally, not by professional psychometricians, and thus do not come with technical reports indicating their test-retest reliability, predictive validity, etc. However, I am able to validate the elementary algebra test by estimating its predictive power vis-à-vis performance in intermediate algebra. Table A1 shows the relationship between student performance in beginning algebra, as measured by final grade and exam score, and outcomes in intermediate algebra. The final grade and exam score are fairly strong correlated (r=0.79) so the multivariate results should be interpreted with some caution. Table A1 indicates that a one-standard-deviation increase in elementary algebra final exam score is correlated with an increase in the probability of taking intermediate algebra of 13 percentage points, an increase in the probability of passing with a C or better of 17 percentage points, and an increase in the intermediate exam score of 0.57 standard deviations. The latter two of these three correlations are still sizeable and statistically significant after controlling for the letter grade received in elementary algebra.

Methodology

I estimate the relationship between student outcomes in elementary and intermediate algebra and the characteristics of their instructors using regression models of the general form:

$$Y_{ijct} = \alpha + \beta * T_{jt} + \delta * X_{it} + \gamma_{ct} + \epsilon_{ijct},$$

where Y_{ijct} is the outcome of student *i* of instructor *j* in course *c* in term *t*, α is a constant, T_{jt} is a vector of instructor characteristics, X_{it} is a set of student control variables, γ_{ct} is a set of course-by-term fixed effects, and ϵ_{ijct} is a zero-mean error term. Standard errors are adjusted for clustering by instructor, as that is the level at which most of the instructor characteristics vary. All models are estimated via ordinary least squares (OLS), but qualitatively similar results are obtained using probit models for binary dependent variables.

The instructor characteristics included in the model are education (highest degree earned), full-time status, and years of experience teaching at GCC. I also include dummy variables identifying instructors with missing data, but only report the coefficients on those variables if there are a non-trivial number of instructors with missing data on a given variable. Student controls, which are included in some but not all models, include race/ethnicity, indicator for receipt of a BOG waiver, age, full-time status, cumulative GPA at the start of the term (set to zero when missing, with these observations identified by a dummy variable), units completed at the start of the term, and math placement level. The course-by-term effects capture differences in the difficulty of the test across terms and algebra levels (elementary and intermediate), as well as any unobserved differences between students in the same algebra level but different courses (i.e. the one- vs. two-semester version).

I also estimate models that replace the instructor characteristics with instructor-specific dummies. These models are estimated separately by semester and include course dummies as well as student control variables. Consequently, the estimated coefficients on the instructor dummies indicate the average outcomes of the students of a given instructor in a given semester compared to similar students that took the same course in the same semester with a different instructor. I also create instructor-level averages of the instructor-by-term estimates that adjust for sampling variability using the Bayesian shrinkage method described by Kane, Rockoff, and Staiger (2007). This adjustment shrinks noisier estimates of instructor effects (e.g., those based on smaller numbers of students) toward the mean for all instructors.

The primary outcomes examined in this paper are: whether the student takes the final exam, whether the student passes the course with a grade of C or better (needed to progress to the next course in the math sequence), and the student's standardized score on the final exam. The estimates thus indicate the correlation between instructor characteristics (or the identity of individual instructors, in the case of the fixed effects models) and student outcomes, conditional on any control variables included in these models. These estimates cannot be interpreted as the causal effect of being taught by an instructor with certain characteristics (or a specific instructor) if student assignment to sections is related to unobserved student characteristics that influence achievement in the course. For example, if highly motivated students on average try to register for a section with full-time (rather than part-time) instructors, then the estimate of the difference between full- and part-time instructors will be biased upwards.

As discussed above, the fact that students in K-12 education are not randomly assigned to classrooms is well-documented, but there is also evidence that "value-added" models that take into account students' achievement prior to entering teachers' classrooms can produce teacher effect estimates that are not significantly different from those obtained by randomly assigning students and teachers. The challenges to the identification of causal effects related to the non-random matching of students and instructors may be more acute in postsecondary education for at least two reasons. First, the prior-year test scores that serve as a proxy for student ability and other unmeasured characteristics in much research on K–12 education are not usually available in higher education. This study is able to partly overcome this concern using a relatively rich set of control variables that include cumulative GPA at the beginning of the semester. Additionally, I am able to estimate results for intermediate algebra that condition on the elementary algebra final exam score (for students who took both courses at GCC during the period covered by my data).

Second, college students often select into classrooms, perhaps based on the perceived quality of the instructor (as opposed to being assigned to a classroom by a school administrator, as is the case in most elementary and secondary schools). At GCC, students are assigned a registration time when they can sign up for classes, and certain populations of students receive priority, including former foster children, veterans, and disabled students. Discussions with administrators at GCC indicate that students sometimes select sections based on instructor ratings on the “Rate my Professor” web site (GCC does not have a formal course evaluation system), but, anecdotally, this behavior has decreased since the use of common final exams has increased consistency in grading standards. An approximate test for the extent to which non-random matching of students to instructors affects the estimates report below is to compare results with and without control variables. The fact that they are generally similar suggests that sorting may not be a significant problem in this context.

To the extent that students do non-randomly sort into classrooms, they may have stronger preferences for classes that meet at certain days/times than they do for specific instructors. However, the descriptive statistics disaggregated by course meeting time shown in Table A2 indicate that any sorting that occurs along these lines is not strongly related to most student characteristics. A few unsurprising patterns appear, such as the proclivity of part-time and older students to enroll in sections that meet in the evening. Table A2 includes a summary measure of student characteristics: the student's predicted score on the final exam based on their characteristics (estimated using data from the same course for all semesters except for the one in which the student is enrolled). This metric indicates that students enrolled in sections with later start times have somewhat more favorable characteristics in terms of their prospective exam performance, but not dramatically so.

I also use these predicted scores to examine whether instructors are systematically matched to students with favorable characteristics. Specifically, I aggregate the predicted scores to the instructor-by-term level and calculate the correlation between the average predicted scores of an instructor's students in a given term and in the previous term. The correlation, while non-zero, is relative weak (r=0.20). Excluding students who drop the course early in the semester (or switch to another section), another source of sorting, further reduces the correlation to r=0.11. In the results section below I show that these correlations are much weaker than the term-to-term correlation in instructors' estimated effects on students' actual exam scores.

