NBER WORKING PAPER SERIES

PROFESSOR QUALITIES AND STUDENT ACHIEVEMENT

Florian Hoffmann
Philip Oreopoulos

Working Paper 12596
http://www.nber.org/papers/w12596

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2006

We are indebted to Marianne Lynham, Matthew Hendrickson, Josie Lalonde, Rick Hayward, and Karel Swift for helping obtain and work with the data for this project and to Rita Ahmadyar, Nicole Corbett, and Paul Ternamian for outstanding research assistance. We are also grateful to Steve Rivkin, Brian Jacob, Thomas Kane, Jonah Rockoff, and Doug Staiger for very helpful discussion, and to conference participants at the Society of Labor Economists Annul Meetings (Boston), the Higher Education Group of the National Bureau of Economic Research, and the Conference on Education, Training and the Evolving Workplace (Vancouver). Oreopoulos thanks the Social Sciences and Research Council for financial support. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Professor Qualities and Student Achievement
Florian Hoffmann and Philip Oreopoulos
NBER Working Paper No. 12596
October 2006
JEL No. H52,I2

ABSTRACT

This paper uses a new administrative dataset of students at a large university matched to courses and instructors to analyze the importance of teacher quality at the postsecondary level. Instructors are matched to both objective and subjective characteristics of teacher quality to estimate the impact of rank, salary, and perceived effectiveness on grade, dropout and subject interest outcomes. Student fixed effects, time of day and week controls, and the fact that first year students have little information about instructors when choosing courses helps minimize selection biases. We also estimate each instructor's value added and the variance of these effects to determine the extent to which any teacher difference matters to short-term academic outcomes. The findings suggest that subjective teacher evaluations perform well in reflecting an instructor's influence on students while objective characteristics such as rank and salary do not. Whether an instructor teaches full-time or part-time, does research, has tenure, or is highly paid has no influence on a college student's grade, likelihood of dropping a course or taking more subsequent courses in the same subject. However, replacing one instructor with another ranked one standard deviation higher in perceived effectiveness increases average grades by 0.5 percentage points, decreases the likelihood of dropping a class by 1.3 percentage points and increases in the number of same-subject courses taken in second and third year by about 4 percent. The overall importance of instructor differences at the university level is smaller than that implied in earlier research at the elementary and secondary school level, but important outliers exist.

Florian Hoffmann
Department of Economics
University of Toronto
140 St. George St. Suite 707
Toronto, Ontario M5S3G7
Canada
flhoffmann@gmx.de

Philip Oreopoulos
Department of Economics
University of Toronto
150 St. George Street
Toronto, ON M5S 3G7
Canada
and NBER
oreo@economics.utoronto.ca

I. Introduction

Teaching is the central task colleges and universities perform for students. Policy administrators often emphasize teaching as the key determinant to a college student's academic experience and successful transition into the labour force. Many university and college mission statements declare that graduates should leave with strong analytical abilities, communication skills, and be primed for fulfilling careers. Students themselves, on the other hand, often complain colorfully in teaching evaluations about how ineffective some instructors have been in helping them meet these goals. "I would rather eat glass," for example, is one student's response to whether he or she would take one of our (anonymous) colleague's courses again. As post secondary school enrolment continues to increase along with tuition, colleges and universities face renewed demands for better teaching and student experience. This paper analyzes the extent to which teaching matters to students' academic achievement and course selection, and whether observable teacher characteristics can predict these outcomes.

At the primary and secondary school level, the literature on the effects of teacher quality and how to measure it is extensive. Starting with the Coleman report in 1966, many have argued that teacher quality matters little and that families and peers are far more important in determining test score and education attainment outcomes. Coleman found little evidence that primary or secondary teachers' subject expertise (measured by test scores and college performance), completion of advanced degrees, or experience relate to students' subsequent performance. Several more recent meta-analyses, however, suggest that teacher quality <u>does</u> in fact lead to higher test scores, but the mixed conclusions across studies may indicate that the size of the influence may depend on the circumstance (Hedges et al., 1994). Studies that examine the relationship between teacher quality and longer-run outcomes, such as earnings, find more consistent evidence that teacher quality matters (e.g. Card and Krueger, 1992, 1996). Rivkin, Hanushek and Kain (2005) also point out that teacher quality may differ in many ways not captured by observable qualifications or experience. Test score improvement differs substantially for students with different teachers, but in the same school and grade. Rivkin, Hanushek and Kain, and Rivkin conclude that although explanations for these differences are not readily captured by common measures of teacher quality, they nevertheless indicate teachers play an influential role. Jacob and Lefgren (2005) find some support for this using subjective principal assessments of teachers – principal evaluations of the best and worst primary school teachers predict future student achievement significantly better than measures of teacher experience, education, and actual compensation.

Research about the connection between teacher quality and student outcomes at the post-secondary level is virtually non-existent. A few studies focus on the effect of particular types of graduate assistants, but these studies rely on relatively small samples and do not have much information on student background. For example, Borjas (2000) analyzes the impact of foreign teaching assistants on economics students' performances at Harvard. More recently, Ehrenberg and Ziang (2005) examine the effects of adjuncts (part-time faculty) on student dropout rates using institutional-level data from a sample of U.S. universities. They find a negative relationship between student persistence and adjunct usage, although they cannot rule out this could be driven by the tendency for schools with higher proportion of adjuncts to also be more likely to have students on the margin of dropping out. The most closely related study to this paper's is by Bettinger and Long (2004, 2005), who estimate from an administrative dataset of public four-year universities in Ohio the effects of adjunct professors on course selection and dropout rates using year-to-year and class-to-class variation in first year instructors. They estimate that adjuncts have very small positive effects on students' picking similar subject courses in subsequent years (relative to full-time faculty), but adversely increase the likelihood that students dropout in the second year.

This paper contributes to the literature about the importance of teacher quality in several ways. It focuses on the effects of teacher quality at the college level. Previous studies usually look at grade-school teachers or measure teacher quality from basic instructor characteristics, such as experience, salary, and career status. Our paper uses both objective and subjective measures of teacher quality. We estimate average effects from ending up with a first year college instructor who is an adjunct professor paid part-time to teach, a lecturer paid full-time to teach, a tenure-track or a tenured professor. We also estimate effects from ending up with an instructor that is highly paid, or that tends to rank high or low in student responses to teacher evaluations. Including teacher evaluations in our analysis allows us to explore Rivkin et al.'s suggestion that observable instructor differences do not correlate with student achievement because they do not correlate with other, less tangible, measures of teacher quality that matter. Our identification strategy also differs from earlier studies. First-year college students take many courses taught by a variety of instructors, and many end up with different instructors teaching the same course because of differences in timetables scheduling or because of year-to-year instructor changes. This set-up facilitates the use of course and student fixed effects so that we can estimate whether differences across a student's first year instructors correlate with differences in her corresponding course or subject-related academic achievement. We also estimate each instructor's value added and the variance of these effects to determine the extent to which any teacher difference matters to short-term academic outcomes.

Using administrative data from a large Canadian university between 1996 and 2005, our findings suggest that whether an instructor teaches full-time or part-time, does research, has tenure, or is highly paid has no influence on a college student's likelihood of dropping a course or taking more subsequent courses in the same subject. Interestingly, these traits are also uncorrelated with an instructor's perceived effectiveness (evaluated by students at the end of a course and averaged over ten years). Subjective teacher evaluations perform much better in reflecting an instructor's influence on students compared to objective characteristics such as rank and salary. This influence, however, is smaller than that implied of elementary and secondary school teachers in earlier research. A one standard deviation increase in an instructor's perceived effectiveness decreases the likelihood of dropping a class by about 1.3 percentage points (compared to a course dropout rate of 9 percent). The same increase in perceived effectiveness is also associated with an average increase in the number of same-subject courses taken in second and third year by 3.0 percent of a standard deviation. The effects are similar among males and females, science and non-science majors, but they are notably more pronounced among students with relatively poor high school grades.