An additional complication in the analysis of learning outcomes in postsecondary education is the censoring of final exam data created by students dropping the course or not taking the final.[8] In the algebra courses at GCC, 47 percent of students enrolled at the beginning of the semester do not take the final exam. Students who do not take the final exam have predicted scores (based on their characteristics) 0.18 standard deviations below students who do take the exam. The censoring of the exam data will bias the results to the extent that more effective instructors are able to encourage students to persist through the end of the course. If the

[8] For an earlier discussion of this issue, see Sheets and Topping (2000).

marginal students perform below average, then the average final exam score of the instructor's students will understate her true contribution.

Data aggregated to the instructor-by-term level indicate that the share of students that take the final is negatively correlated with the average exam score—more students taking the final means a lower score, on average. But the correlation is quite weak (r=-0.15), suggesting that much of the variation in section-level performance on the exam is unrelated to attrition from the course. This would be the case if, for example, dropout decisions often have non-academic causes such as unexpected financial or family issues.

I also address the issue of missing final exam data for course dropouts below by estimating models that impute missing final exam scores in two different ways. First, I make the most pessimistic assumption possible by imputing missing scores as the minimum score of all students in the relevant course and term. Second, I make the most optimistic assumption possible by using the predicted score based on student characteristics. This assumption is optimistic because the prediction is based on students who completed the course, whereas the drop-outs obviously did not and thus are unlikely to achieve scores as high as those of students with similar characteristics who completed the course. Below I show that the general pattern of results is robust to using the actual and imputed scores, although of course the point estimates are affected by imputing outcome data for roughly half the sample.

Results

I begin with a simple analysis of variance that produces estimates of the share of variation in student outcomes in algebra that are explained by various combinations of instructor and student characteristics. I estimate regression models of three outcomes—taking the final exam, passing with a grade of C or better, and final exam score—and report the adjusted r-squared value in Figure 2.[9] The baseline model includes only term-by-course effects, which explain very little of the variation in student outcomes. Adding instructor characteristics (education, full-time status, and experience teaching at GCC) increases the share of variance explained by a small amount, ranging from about 0.5 percent for final taking and successful completion rates to about 1 percent for final exam scores.

Replacing instructor characteristics with instructor fixed effects has a more noticeable effect on the share of variance explained, increasing it by 1.9-2.5 percent for the course completion outcomes and by almost 8 percent for final exam scores. Adding student controls to the model, which themselves explain 10-18 percent of the variation in outcomes, does not alter the pattern of results without controls: instructor education, full-time status, and experience explain much less variation in outcomes than instructor fixed effects.

The estimated relationships between instructor characteristics and student outcomes in pooled data for elementary and intermediate algebra are reported in Table 3.[10] Education is the only variable that is statistically significantly related to the rates at which students take the final exam and successfully complete the course (with a C or better): the students of instructors with doctoral degrees are 5-7 percentage points less likely to experience these positive outcomes as

[9] I obtain qualitatively similar results using unadjusted r-squared.
[10] Results disaggregated by algebra level are available from the author upon request. These results are less precisely estimated and although the coefficients generally point in the same direction as the pooled results, some of the patterns observed in the pooled results are stronger in one course than in the other.

compared to the students of instructors with master's degrees. Students of full-time instructors are 3-4 percentage points more likely to take the final and earn a C or better than students of part-time instructors, but these coefficients are not statistically significant. The point estimates for instructor experience do not follow a consistent pattern.

Instructor characteristics are more consistent predictors of student performance on the final exam. Having an instructor with a doctoral degree, as compared to a master's degree, is associated with exam scores that are 0.15-0.17 standard deviations lower (although only the result with student controls is statistically significant and only at the 10 percent level). The students of full-time instructors scored 0.21-0.25 standard deviations higher than their counterparts in the classrooms of part-time instructors. Returns to instructor experience at GCC are not consistently monotonic, but suggest a large difference between first-time and returning instructors of about 0.20 standard deviations. This result could reflect both returns to experience as well as differential attrition, in which higher quality instructors are more likely to remain at GCC.[11]

The coefficient estimates are not substantially altered by the addition of student-level control variables, suggesting that students do not sort into sections in ways that are systematically related to both their academic performance and the instructor characteristics examined in Table 3. Given the dearth of data on student learning in postsecondary education, the estimated coefficients on the control variables, reported in Table A3, are interesting in their own right. Results that are consistent across all three outcomes include higher performance by Asian students and lower performance by black and Latino students (all compared to white/Anglo students), and better outcomes for older students and for women. One of the

[11] In separate models (not shown) I replace instructor experience at GCC with experience teaching the specific course and do not find any consistent evidence of returns to this measure of experience.

strongest predictors of outcomes is cumulative GPA at the start of the term, with an increase of one GPA point (on a four-point scale) associated with an increase in final exam score of 0.39 standard deviations. Students who are new to GCC (about 10 percent of students), as proxied by their missing a cumulative GPA, outperform returning students by large margins.[12]

Tables 4 and 5 report the results of three robustness checks. First, I include controls for the time of day that the course meets to account for any unobserved student characteristics associated with their scheduling preferences. Adding this control leaves the results largely unchanged (second column of Table 4). Second, using imputed likely minimum and maximum scores for students who did not take the final exam has a larger impact on the point estimates, as would be expected from roughly doubling the sample, but the general pattern of results is unchanged (last two columns of Table 4). The experience results are the most sensitive to this change.

Finally, I estimate a "value-added" type model for intermediate algebra scores only where elementary algebra scores are used as a control variable. The advantage of this model is that the elementary algebra score is likely to be the best predictor of performance in intermediate algebra, but this comes at the cost of only being able to use data from one of the two courses, and only for students who completed the lower-level course at GCC during the period covered by my data. Consequently, the results are much noisier than the main estimates, and Table 5 indicates that simply restricting the sample to students with elementary algebra scores available in the data changes the results somewhat, especially the estimated returns to experience. However, adding the elementary algebra score as a control leaves the pattern of results largely unchanged. In this

[12] The difference between new and returning students may reflect selection into taking math courses early vs. late in the college career, with students who take math courses early being more confident in their math ability and having taken math more recently (i.e. in high school).

analysis, the most robust finding is the substantial difference in student performance between instructors with doctoral and master's degrees (in favor of the latter).