II. Data

The study uses student and instructor administrative data from a large Canadian university. The data cover the Fall and Winter school year periods between 1996 and 2005. We focus on the 41,402 students that entered into a full-time undergraduate Arts and Science program, and were 17 to 20 years old on September 1 in the year of entry. Full-time status means that all students were initially enrolled in courses offering credits that sum to at least 3.5. A typical student must attain 20 credits to complete an undergraduate degree.

A. Student Data

The raw student data come in four files. The Enrolment File contains student application information, including gender, date of birth, mother tongue (English, French, or Other), citizenship, entering program of study, and high school grade average. The High School File contains more detailed secondary school transcript information, including specific courses taken and corresponding grades, as well as high school identifiers. High school data are missing for some out-of-province students and we restrict our sample to the 36,144 students with non-missing high school grades. The Registration File records registration status at the start of each Fall and Winter term between 1996 and 2005. This file indicates the number of credits a student is enrolled in, financial status with the university, whether in good standing (which generally means not on academic probation, enrolled in a full-time program with fees paid), cumulative and current Grade Point Average (GPA), program of study, and graduation status. The Course File contains information on courses enrolled in and credits received for each year on September 1, November 1, January 1, March 1, and the most current status. An advantage of this file is that it allows us to match to courses that students first enrolled in before their first day of class, and regardless of whether they completed the course or not. The Course File also includes course section information and final grade received.

The first set of columns in Table 1 displays descriptive statistics for our population of first-year students. The means are typical for undergraduate students in Canada. Age at entry is 18.6, a majority of students that enrol are female, and high school grade averages are tightly distributed around the mean of about 85 percent. Annual Fall registration status shows that about 10 percent of our first-year students in our sample did not continue to register into the Fall of their second year of the program. The graduation rate among those who started before 2000 was 78 percent. One third of all students report a mother tongue other than English or French, and 10 percent are of Asian citizenship. Program at entry is almost evenly split between Science, Social Science, and undeclared.

B. Course Data

Some students in first year take uncommon courses or upper year courses. Column 2 of Table 1 shows descriptive classroom statistics for all 1,029 courses taken by at least one first year student over the 9 year period. Course selection, however, is concentrated among large first year introductory classes. These 47 largest courses, with average annual enrolment sizes of 200 or more make up 78 percent of a student's curriculum, on average. Appendix Table A1 lists these courses and their characteristics. We focus our main analysis on these core introductory courses to ensure the main results are not driven by particularly small or upper year courses, and for computational reasons. Appendix Table A2 shows that the main results remain virtually unchanged when including all classes.

Many courses are divided up each year into multiple classroom sections. We match course sections to course timetables, drop evening classes that begin after 5PM, and include time of day controls in the analysis: whether a class ever begins before 10AM, after 3PM, on Monday, or on Friday.

C. Instructor Data

Instructor evaluations are taken near the end of semester in each class. The form is anonymous and identical across all Arts and Science courses. The question that administrators focus on most for tenure decisions and teaching review are the teacher effectiveness question: "All things considered, this instructor performs effectively as a university teacher" and the retake question, "Considering your experience with this course, and disregarding your need for it to meet program or degree requirements, would you still have taken this course?" The effectiveness question is on a 7 point scale, ranging from 1 (extremely poor) to 7 (outstanding). Means are generated for teacher-course cells. Across 1,852 first year instructors, the 'mean of the classroom mean' for effective overall, is 5.6, with a standard deviation of .60. The 25th percentile instructor and the 75th percentile instructors differ by almost exactly two standard deviations (5.2 versus 6.2).

Variation around the mean of the instructor effectiveness measure is fairly narrow. Most students give evaluation ratings of 4, 5, or 6. Another way to measure subjective instructor quality is to look at frequency of extremely poor or good ratings. The average fraction of students giving an instructor an overall effectiveness rating of 1 or 2 is 5.5 percent, with one unfortunate large-course instructor receiving 1 or 2 ratings by 46 percent of his or her students. The average fraction of students giving an instructor the highest rating in effectiveness is 25.6 percent, with a high of 91 percent and a low of 0. The mean of the fraction of students that say they would take the course again is 67.8 percent, with a standard deviation of 15.4. We call this variable the retake rate. The 25th percentile instructor has a retake rate of 51 percent, while the 75th percentile instructor has a rate of 80 percent.

The instructor evaluation data include the name of the instructor, the corresponding course, section, and year, the number enrolled in the course and the number who filled out the forms. In a few instances, more than one instructor co-taught a section of a course. In these cases, teacher characteristics are averaged over both, and the pair is counted as one 'instructor'.

We use historical university course calendars to match an instructor's name to his or her corresponding position. We use an instructor's most frequent position over the nine year period to create an indicator variable for 1) whether an instructor is a lecturer, employed full-time primarily to teach (31 percent), 2) whether an instructor is an assistant or associate professor, employed full-time and expected to do research (18 percent), 3) whether an instructor is a full professor, with tenure (27 percent), or 4) whether an instructor falls into an other category (24 percent). We call this category 'part-time' because it mostly includes graduate students and adjunct professors. In addition to information about instructor position, this university also publicly discloses annual earnings for employees paid more than $100,000. We use this data to create a variable (called 'top salary') to indicate what years an instructor earned $100,000 or more, calculated in 2006 real Canadian dollars using Statistics Canada's Consumer Price Index.

Interestingly, instructor type and salary are largely unrelated to perceived effectiveness. Figure 1A shows the distributions of mean perceived effectiveness across instructors in large first year courses by instructor type. The average perceived effectiveness among lecturers, junior and full professors are similar (5.8, 5.6, and 5.6 respectively). Part-time instructors tend to receive lower evaluations (the mean is 5.3), but the variance of subjective quality within each type remains high. Figure 1B shows that the distributions of mean perceived effectiveness measure by income category are also similar.

E. Outcomes Data

Column 4 of Table 1 shows outcome data categorized by student and first year class. This is our baseline dataset used for in the analysis. Most classes taken in freshman year last two semesters and are worth 1 credit. About 15 percent of courses are Fall semester courses and another 15 percent are Winter courses, worth .5 credits. Students take, on average, 4.5 course credits. As students specialize in higher years, the average number of upper year courses in the same subject as the first year course declines, while the standard deviation increases.

Unlike standardized test scores often used in primary and secondary teacher quality studies, college course grades as outcomes are problematic because they may be adjusted by the instructor to normalize across sections or even to encourage better teacher evaluations. We include grade outcomes for interest, but focus instead on whether a student drops a course and whether a student enrols in subsequent courses in the same subject.