The outcomes examined thus far are all measured during the term that the student is in the instructor's class. It could be the case that instructors work to maximize performance on the final exam at the expense of skills that have longer-term payoffs, as in Carrell and West's (2009) study of the U.S. Air Force Academy (although in their case the short-term outcome was course evaluations). In Table A4, I show the estimated relationship between the characteristics of elementary algebra instructors and three outcomes of their students after taking the course: whether they take intermediate algebra, whether they pass with a C or better, and their score on the intermediate algebra exam.[13]

The results are imprecisely estimated given the reduced sample size. The point estimates for full-time instructors are generally positive, but usually not large enough to be statistically significant. The results for experience indicate, for the final exam only, that students of first-year instructors fare better in the follow-on course than those of veteran instructors, but this result is fairly sensitive to the inclusion of control variables and is based on only 27 percent of the students who took elementary algebra. In sum, the results in Table A4 do not bolster the results based on immediate learning outcomes, but they are not convincing enough to undermine them either.

The analysis of variance indicated that instructor fixed effects explain a much greater share of the variation in learning outcomes than the handful of instructor characteristics available in the administrative data. As explained in the methodology section above, I estimate instructor-

[13] The intermediate algebra taking and completing variables are defined for all students, whereas the final exam is only defined for students who took the final at some point in the period covered by my data. Additionally, I will misclassify as non-takers (and non-completers) students who took intermediate algebra during the summer or in a self-paced version (both of which are not included in my data extract).

by-section effects separately by term and include the same student-level controls used in the analysis of instructor characteristics. The standard deviation of the estimated instructor effects on taking the final exam and the final exam score are shown in Table 6. The 267 estimated instructor-by-term effects have a standard deviation of 0.11 for taking the final (i.e. 11 percentage points) and 0.37 for the exam score (i.e. 0.37 student-level standard deviations). The correlation between the two is -0.10, similar to the correlation for the unadjusted data discussed above.

Part of the variability in the instructor-by-term effect estimates results from sampling variation, especially with the relatively small numbers of students enrolled in individual classrooms (and even smaller number that take the final). This variability will tend to average out over multiple terms. The second row of Table 6 shows that averaging all available data for the 76 instructors in the data produces a standard deviation of instructor-level effects of 0.09 for taking the final and 0.31 for the exam score. Shrinking these estimates to take into account the signal-to-noise ratio further reduces the standard deviations to 0.05 and 0.21, respectively. These estimates are also robust to using a random effects specification (estimated using hierarchical linear models), as shown by the last two rows of Table 6.

The fact that the standard deviations of the estimated instructor effects remain substantial after adjusting for sampling variability in the fixed effects specifications is due in part to the relatively strong correlation between the estimated effects of the same instructor over time. Figure 3 shows the relationship between the effect estimate for each instructor and the estimate for the same instructor in the prior term that she taught, which have a correlation of r=0.56.

The relative stability of instructor effects over time suggests that they are capturing something persistent about the quality of instruction being delivered. As a further check on these results, I estimate the relationship between student performance in algebra and the effectiveness of the instructor measured using data from all semesters other than the current one.[14] This means that idiosyncratic variation in student performance specific to a student's section will not be included in the estimated instructor effect. Table 7 shows that the instructor effect is a powerful predictor of student performance on the final exam, but not on the likelihood that the student will take the final or pass the course. An increase in the estimated instructor effect of one standard deviation (measured in student scores) is associated with a 0.95-standard-deviation in student scores. Given the standard error, I cannot reject the null hypothesis of a one-for-one relationship.

Table 7 also shows the relationship between the estimated effects of elementary instructors and student outcomes in the follow-on course (intermediate algebra). Students who had an elementary instructor with a larger estimated effect are no more likely to take intermediate algebra, but are more likely to complete the course successfully. These students are also predicted to score higher on the intermediate algebra common final, but this relationship is imprecisely estimated and its statistical significance is not robust to excluding the eight percent of test-takers who had the same instructor in both elementary and intermediate algebra.

Limitations

The results reported in this study suggest that individual instructor effects are potentially more important for student learning in developmental algebra courses than observed instructor characteristics. It is not surprising that variation in the quality of instruction related to both observed and unobserved characteristics is greater than the variation explained by observed characteristics on their own, but these findings should still be interpreted cautiously given that

[14] Specifically, I average the instructor-by-term effects for all terms except for the one during which the student is enrolled.

they are based on data from a relatively modest number of instructors (76) in a handful of courses at a single institution.

This limitation highlights a significant challenge of conducting research on student learning in higher education. Unlike in K-12 education, where it is possible to obtain administrative data on standardized test performance for an entire district or state, research efforts in higher education will usually be limited to a single course or handful of courses that use common assessments. This means that the total number of instructors will likely not be very large, and that the results will be based on a very specific context—no more than a handful of courses on a single campus.

This study examined data from developmental algebra courses on a community college campus in California that enrolls a unique mix of students (including a large Armenian population). Clearly this limits the extent to which the results can be extrapolated to other contexts, such as higher-level courses, campuses in other states, four-year campuses, and courses in other subjects, just to name a few potentially relevant dimensions. A college would have to make a systematic effort to encourage the use of common finals across a wide variety of courses, and incorporate the resulting data into its administrative records, for a more systematic single study to be possible.

Another limitation of this study is that the measure of instructional quality, performance on common assessments, does not capture instructor actions. Gathering any objective information on student learning that is consistent across different instructors is an important first step toward better understanding the role of instructional quality. Whether the quality of instruction can be measured using various non-test-based methods, such as student surveys and observations by faculty and administrators, is a ripe subject for future research. Some of the multi-section validity studies discussed above indicate that there is a positive relationship between student course ratings and student achievement, and there is also some research from K-12 education documenting a relationship between certain types of student survey questions and performance on standardized tests (MET Project, 2012). Expanding this line of research to include modern methods and raters other than students has clear potential value.