Across all large classes, 9.1 percent initially enrol by September 1 in a given school year but cancel before receiving a final grade. Dropping a course is significantly associated with short and long term academic achievement. Every full-year course dropped is related, on average, with a 0.5 decline in GPA among courses completed that year. Dropping a course also leads to longer completion times. Each dropped first-year course is associated with a 7 percentage point lower probability of spending at least four years in college and a 12 percentage point lower probability of graduating within five years. Students themselves choose whether to dropout so that, more likely, an instructor's perceived effectiveness may influence the dropout rate rather than the other way around.[1]

[1] It is possible that an instructor's evaluation (averaged over the 9 year period) could be worse were students allowed to evaluate instructors before dropping out. In one extreme case, all disgruntled students drop, leaving end-of-course evaluations equal and positive across courses with large and small dropout rates. The difference between end-of-class and initial class mean evaluation, however, is likely to be small

We also examine how various instructor qualities affect subject specialization. The administrative data used in this paper does not measure well a student's program or choice of major, but the number of subject-courses enrolled in is strongly related to these variables. To major in Economics, for example, a student is expected to complete three (typically out of four or five courses) in second year in this subject, after completing introductory Economics in first year. Program requirements in other subjects are similar.

III. Regression Methodology

The main analysis uses data at the student-by-course-section level. Consider the outcome of interest, y_{iktps}, for student i, in course k, in year t, with instructor (professor) p, in section s. We decompose y_{iktps} by the following:

$$(1) \qquad y_{iktps} = \beta Q_p + \gamma X_{kts} + \delta_i + \delta_k + \delta_{iktps},$$

where Q_p is a measure of subjective or objective teacher quality for instructor p, X_{kts} are time of day and time of week controls, δ_i and δ_k are student and course fixed effects respectively, and δ_{iktps} is the statistical error term. All standard error estimates incorporate residual clustering grouped by instructor.

since most students finish and the difference in perceived effectiveness between dropouts and non-dropouts is probably small. Even if it were not, we simply estimate whether variation in perceived effectiveness, measured by students who remain in class, correlates with any academic achievement. This is an empirical question, and one possibility explaining a weak correlation could be because of this type of measurement error from missing evaluations.

We are interested in estimating β, the average impact from being taught by an instructor whose quality differs by one unit of Q_p. Course fixed effects account for course specific outcome differences, so that we are identifying off the within-course variation. The dropout rate for introductory Calculus, for example, is higher than the dropout rate for Psychology and Calculus instructors tend to receive worse evaluations than Psychology instructors.

The key identification strategy for estimating β is to use instructor quality variation across each student's set of first year classes. Student fixed effects absorb tendencies for some types of individuals to enrol in particular sets of classes or take classes with particular types of instructors. A remaining bias may arise if these tendencies are not equally weighted across all courses – for example, if students who major in Economics (and are less likely to drop economics courses) care about taking the introductory course with a highly ranked instructor, but care less about who their instructors are in other courses. We focus on first year courses to reduce the likelihood of this behavior. Incoming first year students are less likely to select classes based on instructor because little is known about instructors when selecting courses before starting university, and instructors are often not listed in course calendars. Time of day and week controls also help remove a possible correlation with certain types of individuals preferring to attend or teach classes early or late in the day or preferring to avoid classes taught on Mondays or Fridays. Individual fixed effects control for student-specific selection selection behavior typical across all courses.

IV. Regression Results

Table 2 presents our main results. In column 1 of the first panel, we regress an indicator for whether a student dropped a course on year fixed effects, course fixed effects, and perceived instructor effectiveness (averaged over all student evaluations recorded between 1995 and 2004). The estimated standard errors account for clustering of residuals by instructor. The sample includes first year students between 1995 and 2004 initially enrolled in large first year classes.

A student with an instructor who receives an average perceived effectiveness evaluation of 4 is 1.3 percentage points more likely to drop a course compared to taking it with an instructor who receives an average evaluation of 5 (about a 2 standard deviation difference in instructor quality). Adding student fixed effects in column 2 and time of day and week controls in column 3 does not change the point estimate very much, which is expected, since few first year students likely choose courses based on its instructor. An instructor's experience and faculty position are insignificantly related to whether students drop a course. Students taught by a lecturer, hired full-time to teach, are 0.8 percentage points less likely to drop a course than if taught by research faculty (mostly full-time professors). We cannot reject the possibility that course dropout rates are unrelated to lecturer, faculty, or salary status. However, even conditioning on instructor rank in column 7, students are significantly more likely to cancel a course if their instructor tends to rank poorly on perceived effectiveness.

Panel 2 of Table 2 uses the number of same-subject courses taken in year 2 as the outcome variable. We use this variable as a proxy for a students' interest in specializing in a subject that corresponds with a first year introductory course. The results suggest that subjective and objective instructor qualities have minimal influence on subject specialization. From column 3, a two standard deviation increase in perceived instructor effectiveness increases the number of courses taken in second year in the same subject by 0.03 courses – 3 percent of the outcome variable's standard deviation. Instructors paid high salaries (more than $100,000 in 2005 Canadian dollars) slightly decrease the likelihood of taking more subject courses. All estimated instructor quality effects combined in column 7 are insignificantly different from zero.

Table 3 shows the same regression specification in column 7 of Table 2, but also includes a wider set of course and subject-related outcomes. Column two indicates students with instructors that tend to receive better evaluations also receive significantly higher grades, on average. Of course, one explanation is that instructors that grade easier may tend to receive better evaluations. We cannot rule out this possibility and, thus, do not place much emphasis on the grade outcome results.[2] Students taking classes with lecturers and younger professors receive a final grade about 1.5 percentage points lower than students taking classes with full professors. This estimate may only reflect the possibility that lecturers and younger professors tend to grade students worse. None of the subsequent subject grade or subject course selection outcome variables in second or third year are significantly related to the observed instructor characteristics. This is also true for year 2 and year 3 subject credits received. Receiving a subject credit requires both enrolling in a subject-related course and passing it.

---

[2] We also used Heckman's two-step consistent estimator with year fixed effects and without individual fixed effects to account for selection into the grade regression samples (which depends on enrolling in a course and completing it). We modeled being in the sample determined by high school grade, gender, and/or first year GPA. This alternative specification produces very similar results to those shown here.

Students evaluate instructors across a variety of traits, and measuring subjective quality in terms of the portion of students that provides particularly high or low evaluations may be more informative than summarizing quality using the mean across students, as done in the earlier tables. Table 4 shows estimates of the effects of alternative measures of subjective quality on student achievement. Each coefficient shown is from a separate regression of the outcome variable on instructor quality, course and student fixed effects, and time of day and week controls. Interestingly, it appears to not make much difference whether we summarize subjective quality as the mean reported overall effectiveness, the fraction of students ranking the instructor 7 out of 7 in terms of effectiveness, or the fraction of students ranking the instructor 1 or 2 out of 7. A two standard deviation increase in these instructor quality measures is associated with reduced dropout rate by 1.7, 1.4, and 1.7 percentage points respectively. The implied effects from changes to the other instructor qualities evaluated are also similar. This is not surprising because the assessed qualities are highly correlated – the correlations between instructor effectiveness and the other quality variables (e.g. organized, enthusiastic, answers questions clearly) ranges from 0.65 to 0.95. The multicolinearity is too pronounced to regress the outcome variables on multiple quality traits and generate precise coefficient estimates.

The point estimates in Table 4, with few exceptions, suggest that subjective instructor quality plays some role in determining subsequent subject courses and credits. Only a few estimates, however, are significant, and the implied magnitude of the effect is small. A two standard deviation increase in perceived enthusiasm, for example, leads to an increase in the number of subject-related courses taken the following year by 4 percent of its standard deviation. The most informative variation comes from the retake rate– the fraction of students who report, 'considering my experience with this course, and disregarding the need to take it to meet program or degree requirements, I would have still taken this course'. A two standard deviation increase in the retake rate is associated with a 3 percentage point reduction in the course dropout rate and an increase in the number of subject-related courses taken in year 2 by 7.7 percent of its standard deviation.