Any single measure of student learning or instructional quality is going to be limited, both by the inherent limitations of the measure itself (e.g., the properties of a test) and by associated methodological challenges (e.g., missing values, data censoring, selection bias issues, etc.). Efforts to incorporate these measures into policy and practice should acknowledge these limitations and avoid the dangers of focusing too much on any single measure.

Conclusion

The limitations discussed above make clear that the results of this study should not be assumed to hold across a wider array of courses and institutions. At the same time, the findings raise interesting questions about teaching and learning in higher education. The finding that full-time instructors outperform part-time instructors, on average, suggests that providing full-time employment (and the associated compensation and benefits) either attracts better employees, retains them for longer, or enhances quality by enabling them to focus on their teaching responsibilities at a single institution rather than cobbling together work across multiple institutions.

The finding that instructors with master’s degrees are more effective in the classroom, on average, than instructors with doctorates is more difficult to explain. One speculative possibility is that PhD-level instructors are more interested in teaching upper-level math courses, or students with strong interest and ability in math, which is likely very different from teaching remedial algebra courses. Another possibility is that instructors with master's degrees are more likely to have sought employment at a community college, whereas those with PhDs may have preferred other options (such as a more research-focused institution). Descriptive empirical work on the labor market for postsecondary instructors is a ripe area for future research.

More generally, despite the limitations of any analysis of data from a small number of courses, this paper exemplifies the kind of work that can be done with data on student learning that are comparable across sections of a course taught by different instructors—data that are rarely available in American higher education. Importantly, it shows that examining only course completion rates can miss important variation in student learning. Students who complete a course vary widely in their mastery of the material, which influences their likelihood of success in follow-on courses. The absence of random assignment of students and teachers is likely to be a challenge in all research on this subject, although the GCC data offer preliminary evidence that this is not as important a concern as it is in other contexts.

The study does not examine instructional practices, and thus cannot offer guidance to individual instructors on how to improve their craft. But the fact that meaningful differences in the quality of instruction delivered exist suggests that faculty and administrators might be able to improve student outcomes through evaluating instructors and providing targeted feedback based on data such as performance on common assessments. The new generation of K-12 teacher evaluation systems currently, which combine information on student growth on standardized tests with observation-based feedback from administrators and master teachers (Whitehurst, Chingos, & Lindquist, 2014), might serve as a starting point for similar efforts in higher education. The observation component can serve both to provide actionable feedback to instructors and make possible research that links specific teacher behaviors to student learning.

References

American Association of Community Colleges. (2012). 2012 Community College Fast Facts. Available online at http://www.aacc.nche.edu/AboutCC/Documents/FactSheet2012.pdf.

Bettinger, E., & Long, B. T. (2004). Do college instructors matter? The effects of adjuncts and graduate assistants on students' interests and success. Cambridge, MA: National Bureau of Economic Research Working Paper No. 10370.

Carrell, S. E., & West, J. E. (2009). Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Public Economics,* 118, 409–432.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the impacts of teacher I: Evaluating bias in teacher value-added estimates. *American Economic Review,* 104, 2593–2632.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the impacts of teacher II: Teacher value-added and student outcomes in adulthood. *American Economic Review,* 104, 2633–2679.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, 51, 281–309.

Feldman, K. A. (1989). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education,* 30, 583–645.

Galbraith, C. S., Merrill, G. B., & Kline, D. M. (2012). Are student evaluations of teaching effectiveness valid for measuring student learning outcomes in business related classes? A neural network and Bayesian analyses. *Research in Higher Education,* 53, 353–374.

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review*, 100, 267–271.

Hoffman, F., & Oreopoulous, P. (2009). Professor qualities and student achievement. *Review of Economics and Statistics*, 91, 83–92.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). Have we identified effective teachers? Validating measures of effective teaching using random assignment. Seattle, WA: Bill and Melinda Gates Foundation.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2007). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, 27, 615–631.

Kane, T. J., & Staiger, D. O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation. Cambridge, MA: National Bureau of Economic Research Working Paper No. 14607.

MET Project (2012). Asking students about teaching: Student perception surveys and their implementation. Seattle, WA: Bill and Melinda Gates Foundation.

National Center for Education Statistics. (2012). *Digest of education statistics 2011*. Washington, DC: US Department of Education.

Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy*, 4, 537–571.

Sheets, D. F., & Topping, E. E. (2000). Assessing the quality of instruction in university economics courses: Attrition as a source of self-selection bias in mean test scores. *The Journal of Economics*, 26, 11–21.

Sheets, D. F., Topping, E. E., & John Hoftyzer, J. (1995). The relationship of student evaluations of faculty to student performance on a common final examination in the principles of economics courses. *The Journal of Economics*, 21, 55–64.

Umbach, P. D., & Wawrzynski, M. R. (2005). Faculty do matter: The role of college faculty in student learning and engagement. *Research in Higher Education*, 46, 153–184.

Watts, M., & Bosshardt, W. (1991). How instructors make a difference: Panel data estimates from principles of economics courses. *Review of Economics and Statistics*, 73, 336–340.

Weinberg, B. A., Hashimoto, M. & Fleisher, B.M. (2010). Evaluating teaching in higher education. *The Journal of Economic Education*, 40, 227–261.

Whitehurst, G. J., Chingos, M. M, & Lindquist, K. M. (2014). Evaluating teachers with classroom observations: Lessons learned in four districts. Washington, DC: Brookings Institution.

Figures

Figure 1. Distribution of Scores on Final Exams with 25 Questions

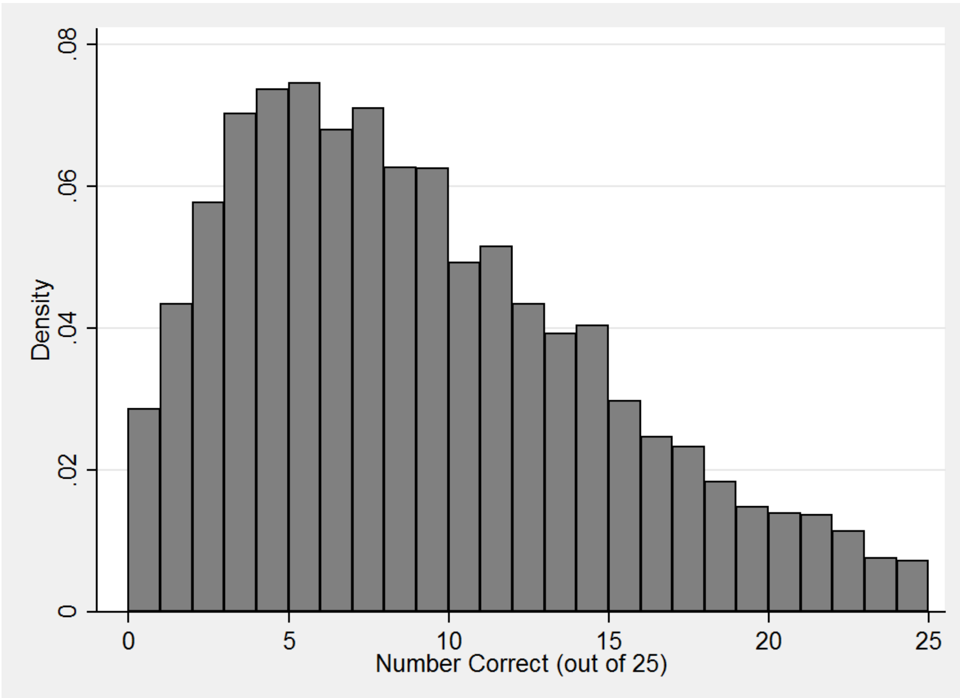
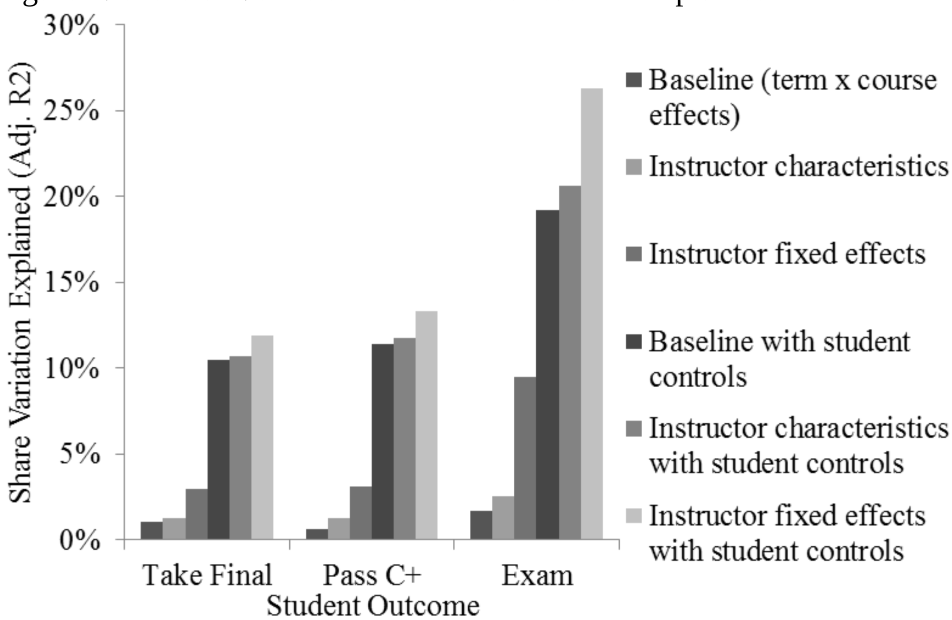


Figure 2. Share of Variation in Student Outcomes Explained



Notes: Instructor characteristics include education, full-time status, and experience teaching at GCC. Student controls include race/ethnicity, gender, BOG waiver, age, full-time status, cumulative GPA at start of semester (set to zero when missing, with these observations identified by a dummy variable), units completed at the start of the term, and math placement level.

Figure 3. Semester-to-Semester Stability in Instructor Effect Estimates Based on Exam Scores (Correlation=0.56)