Tables 5 and 6 explore how the main results differ by gender, mother tongue, high school grade, and program of study. Table 5 shows results using class dropout as the outcome variable, Table 6 shows results using the number of subject-related courses taken in second year. The estimated effects from perceived effectiveness on class dropout are very similar regardless of gender, whether English is a student's mother tongue or not, and whether a student enters university as a science major or not. Lecturer status, tenure status, and top salary status are insignificantly related to course dropout for each these sub-groups.

The tables also suggest that instructor quality impacts students differently depending on their past performance. From column 5 of Table 5, the effect of perceived effectiveness on the likelihood of dropping a course is driven almost entirely by university students with poorer high school grades. Among students in the lowest high school grade quartile, a two standard deviation increase in subjective instructor quality lowers the likelihood of dropping a course by 5.6 percentage points, whereas the estimated effect among students from the top high school grade quartile is zero. Dropping a course is still insignificantly related to lecturer and salary status for both groups. But additional differences arise when looking a subsequent course selection.

Those from the lowest quartile with a better instructor in terms of perceived effectiveness enrol in more related courses the following year – a two standard deviation increase in perceived effectiveness is associated with a higher count of subject-related courses in second year by 7 percent of a standard deviation. The estimated effect among students in the highest quartile is about zero. The influence of instructor rank and salary also differs by students' high school grade quartile. Lecturers have a significant negative impact on subject interest for students among the lowest quartile, but a positive impact among students from the highest quartile. Compared with full professors, students from the lowest high school grade quartile are less likely to be interested in a subject after taking an introductory course with an assistant or associate professor, or an adjunct or emeritus professor. Highly paid professors, however, have a positive influence on subject interest among students with better high school grades.

V. Instructor Value Added

University teachers may matter in ways not captured by the observable differences we use above to estimate effects on student achievement. At the primary and secondary level, several researchers have measured teacher specific fixed effects (or value added) on student performance, and conclude that the effects differ across teachers in significant ways. Value added is the expected increase or decrease in a student's outcome from attending a class with a particular instructor, relative to the mean outcome. The value-added standard deviation indicates the extent to which any teacher differences matter in determining student performance. A zero standard deviation implies that it makes no difference, on average, to a student's performance which teacher she is assigned to.

Kane, Rockoff, and Staiger (2006) estimate teachers' value added to elementary and secondary student test scores. The results suggest that a change from the teacher with the 25th percentile value added to the teacher with the 75th percentile value added (about a 2 standard deviation difference) would affect a student's test score by about one quarter of its standard deviation. Aaronson, Barrow, and Sander (2002) estimate that one semester with a high school teacher rated two standard deviations higher in value added would add 0.3 to 0.5 grade equivalents, or 25 to 45 percent of an average school year, to a student's math score performance. Rockoff (2004) finds that the standard deviation in the value added for New Jersey elementary students' normalized test scores is 0.11. Rivkin, Hanushek, and Kain (2005) also arrive at similar estimates for Texas elementary students.

Consider a more general version of equation (1) where achievement depends simply on who teaches the course. We decompose y_{iktps} by the following:

(2) $$y_{iktps} = \delta_p + \delta_k + \delta_{kts} + \delta_{iktps}$$

where δ_p is an instructor fixed effect and the variance (or standard deviation) of δ_p is the primary parameter of interest. The outcome may depend on a class specific effect, δ_{kts}, independent of instructor. This may be because the class is held early in the morning, or because of a particular disruption (a fire alarm, for example). The outcome also depends on student specific factors, δ_{iktps}, and factors other than instructor, course, time, or section.

Even in the absence of student selection, estimating each instructor's value added, $\hat{\delta}_p$, from regression equation (2) will lead to measurement error, u: $\hat{\delta}_p = \delta_p + u$. This makes interpretation of the magnitude and variance of the fixed effects difficult. Measurement error will bias the calculation of $\text{var}(\delta_p)$ upwards [$\hat{\sigma}^2_{\delta_p} = \sigma^2_{\delta_p} + \sigma^2_u$] and previous researchers suggest this bias is often substantial. To avoid it, alternative methods or adjustments are used.

We adopt an approach similar in spirit to Kane, Rockoff, and Staiger (2006), who look at the year-to-year teacher covariances in mean student test scores across classes, schools, and time. The covariance estimation approach is also similar to one used by Page and Solon (2003) to estimate 'value added' to earnings from growing up in a particular neighborhood.[3]

For each course, the covariance between two students in the same section is:

(3)

$$C(y_{\mathrm{i=1,k^*tps}}, \mathrm{y}_{\mathrm{i\neq1,k^*tps}}) = V(\delta_p) + V(\delta_{\mathrm{k}^*ts}) + 2\mathrm{C}(\delta_p, \delta_{\mathrm{k}^*ts}) + 2\mathrm{C}(\delta_p, \delta_{i\mathrm{k}^*tps}) + 2\mathrm{C}(\delta_{\mathrm{k}^*ts}, \delta_{i\mathrm{k}^*tps}),$$

[3] For comparison, Page and Solon estimate that a one standard deviation in value added neighbourhood quality would increase expected earnings by up to 18 percent. Also for comparison, the authors estimate sibling earnings covariances to be about 0.35. This would imply a one standard deviation in 'family value added' would increase earnings by 59 percent!

Students may select by section to be with friends or because students in similar programs end up in similar classes. The within section covariance will be biased upwards by this type of within-section sorting. To better measure the permanent component of the instructor value-added variance, we also calculate the covariance between two students from the same course with the same instructor but in different sections, possibly in different years. If instructors are assigned to sections in a manner that is independent of section fixed effects[4], then the covariance across sections is:

$$(4) \qquad C(y_{\mathrm{i=1,k^*,t,p,s=1}}, y_{\mathrm{i\neq1,k^*,t,p,s\neq1}}) = V(\delta_p) + 2\mathrm{C}(\delta_p, \delta_{ik^*tps}) .$$

If students take classes independently of who teaches (so that $\mathrm{C}(\delta_p, \delta_{\mathrm{ik}^*tps}) = 0$), the covariance across sections measures the permanent value added variance, $V(\delta_p)$. To address possible selection if this is not true, we also regress y_{iktps} on course and individual fixed effects and use the residuals to compute the across and within section covariances in equations 4 and 5. Individual fixed effects should help control for overall tendencies of individuals to choose classes based on instructor or program.

We follow the covariance estimation procedure used by Solon and Page (2003) for common neighbors. Let $\mathrm{y}'_{\mathrm{iktps}}$ be the residual from regressing y_{iktps} on course and year fixed effects or course and individual fixed effects. This first step adjusts for possible outcome differences by courses or time. The covariance is calculated as follows:

[4] This assumption may not be true if instructors prefer teaching at a particular time of day, taught over less than the 9 year period the data span. In these cases, the across section covariance estimate will be biased upwards.

(5) $$\hat{C}(y'_{i,ktps}, y'_{i',ktps}) = \left[\sum_{k'=1}^{K} \sum_{t'=1}^{T} \sum_{p'=1}^{P} \sum_{s'=1}^{S_{ktp}} \sum_{i=1}^{I_{ktps}} \sum_{i'=1}^{I_{ktps}} y'_{i,k't'p's'} y'_{i'\neq i,k't'p's'} \right] / N' ,$$

where I_{ktps} is the number of individuals within section ktps, S_{ktp} is the number of sections in a course, T is the number of courses, K is the number of years, and N' is the number of observations in the numerator. We calculate the square root of the covariance for an upper-bound estimate of the standard deviation of instructor value added fixed effects. Ninety-five percent confidence regions are estimated by bootstrap.