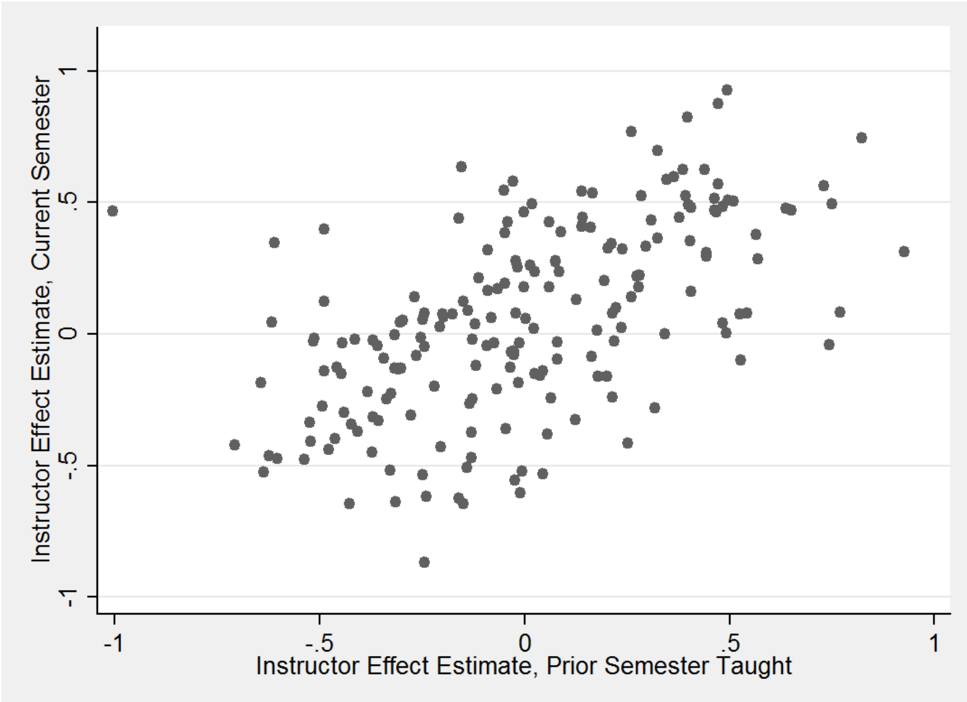


Table 1. Student and Instructor Summary Statistics, by Algebra Course

	Including Early Drops		Excluding Early Drops	
	Elementary	Intermediate	Elementary	Intermediate
Student race				
Armenian	34%	35%	34%	36%
Asian	4%	8%	4%	8%
Black	2%	2%	2%	2%
Filipino	4%	5%	4%	5%
Latino	32%	25%	32%	24%
White	8%	10%	8%	10%
Other/missing	16%	15%	15%	15%
Female	58%	55%	59%	55%
Sex missing	1%	1%	1%	1%
BOG waiver	67%	61%	68%	61%
Age	24.3	23.4	24.0	23.2
Full-time student	40%	49%	45%	55%
Cum. GPA, start of semester	2.29	2.42	2.34	2.44
Cum. GPA missing	10%	10%	10%	9%
Units completed, start of semester	27.1	30.7	26.9	30.6
Math placement level				
Level 1	13%	5%	13%	5%
Level 2	17%	8%	18%	7%
Level 2.5	3%	0%	3%	0%
Level 3	31%	12%	32%	12%
Level 3.5	6%	10%	6%	10%
Level 4+	0%	31%	0%	32%
Missing	29%	34%	28%	33%
Predicted final exam score	-0.11	-0.09	-0.09	-0.08
Section size	55.1	52.3	54.8	52.1
Section size after early drops	42.0	40.5	42.3	40.7
Instructor education				
Master's	70%	66%	71%	66%
Doctorate	10%	14%	10%	14%
Unknown	19%	20%	19%	20%
Instructor full-time	16%	19%	17%	20%
Instructor exp, course	4.5	4.6	4.5	4.6
Exp in course missing	4%	3%	4%	3%
Instructor exp, college	6.3	6.6	6.4	6.6
Exp at college missing	2%	2%	2%	2%
Observations (student records)	5,600	8,620	4,298	6,690
Observations (unique students)	4,014	6,146	3,518	5,459
Observations (unique sections)	113	168	113	168
Observations (unique instructors)	49	67	49	67

Table 2. Student Outcomes

	Elementary	Intermediate
Don't drop course early	77%	78%
Complete course	57%	58%
Take final	52%	54%
Pass course	45%	47%
Pass with C or better	36%	38%
Conditional on not dropping early		
Complete course	74%	74%
Take final	64%	65%
Pass	59%	60%
Pass with C or better	47%	48%
Final, percent correct	38%	32%
Final, std. dev.	24%	22%

Table 3. Relationship Between Instructor Characteristics and Student Outcomes, Elementary and Internmediate Algebra

	No Controls			With Controls		
	TakeFinal	Pass C+	Score	TakeFinal	Pass C+	Score
Instructor's education (relative to Master's)						
Doctorate	-0.055	-0.070	-0.145	-0.048	-0.063	-0.167
	(0.021)*	(0.020)**	(0.098)	(0.018)**	(0.019)**	(0.093)+
Unknown	-0.010	-0.042	0.089	0.001	-0.029	0.110
	(0.023)	(0.023)+	(0.097)	(0.021)	(0.022)	(0.091)
Full-time instructor	0.038	0.042	0.205	0.026	0.037	0.250
	(0.026)	(0.029)	(0.088)*	(0.022)	(0.025)	(0.077)**
Instructor's exp, GCC (relative to 0 years)						
1-2 years	0.018	0.011	0.186	-0.013	0.000	0.212
	(0.035)	(0.050)	(0.128)	(0.039)	(0.052)	(0.091)*
3-5 years	0.006	-0.003	0.236	-0.029	-0.014	0.280
	(0.036)	(0.047)	(0.139)+	(0.037)	(0.048)	(0.120)*
6+ years	-0.043	-0.079	0.228	-0.054	-0.071	0.281
	(0.039)	(0.048)	(0.142)	(0.043)	(0.051)	(0.117)*
Observations	14,218	14,218	7,133	14,217	14,217	7,133
R-squared	0.016	0.016	0.031	0.110	0.122	0.212

Notes: ** p<0.01, * p<0.05, + p<0.1; robust standard errors adjusted for clustering by instructor appear in parentheses. All regressions include course-by-term fixed effects and dummies identifying instructors with missing full-time status and experience. Controls include student race/ethnicity, BOG waiver, age, full-time status, cumulative GPA at start of semester (set to zero when missing, with these observations identified by a dummy variable), units completed at the start of the term, and math placement level.

Table 4. Relationship Between Instructor Characteristics and Exam Scores, Elementary and Intermediate Algebra, Robustness Checks

	Preferred	Time Controls	Impute Min	Impute Max
Instructor's education (relative to Master's)				
Doctorate	-0.167	-0.138	-0.158	-0.087
	(0.093)+	(0.086)	(0.058)**	(0.050)+
Unknown	0.110	0.099	0.062	0.061
	(0.091)	(0.086)	(0.055)	(0.044)
Full-time instructor	0.250	0.207	0.166	0.120
	(0.077)**	(0.077)**	(0.063)*	(0.042)**
Instructor's exp, GCC (relative to 0 years)				
1-2 years	0.212	0.223	0.089	0.090
	(0.091)*	(0.109)*	(0.076)	(0.041)*
3-5 years	0.280	0.300	0.113	0.113
	(0.120)*	(0.121)*	(0.078)	(0.062)+
6+ years	0.281	0.343	0.059	0.120
	(0.117)*	(0.120)**	(0.077)	(0.059)*
Observations	7,133	7,133	14,217	14,217
R-squared	0.212	0.216	0.159	0.320

Notes: ** p<0.01, * p<0.05, + p<0.1; see notes to Table 3. All models include student controls.