The results are presented in Table 7. All outcome variables, except course dropout, are standardized to have mean 0 and standard deviation 1. The within-section covariance for first year standardized course grade is 0.019. Column 1 shows the square root of this value, 0.138, which is an estimate for the standard deviation of instructor fixed effects on grades for a given course. A two standard deviation difference in instructor fixed effects between students would thus account for approximately a 3.4 percentage point difference in grade performance (2*0.138*13, with 13 being the standard deviation of a first year student's grade). Since grading is often at the discretion of the instructor, we cannot attribute this variation to instructor value added. We can only conclude that a small portion of a courses's grade distribution occurs from being taught by different instructors.

The covariance estimates fall when using residuals after including individual fixed effects (in column 2), and when calculating the across-section covariance (in column 4) instead of within-section. This suggests at least some sorting of students by section and by instructor. Our preferred estimates are from columns 3 and 4. Column 3 may overstate the instructor value added variation because of possible student sorting. Column 4 may understate the variation because the individual fixed effects may absorb some of the instructor effects before we calculate the instructor effect variation from the residuals. The true variation probably lies somewhere in the middle. Fortunately for interpretation, the covariance estimates in both cases are similar for all but one of the student outcomes.[5]

The standard deviation of value added for a student's likelihood of dropping out, from columns 3 and 4, is about 0.012. This suggests, for a class size of 200, about an additional 5 students would drop the course if an instructor's value added was two standard deviations lower. This estimate is not much higher compared to that implied from the 1.4 percentage point effect due to a two standard deviation decrease in perceived instructor effectiveness.

Turning to subsequent outcomes in following years, a two standard deviation difference in the effect an instructor has on the number of same-subject courses taken in year 2 can explain about 0.07 to 0.20 of an extra subject-course, on average (6 to 16 percent of the standard deviation in the number of subject courses). The average effect on third year course selection is lower, and the year 2 and year 3 grade covariance estimates between students enrolled in same-subject courses and with the same first-year instructor are negative and not significantly different from zero (not shown in table).

[5] The estimated standard deviations for instructor fixed effects using the larger sample of first year students taking any course (not just large first year courses) are almost identical to the ones shown here.

VI. Conclusion

In this paper, we address the question whether differences in teacher quality among currently acceptable university instructors matter to students' achievement and course interest. To analyze this, we use a new administrative dataset of students at a large Canadian university matched to first year courses and corresponding instructors. Instructor quality is measured by objective, subjective, and value added measures. We identify our estimates using variation across different sections within the same course. The within-course analysis provides an intuitive counterfactual estimate of how different a student's subsequent achievement would be expected to be if she enrolled in the same course but with a different type of instructor. To control for individual specific characteristics and selection behavior we include student fixed effects. Remaining selection on teacher quality is likely to be small since, for many first-year courses, instructors are not listed in course calendars and students must pick the courses we match to (as of September 1) with little or no prior knowledge about instructors. We also control for time of day and week controls to minimize remaining selection issues.

We find that differences in commonly observed instructor traits, such as rank, faculty status, and salary, have virtually no effect on student outcomes. There are no average differences in students' dropout, subsequent grade, and course selection outcomes by instructor tenure or tenure-track status, full-time or part-time lecturer status, and salary status (whether an instructor earns more than CDN$100,000 in the year taught). These results are consistent across alternative model specifications and student populations. They suggest that, at the margin, universities can hire cheaper part-time or full-time lecturers without expecting adverse impacts to student achievement. The findings are similar to Bettinger and Long (2004), who find small and often insignificant effects on subsequent course interest from taking a first year class with an adjunct or graduate student instructor. They are also similar to Jacob and Lefgren (2005) and others who find elementary and secondary teacher experience, education status, and salary have little impact on test scores.

What does matter is instructors' perceived effectiveness and related subjective measures of quality evaluated by students. Interestingly, subjective instructor evaluations have almost no correlation with instructor rank or salary, yet vary widely within these categories. Students with instructors that tend to receive high evaluations are less likely to cancel a course, more likely to receive better grades, and somewhat more likely to take similar courses in following years. To help quantify this, consider that the average instructor ranking in perceived effectiveness among the instructors ranking in the bottom quarter is 4.8 on a 7 point scale, and the average among instructors ranking in the top quarter is 6.3. If first-year instructors ranked in the bottom quarter could be replaced with instructors ranked in the top, we estimate that the course dropout rate would fall by 2 percentage points, and the number of related courses taken in second year would increase by about 4 percent. For comparison, if we were to replace entering first-year students in this university from the bottom quarter of high school grade averages with students from the top quarter, the dropout rate would fall by 6.4 percentage points.

The second and third year effects suggest that any one instructor's influence on student achievement and behaviour fades over time. But exposure to consistently effective teachers each year or consistently ineffective ones could have a large cumulative impact. The impact is greater for students with initially poor academic achievement in high school, but about the same for males and females, science and non-science majors, and minority and non-minority students.

We also estimate the value added variance to indicate the extent to which any observable or unobservable teacher difference matters to academic outcomes. The findings suggest that subjective teacher evaluations perform well in reflecting an instructor's influence on students. Increasing an instructor's value added by one standard deviation reduces a course's dropout rate by 1.2 percentage points, which is not much larger in magnitude compared to the effect of a standard deviation increase in perceived instructor effectiveness. We interpret the overall college instructor influence on student achievement as smaller than the overall influence suggested in earlier value added research for elementary and secondary school teachers. Class grade distributions and dropout rates differ across college instructors teaching the same course, but less so compared to class grade distributions across elementary and secondary school instructors. Standardized effects from a change in instructor quality are about half the size or lower for college instructors than for elementary and secondary teachers (e.g. Kane, Rockoff, and Staiger, 2006, Rivkin, Hanushek, and Kain, 2005, and Jacob and Lefgren, 2006). Perhaps by the time students enter college cognitive ability and motivation are less malleable than in early childhood and, consequently, teachers have less impact.

References

Aaronson, Daniel, Lisa Barrow and William Sander (2002). "Teachers and student Achievement in the Chicago public high schools." Working Paper Series WP-02-28, Federal Reserve Bank of Chicago

Bettinger, Eric and Bridget Terry Long, "Does It Matter Who Teaches the Course? The Impact of Adjuncts and Graduate Instructors' Case Western working paper.

Bettinger, Eric and Bridget Terry Long, "Help or Hinder? Adjunct Professors and Student Outcomes," mimeo, 2005.

Borjas, George, "Foreign-Born Teaching Assistants and the Academic Performance of Undergraduates," NBER Working Paper No. 7635, 2000.

Card, David, and Alan Krueger, "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," Journal of Political Economy, Vol. 100 (February), 1992, pp. 1-40.

Card, David, and Alan Krueger, "School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina, NBER Working Paper #5708, 1996.

Coleman, J.S., E.Q. Campbell, C.J. Hobson, J. McPartland, A.M. Mood, F.D. Weinfeld, and R.L. York. "Equality of Educational Opportunity," Washington, D.C.: U.S. Government Printing Office, 1966.

Ehrenberg, Ronald G. and Zhang, Liang, "Do Tenure and Tenure-Track Faculty Matter?" Journal of Human Resources, forthcoming, 2005.

Hedges, Larry V., Richard Laine, and Rob Greenwald. "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes," Education Researcher 23, no. 3 (April): 1994, pp. 5-14.

Heckman, J. "Sample Selection Bias as a Specification Error," Econometrica, Vol. 47, 1979, pp. 153-161.

Jacob, Brian A., and Lars Lefgren. "Pricnipals as Agents: Subjective Performance Measurement in Education," Harvard Working Paper, June 2005.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. "What Does Certification Tell Us About Teacher Effectiveness?: Evidence from New York City", NBER working paper (forthcoming), April 2006

Rivkin, Steven G., Erik Hanushek, and John Kain. “Teachers, Schools and Academic Achievement,” Econometrica, Vol. 73, no. 2, 2005.

Rockoff, Jonah. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," American Economic Review, Papers and Proceedings, May 2004.

Table 1
Descriptive Means and Standard Deviations

Student Data, 1995 - 2004

	(1)
age at entry	18.58
	(0.731)
female	0.598
highschool grade	85.72
	(5.586)
GPA year1	2.529
	(0.910)
GPA year 2	2.592
	(0.866)
GPA year 3	2.733
	(0.807)
total credits year 1	4.450
	(0.925)
total credits year 2	3.949
	(1.514)
total credits year 3	3.917
	(1.557)
registered in fall, year 1	0.977
	(0.148)
registered in fall, year 2	0.893
	(0.375)
undergraduate degree: all observations	0.443
undergraduate degree:	0.779
entered program before Fall 2000	
number of students	34,993

1st Year Class Data, 1995 - 2004

	All courses	Large courses
Class Characteristics	**(2)**	**(3)**
number of students in class	273.0	340.0
	(309.3)	(320.7)
number of sections in course x year	2.589	2.908
	(2.077)	(2.166)
class on Monday	0.391	0.388
class on Friday	0.258	0.269
class begins before 10AM	0.084	0.100
class begins after 4PM	0.024	0.022
student evaluation completion rate	0.572	0.565
	(0.167)	(0.173)
Instructor Characteristics		
effective overall (scale is 1-7)	5.568	5.541
	(0.566)	(0.571)
fraction of evaluations with values 1-2 (bad evaluations)	0.055	0.058
	(0.062)	(0.062)
fraction of evaluations with value 7 (good evaluations)	0.259	0.256
	(0.170)	(0.171)
instructor is lecturer (non tenure or non tenure-track)	0.280	0.310
	(0.423)	(0.434)
instructor is assistant or associate professor	0.217	0.180
	(0.359)	(0.320)
instructor is full professor	0.261	0.268
	(0.375)	(0.369)
part time instructor (grad student, emeritus, adjunct, or missing info)	0.242	0.242
	0.390	(0.385)
instructor's real income > Cdn$100,000	0.188	0.202
	(0.332)	(0.332)
number of courses	1029	47
number of classes (course x section x year)	4124	784
number of different instructors	1849	397
Fraction of students in large courses	0.78	

Student x 1st Year Class Data, 1995 - 2004

	Large Courses
	(4)
dropped class	0.091
	(0.288)
grade in class	69.184
	(13.974)
number of same-subject classes year1	1.164
	(0.557)
number of same-subject classes year 2	0.704
	(1.189)
number of same-subject classes year3	0.501
	(1.251)
avg. grade in same-subject classes year2	68.88
	(13.36)
avg. grade in same-subject classes year3	70.41
	(12.60)
number of observations	103780

Note: standard deviations in parentheses.

Table 2
Regressions of Class Dropout and Number of Same-Subject Classes in Year 2 on Year 1 Instructor Quality

	Mean and s.d. of instructor quality variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Dependent variable: Dropped class (mean = 0.091, se=0.288)						
instructor effectiveness	*5.54*	-0.013	-0.016	-0.015				-0.013
	(0.572)	[0.004]**	[0.005]***	[0.005]***				[0.005]**
lecturer	*0.310*				-0.008	-0.008		-0.003
	(0.434)				[0.008]	[0.009]		[0.012]
assistant or associate	*0.180*					0.006		0.007
professor	*(0.320)*					[0.010]		[0.013]
part-time instructor	*(0.242)*					0		0.004
	(0.385)					[0.011]		[0.014]
top salary	*0.20*						0.011	0.011
	(0.332)						[0.010]	[0.012]
		Dependent variable: Number of same-subject courses in year 2 (mean = 0.704 , se=1.187)						
instructor effectiveness	*5.54*	0.041	0.035	0.03				0.022
	(0.572)	[0.025]*	[0.020]*	[0.020]				[0.021]
lecturer	*0.310*				0.024	-0.002		-0.005
	(0.434)				[0.038]	[0.041]		[0.054]
assistant or associate	*0.180*			0.034		0.026		0.004
professor	*(0.320)*			(0.029)		[0.050]		[0.060]
part-time instructor	*(0.242)*					-0.078		-0.077
	(0.385)					[0.045]*		[0.057]
top salary	*0.20*						-0.057	-0.051
	(0.332)						[0.049]	[0.055]
Year Fixed Effects		Yes	No	No	No	No	No	No
Course Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Fixed Effects		No	Yes	Yes	Yes	Yes	Yes	Yes
Time of Day Controls		No	No	Yes	Yes	Yes	Yes	Yes
Number of observations		103780	103780	103780	103780	103780	103780	103780

Notes: The rank coefficients, lecturer, assistant or associate professor, and other are relative to the omitted rank variable, full professor. The top salary variable indicates an instructor earns more than $100,000 in 2005 Canadian dollars. Standard Errors in parantheses. *** significant on 1%-level; ** significant on 5%-level; * significant on 10%-level. Sample includes students initially enrolled in courses with average class sizes greater than 200 between 1995 and 2004.

Table 3
Regressions of Student-Class Outcomes on Instructor Quality

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Mean and s.d. of instructor quality variable	Dropped Class	Grade	Same-subject Courses, year2	Same-subject Credits, year2	Same-subject Avg, year2	Same-subject Courses, year3	Same-subject Credits, year3	Same-subject Avg, year3
Mean and S.d. of		*0.091*	*69.18*	*0.704*	*0.352*	*68.881*	*0.501*	*68.88*	*70.41*
dependent variables		*(0.288)*	*(13.974)*	*(1.189)*	*(0.595)*	*(13.356)*	*(0.629)*	*(13.356)*	*(12.605)*
instructor effectiveness	*5.54*	-0.013	1.078	0.022	0.011	0.25	0.029	0.014	0.093
	(0.572)	[0.005]**	[0.223]***	[0.021]	[0.011]	[0.223]	[0.019]	[0.009]	[0.271]
lecturer	*0.310*	-0.003	-1.171	-0.005	-0.002	0.046	-0.035	-0.012	-0.009
	(0.434)	[0.012]	[0.515]**	[0.054]	[0.027]	[0.508]	[0.041]	[0.020]	[0.489]
assistant or associate	*0.180*	0.004	-1.1	-0.077	-0.038	1.197	-0.039	-0.011	-0.058
professor	*(0.320)*	[0.014]	[0.660]*	[0.057]	[0.028]	[0.606]**	[0.051]	[0.025]	[0.521]
part-time instructor	*(0.242)*	0.011	-0.415	-0.051	-0.025	0.046	-0.023	-0.007	0.106
	(0.385)	[0.012]	[0.476]	[0.055]	[0.027]	[0.448]	[0.039]	[0.019]	[0.473]
top salary	*0.20*	0.007	-0.097	0.004	0.002	0.473	-0.011	-0.002	-0.235
	(0.332)	[0.013]	[0.471]	[0.060]	[0.030]	[0.509]	[0.048]	[0.023]	[0.436]
Number of observations		103780	94301	103780	103780	38063	103780	38063	21258

Notes: Each column reports estimates from regressing the student outcome variable on measures of instructor quality plus course and student fixed effects, time of day and week controls The rank coefficients, lecturer, assistant or associate professor, and other are relative to the omitted rank variable, full professor. The top salary variable indicates an instructor earns more than $100,000 in 2005 Canadian dollars. Standard Errors in parantheses. *** significant on 1%-level; ** significant on 5%-level; * significant on 10%-level. Sample includes students initially enrolled in courses with average class sizes greater than 200 between 1995 and 2004.