Table 5. Relationship Between Instructor Characteristics and Student Outcomes, Intermediate Algebra, Controlling for Elementary Scores

	Controls		Restrict Sample		Control Elem Score	
	TakeFinal	Score	TakeFinal	Score	TakeFinal	Score
Instructor's education (relative to Master's)						
Doctorate	-0.061	-0.211	-0.043	-0.171	-0.047	-0.187
	(0.017)**	(0.094)*	(0.038)	(0.095)+	(0.037)	(0.090)*
Unknown	0.009	0.104	-0.054	0.176	-0.054	0.149
	(0.025)	(0.104)	(0.046)	(0.102)+	(0.043)	(0.099)
Full-time instructor	0.010	0.128	-0.079	0.129	-0.083	0.084
	(0.023)	(0.071)+	(0.046)+	(0.082)	(0.046)+	(0.080)
Instructor's exp, GCC						
1-2 years	-0.086	0.258	0.106		0.163	
	(0.030)**	(0.160)	(0.091)		(0.103)	
3-5 years	-0.113	0.321	0.059	0.009	0.127	0.103
	(0.030)**	(0.192)+	(0.091)	(0.142)	(0.104)	(0.140)
6+ years	-0.149	0.294	0.056	0.164	0.102	0.126
	(0.029)**	(0.191)	(0.091)	(0.150)	(0.103)	(0.152)
Elementary algebra					0.077	0.464
score (standardized)					(0.013)**	(0.028)**
Observations	8,620	4,371	1,870	1,060	1,870	1,060
R-squared	0.119	0.206	0.148	0.303	0.179	0.449

Notes: ** p<0.01, * p<0.05, + p<0.1; see notes to Table 3. All models include student controls. Elem score controls include standardized score on elementary algebra final, whether elementary algebra course was one or two semesters, and number of semesters since taking elementary algebra.

Table 6. Standard Deviations of Instructor Effect Estimates

	TakeFinal	Score
Fixed effects		
Instructor*term (N=267)	0.113	0.368
Instructor (N=76), unshrunk	0.088	0.307
Instructor (N=76), shrunk	0.051	0.213
Random effects		
Instructor*term (N=267)	0.086	0.323
Instructor (N=76)	0.071	0.305

Notes: Instructor*section fixed effects are estimated separately by term, and include student controls listed in notes to Table 3. Random effects are estimated in an HLM with data pooled across all terms. The instructor*term HLM includes a single level (instructor*term/course) and the instructor HLM includes two levels (instructor*term/course and instructor). Both HLMs include the same control variables as in the fixed effects model.

Table 7. Relationship Between Elementary Algebra Instructor Effect Estimate and Student Outcomes in Elementary and Intermediate Algebra

	Elementary and Intermediate		
	TakeFinal	Pass C+	Score
Estimated effect of instructor	-0.048	-0.011	0.952
	(0.046)	(0.046)	(0.065)**
Observations	7,827	7,827	3,919
R-squared	0.114	0.125	0.261

	Intermediate, Terms Prior to Fall 2011				
	TakeCourse	TakeFinal	Pass C+	Score	Score
Estimated effect of elementary instructor	-0.011	0.071	0.076	0.446	0.338
	(0.056)	(0.039)+	(0.038)+	(0.204)*	(0.246)
Exclude same instructor?	No	No	No	No	Yes
Observations	2,430	2,430	2,430	653	604
R-squared	0.143	0.131	0.132	0.173	0.166

Notes: ** p<0.01, * p<0.05, + p<0.1; see notes to Table 3. All models include student controls. The estimated effect of elementary instructor excludes data from the semester the student took the course. Data from final semester (fall 2011) are excluded from the intermediate algebra results due to lack of next-semester data. "Exclude same instructor?" indicates whether intermediate algebra students who had the same instructor in beginning and intermediate algebra are excluded. All regressions include course-by-term fixed effects. In the bottom panel, controls are based on those at the beginning of the elementary algebra semester.

Table A1. Relationship Between Elementary Algebra Performance and Student Outcomes in Intermediate Algebra

	Take Intermdiate Algebra			Pass with C or better			Score on Common Final Eaxm		
Student's grade (relative to A [17%])									
B (23%)	-0.020		-0.008	-0.121		-0.046	-0.863		-0.401
	(0.030)		(0.036)	(0.034)**		(0.042)	(0.074)**		(0.099)**
C (33%)	-0.067		-0.047	-0.307		-0.175	-1.245		-0.496
	(0.029)*		(0.041)	(0.033)**		(0.050)**	(0.085)**		(0.128)**
D (17%)	-0.427		-0.400	-0.503		-0.330	-1.286		-0.309
	(0.030)**		(0.049)**	(0.029)**		(0.051)**	(0.127)**		(0.179)+
F (9%)	-0.536		-0.504	-0.551		-0.345	-1.076		0.037
	(0.033)**		(0.052)**	(0.030)**		(0.057)**	(0.194)**		(0.232)
Elem algebra final		0.125	0.012		0.172	0.079		0.572	0.454
exam score (std)		(0.010)**	(0.017)		(0.010)**	(0.018)**		(0.028)**	(0.053)**
Mean of dep var	0.61	0.61	0.61	0.34	0.34	0.34	0.00	0.00	0.00
Observations	2,383	2,383	2,383	2,383	2,383	2,383	1,027	1,027	1,027
R-squared	0.201	0.112	0.201	0.167	0.146	0.177	0.284	0.341	0.370

Notes: ** p<0.01, * p<0.05, + p<0.1; see notes to Table 3. All regressions include course-by-term fixed effects. Data from final semester (fall 2011) are excluded due to lack of next-semester data. Models with grades also include a dummy identifying the 1% of students that did not receive a letter grade. The linear correlation between grade (on a four-point scale) and exam score (standardized) in elementary algebra is 0.79.