Table 4
Regressions of Student-Class Outcomes on Subjective Quality

	Mean and s.d. of instructor quality variable	(1) Dropped Class	(2) Grade	(3) Same-subject Courses, year2	(4) Same-subject Credits, year2	(5) Same-subject Avg, year2
Mean and S.d. of dependent variables		*0.091*	*69.18*	*0.704*	*0.352*	*68.881*
		(0.288)	*(13.974)*	*(1.189)*	*(0.595)*	*(13.356)*
Average insturctor's perceived effectiveness	*5.54*	-0.015	0.939	0.03	0.015	0.217
	(0.572)	[0.005]***	[0.209]***	[0.020]	[0.010]	[0.197]
Fraction that give instructor highest rating	*0.256*	-0.042	3.242	0.084	0.042	0.912
	(0.171)	[0.017]**	[0.558]***	[0.067]	[0.033]	[0.669]
Fraction that give instructor lowest rating	*0.057*	0.141	-6.465	-0.211	-0.105	-0.655
	(0.061)	[0.046]***	[2.354]***	[0.163]	[0.081]	[1.909]
(mean) hotness	*4.431*	0	-0.018	0	0	-0.012
	(11.556)	[0.000]	[0.016]	[0.001]	[0.001]	[0.010]
(mean) avg_clarity	*3.348*	-0.011	0.373	0.025	0.012	-0.188
	(0.887)	[0.004]***	[0.215]*	[0.016]	[0.008]	[0.178]
(mean) avg_easiness	*3.014*	-0.004	0.229	-0.013	-0.006	-0.702
	(0.612)	[0.006]	[0.452]	[0.021]	[0.011]	[0.349]**
Provides helpful comments and feedback	*4.909*	-0.025	1.645	0.035	0.017	0.17
	(0.463)	[0.009]***	[0.352]***	[0.031]	[0.016]	[0.292]
Available to meet	*5.231*	-0.013	1.06	-0.002	-0.001	0.312
	(0.488)	[0.006]**	[0.352]***	[0.027]	[0.014]	[0.286]
Answers questions clearly and effectively	*5.316*	-0.018	0.974	0.032	0.016	0.106
	(0.495)	[0.006]***	[0.293]***	[0.023]	[0.011]	[0.237]
Communicates enthusiasm and interest	*5.525*	-0.014	0.657	0.041	0.02	0.17
	(0.639)	[0.005]***	[0.205]***	[0.021]**	[0.010]**	[0.201]
Explains concepts clearly	*5.393*	-0.015	0.911	0.032	0.016	0.096
	(0.563)	[0.005]***	[0.194]***	[0.021]	[0.011]	[0.221]
Presents material in organized manner	*5.404*	-0.013	0.848	0.017	0.008	0.079
	(0.557)	[0.005]**	[0.183]***	[0.024]	[0.012]	[0.178]
Provides fair evaluation of student learning	*4.980*	-0.024	2.21	0.064	0.032	0.087
	(0.421)	[0.009]***	[0.357]***	[0.037]*	[0.018]*	[0.316]
Would take course again given experience	*65.637*	-0.001	0.08	0.003	0.002	-0.016
	(15.175)	[0.000]***	[0.015]***	[0.002]*	[0.001]*	[0.015]
Number of observations		103780	94301	103780	103780	38063

Notes: Each value is from a separate regression from regressing the student outcome variable on the subjective instructor quality measure plus course and student fixed effects, time of day and week controls With the exception of the variables, 'fraction that give instructors highest or lowest rating' and 'would take course again given experience', all quality measures are on a 7 point scale, with 1 meaning strongly disagree and 7 meaning strongly agree. Standard Errors in parantheses. *** significant on 1%-level; ** significant on 5%-level; * significant on 10%-level. Sample includes students initially enrolled in courses with average class sizes greater than 200 between 1995 and 2004.

Table 5
Regressions of Class Dropout on Instructor Quality Measures, by Student Characteristic

	Mean and s.d. of instructor quality variable	(1) Male	(2) Female	(3) Mother Tongue English	(4) Mother Tongue Non-English	(3) Lowest HS Quartile	(4) Highest HS Quartile	(5) Science Majors	(6) Arts and Social Science Majors
Mean and S.d. of dependent variables		*0.096*	*0.088*	*0.125*	*0.062*	*0.125*	*0.062*	*0.082*	*0.108*
		(0.295)	*(0.283)*	*(0.331)*	*(0.241)*	*(0.331)*	*(0.241)*	*(0.274)*	*(0.310)*
instructor effectiveness	*5.54*	-0.013	-0.012	-0.014	-0.01	-0.049	0.002	-0.01	-0.018
	(0.572)	[0.006]**	[0.006]**	[0.006]**	[0.006]*	[0.012]***	[0.004]	[0.005]*	[0.007]**
lecturer	*0.310*	-0.005	-0.003	-0.005	-0.001	0.003	-0.009	0	0
	(0.434)	[0.013]	[0.013]	[0.012]	[0.014]	[0.017]	[0.010]	[0.014]	[0.014]
assistant or associate	*0.180*	0.008	0	0.001	0.007	0.023	-0.011	0.003	0.023
professor	*(0.320)*	[0.014]	[0.016]	[0.014]	[0.017]	[0.021]	[0.011]	[0.018]	[0.014]
part-time instructor	*(0.242)*	0.013	0.009	0.011	0.012	0.026	-0.001	0.01	0.016
	(0.385)	[0.012]	[0.013]	[0.012]	[0.013]	[0.014]*	[0.010]	[0.014]	[0.012]
top salary	*0.20*	0.015	0	0.006	0.01	0.011	-0.009	-0.02	0.02
	(0.332)	[0.013]	[0.014]	[0.013]	[0.015]	[0.015]	[0.009]	[0.015]	[0.014]
Number of observations		42198	61582	65820	37960	25945	25945	41768	62012

Notes: Each column reports estimates from regressing the student outcome variable on measures of instructor quality plus course and student fixed effects, time of day and week controls The rank coefficients, lecturer, assistant or associate professor, and other are relative to the omitted rank variable, full professor. The top salary variable indicates an instructor earns more than $100,000 in 2005 Canadian dollars. Standard Errors in parantheses. *** significant on 1%-level; ** significant on 5%-level; * significant on 10%-level. Sample includes students initially enrolled in courses with average class sizes greater than 200 between 1995 and 2004.