Table A2. Student and Instructor Summary Statistics, Excluding Early Drops, by Time of Day

	Early	Morning	Afternoon	Evening
Student race				
Armenian	33%	33%	35%	41%
Asian	6%	6%	8%	5%
Black	2%	2%	2%	1%
Filipino	5%	5%	5%	4%
Latino	29%	31%	25%	23%
White	8%	9%	10%	10%
Other/missing	16%	14%	15%	16%
Female	54%	54%	59%	56%
Sex missing	1%	0%	1%	1%
BOG waiver	66%	65%	63%	62%
Age	22.5	22.2	23.8	26.3
Full-time student	54%	58%	53%	32%
Cum. GPA, start of semester	2.27	2.39	2.45	2.43
Cum. GPA missing	7%	11%	9%	9%
Units completed, start of semester	29.5	26.3	30.3	32.5
Math placement				
Level 1	3%	8%	8%	11%
Level 2	6%	12%	13%	11%
Level 2.5	2%	1%	1%	2%
Level 3	14%	21%	19%	23%
Level 3.5	10%	9%	9%	7%
Level 4+	22%	22%	19%	14%
Missing	43%	26%	31%	33%
Predicted final exam score	-0.15	-0.11	-0.06	-0.02
Section size	52.4	52.9	53.5	53.6
Section size after early drops	40.7	42.2	41.1	40.3
Instructor education				
Master's	77%	63%	71%	65%
Doctorate	9%	18%	7%	15%
Unknown	15%	19%	22%	20%
Instructor full-time	3%	36%	17%	0%
Instructor exp, course	5.1	5.0	3.9	4.4
Exp in course missing	3%	3%	2%	6%
Instructor exp, college	8.6	6.6	5.8	5.7
Exp at college missing	0%	2%	0%	6%
Observations (student records)	1,552	3,991	3,527	1,918
Observations (unique sections)	39	99	92	50

Notes: Early classes conclude at or before 9am, morning classes start before noon (but conclude after 9am), afternoon classes start after noon but before 6pm, and evening classes start at 6pm or later.

Table A3. Coefficients on Control Variables

	Take Final	Pass C+	Score
Student race/ethnicity (relative to white/Anglo)			
Asian	0.049	0.054	0.225
	(0.018)**	(0.019)**	(0.060)**
Black	-0.139	-0.106	-0.256
	(0.031)**	(0.029)**	(0.089)**
Filipino	0.064	0.044	-0.031
	(0.021)**	(0.021)*	(0.074)
Latino	-0.044	-0.032	-0.086
	(0.016)**	(0.014)*	(0.046)+
White/Armenian	0.053	0.035	0.054
	(0.016)**	(0.017)*	(0.040)
Other/missing	0.006	0.021	0.020
	(0.018)	(0.017)	(0.048)
Female	0.063	0.062	0.050
	(0.009)**	(0.008)**	(0.022)*
Gender missing	-0.062	-0.019	0.039
	(0.053)	(0.052)	(0.173)
BOG waiver	0.025	0.016	0.020
	(0.010)*	(0.010)	(0.021)
Student age (years)	0.001	0.005	0.026
	(0.001)+	(0.001)**	(0.002)**
Full time	0.146	0.080	-0.036
	(0.011)**	(0.010)**	(0.022)+
Cumulative GPA at	0.103	0.139	0.388
start of term	(0.007)**	(0.008)**	(0.025)**
Cum GPA missing	0.230	0.303	0.879
	(0.022)**	(0.022)**	(0.087)**
Units completed at start	0.001	-0.000	-0.005
of term	(0.000)**	(0.000)	(0.001)**
Math placement level (relative to Level 1)			
Missing	-0.053	-0.053	-0.017
	(0.025)*	(0.023)*	(0.056)
Level 2	0.051	0.026	0.214
	(0.021)*	(0.023)	(0.045)**
Level 2.5	0.055	0.020	0.058
	(0.047)	(0.044)	(0.106)
Level 3	0.055	0.027	0.048
	(0.022)*	(0.024)	(0.043)
Level 3.5	0.021	0.025	0.144
	(0.023)	(0.027)	(0.059)*
Level 4+	0.097	0.077	0.345
	(0.022)**	(0.023)**	(0.061)**
Observations	14,217	14,217	7,133
R-squared	0.110	0.122	0.212

Notes: ** p<0.01, * p<0.05, + p<0.1; robust standard errors adjusted for clustering by instructor appear in parentheses. All regressions include course-by-term fixed effects and instructor characteristics (education, full-time status, and experience at GCC).

Table A4. Relationship Between Elementary Algebra Instructor Characteristics and Student Outcomes in Intermediate Algebra

	No Controls			With Controls		
	Take	Pass C+	Score	Take	Pass C+	Score
Instructor's education (relative to Master's						
Doctorate	-0.013	0.046	-0.032	-0.009	0.045	0.017
	(0.032)	(0.023)+	(0.082)	(0.039)	(0.028)	(0.081)
Unknown	-0.009	0.017	0.053	0.013	0.039	0.111
	(0.025)	(0.021)	(0.082)	(0.024)	(0.020)+	(0.068)
Full-time instructor	0.054	0.012	0.101	0.039	0.009	0.079
	(0.027)*	(0.021)	(0.093)	(0.027)	(0.019)	(0.094)
Instructor's exp, GCC (relative to 0 years)						
1-2 years	0.033	0.006	-0.180	0.032	0.017	-0.060
	(0.043)	(0.038)	(0.067)*	(0.045)	(0.032)	(0.059)
3-5 years	0.008	-0.006	-0.362	0.018	0.017	-0.241
	(0.037)	(0.033)	(0.085)**	(0.037)	(0.024)	(0.083)**
6+ years	-0.000	0.007	-0.288	0.015	0.036	-0.153
	(0.037)	(0.033)	(0.100)**	(0.035)	(0.025)	(0.097)
Observations	4,823	4,823	1,299	4,822	4,822	1,299
R-squared	0.045	0.021	0.017	0.134	0.125	0.144

Notes: ** p<0.01, * p<0.05, + p<0.1; see notes to Table 3. All models include student controls. Data from final semester (fall 2011) are excluded due to lack of next-semester data. All regressions include course-by-term fixed effects and dummies identifying instructors with missing full-time status and experience.