Table 6

Regressions of Same-Subject Courses on Instructor Quality Measures, by Student Characteristic

	Mean and s.d. of instructor quality variable	(1) Male	(2) Female	(3) Mother Tongue English	(4) Mother Tongue Non-English	(3) Lowest HS Quartile	(4) Highest HS Quartile	(5) Science Majors	(6) Arts and Social Science Majors
Mean and S.d. of dependent variables		*0.777*	*0.654*	*0.704*	*0.704*	*0.630*	*0.729*	*0.646*	*0.743*
		(1.249)	*(1.143)*	*(1.193)*	*(1.182)*	*(1.161)*	*(1.178)*	*(0.274)*	*(1.222)*
instructor effectiveness	*5.54*	0.019	0.024	0.034	0.001	0.067	-0.008	-0.028	0.032
	(0.572)	[0.022]	[0.023]	[0.022]	[0.024]	[0.033]**	[0.021]	[0.023]	[0.030]
lecturer	*0.310*	-0.023	0.014	-0.022	0.034	-0.134	0.099	0.104	-0.002
	(0.434)	[0.052]	[0.059]	[0.049]	[0.068]	[0.061]**	[0.052]*	[0.085]	[0.066]
assistant or associate professor	*0.180*	-0.128	-0.037	-0.075	-0.068	-0.16	-0.025	0.093	-0.116
	(0.320)	[0.053]**	[0.067]	[0.052]	[0.076]	[0.087]*	[0.058]	[0.094]	[0.079]
part-time instructor	*(0.242)*	-0.098	-0.01	-0.07	-0.016	-0.093	0	0.138	-0.086
	(0.385)	[0.051]*	[0.061]	[0.052]	[0.067]	[0.056]*	[0.052]	[0.110]	[0.052]
top salary	*0.20*	-0.035	0.039	0.011	-0.012	-0.103	0.136	0.11	0.012
	(0.332)	[0.056]	[0.068]	[0.057]	[0.072]	[0.070]	[0.048]***	[0.078]	[0.072]
Number of observations		42198	61582	65820	37960	25945	25945	41768	62012

Notes: Each column reports estimates from regressing the student outcome variable on measures of instructor quality plus course and student fixed effects, time of day and week controls The rank coefficients, lecturer, assistant or associate professor, and other are relative to the omitted rank variable, full professor. The top salary variable indicates an instructor earns more than $100,000 in 2005 Canadian dollars. Standard Errors in parantheses. *** significant on 1%-level; ** significant on 5%-level; * significant on 10%-level. Sample includes students initially enrolled in courses with average class sizes greater than 200 between 1995 and 2004.

Table 7
Estimates of Standard Deviations of Instructor Fixed Effects

	Square root of covariance between students in same section		***Square root of covariance between students with same instructor, different section***	
	(1)	**(2)**	**(3)**	**(4)**
grade	0.1379	0.0937	0.0675	0.0450
(with mean=0, s.d.=1)	(0.0036)	(0.0018)	(0.0074)	(0.0025)
dropped course	0.0404	0.0256	0.0137	0.0097
(with mean=0.09, s.d.=0.2	(0.0014)	(0.0009)	(0.0029)	(0.0021)
subject courses, year 2	0.1408	0.0970	0.0787	0.0296
(with mean=0, s.d.=1)	(0.0051)	(0.0030)	(0.0091)	(0.0073)
subject courses, year 3	0.1687	0.0876	0.0615	0.0000
(with mean=0, s.d.=1)	(0.0042)	(0.0038)	(0.0095)	(0.0003)
Course Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	No	Yes	No
Individual Fixed Effects	No	Yes	No	Yes

Notes: The student-course outcome variable is regressed on course fixed effects and individual or year fixed effects where indicated. The residuals are used to estimate within section and across section covariances, as detailed more in the text. Bootstrapped standard Errors in parantheses. *** significant on 1%-level; ** significant on 5%-level; * significant on 10%-level. Sample includes students initially enrolled in courses with average class sizes greater than 200 between 1995 and 2004.

Summary statistics of big courses, Part 1

	number of obs	section size	Avg. number of sections per year	grade	response rate	Effective overall	Would take course again
Introduction to Antropology	377	188.6 (4.505)	1.000 (0.000)	68.18 (13.00)	0.618 (0.000)	5.419 (0.082)	77.07 (0.147)
Introduction to Architecture	939	142.1 (32.77)	1.000 (0.000)	71.22 (11.31)	0.600 (0.012)	5.081 (0.181)	77.71 (5.547)
Contemporary Architecture	766	134.8 (30.62)	1.000 (0.000)	73.45 (12.99)	0.665 (0.039)	4.939 (0.102)	73.17 (4.752)
Introduction to Astronomy	1934	152.0 (61.84)	1.689 (0.463)	69.04 (13.34)	0.359 (0.061)	5.318 (0.215)	72.37 (4.980)
Introduction to Biology	9342	1057.3 (143.4)	1.000 (0.000)	70.44 (11.84)	0.765 (0.030)	5.407 (0.078)	68.16 (1.141)
Introduction to General Chemistry	3515	220.1 (9.874)	4.000 (0.000)	70.03 (13.66)	0.519 (0.198)	5.737 (0.431)	65.57 (2.909)
Introduction to Organic Chemistry	5471	563.8 (91.47)	2.000 (0.000)	72.20 (14.08)	0.577 (0.013)	5.727 (0.231)	68.64 (3.712)
Introduction to Physical Chemistry	5321	547.4 (85.80)	2.000 (0.000)	71.11 (13.19)	0.460 (0.053)	5.211 (0.110)	52.37 (3.894)
Introduction to Computers	561	53.34 (9.637)	1.390 (0.488)	69.79 (12.77)	0.480 (0.102)	5.197 (0.654)	71.56 (7.988)
Introduction to Computer Programming	2743	91.82 (24.61)	3.895 (1.261)	70.36 (17.22)	0.562 (0.102)	5.780 (0.479)	75.43 (4.645)
Introduction to Computer Science	2169	81.11 (17.60)	3.610 (1.223)	67.10 (18.47)	0.538 (0.073)	5.700 (0.522)	71.54 (7.014)
Mathematics for Computer Science	403	89.49 (39.73)	2.184 (0.920)	64.08 (19.95)	0.509 (0.029)	5.474 (0.628)	46.78 (7.675)
Introduction to Economics	6244	266.4 (39.97)	3.000 (0.000)	67.12 (14.64)	0.520 (0.095)	5.477 (0.454)	72.73 (8.245)
Introduction to Economics for Non-Specialists	122	122.0 (0.000)	1.000 (0.000)	70.07 (12.99)	0.544 (0.000)	5.800 (0.000)	70.00 (0.000)
Effective Writing	1616	30.85 (7.903)	6.283 (1.233)	73.22 (9.485)	0.694 (0.092)	5.913 (0.356)	84.45 (8.344)
Narrative	1404	52.86 (14.13)	4.088 (1.081)	70.75 (10.82)	0.641 (0.100)	5.999 (0.388)	83.80 (7.982)
Modern Literature	1375	120.2 (75.40)	2.303 (1.417)	70.05 (12.54)	0.538 (0.071)	5.793 (0.582)	81.95 (11.832)
Introduction to History of Art	646	85.90 (22.22)	1.000 (0.000)	72.78 (10.95)	0.520 (0.013)	5.783 (0.078)	89.22 (0.538)
Introduction to Geography	788	105.5 (22.46)	1.000 (0.000)	69.81 (10.36)	0.582 (0.039)	5.645 (0.468)	75.35 (3.964)
Introduction to History	1869	250.0 (60.75)	1.000 (0.000)	69.07 (12.94)	0.498 (0.062)	6.049 (0.291)	78.27 (2.806)
European History before 1945	1784	215.7 (59.57)	1.000 (0.000)	68.44 (12.53)	0.632 (0.023)	6.041 (0.412)	83.81 (2.889)
Civilization and Culture of Asia	1140	156.7 (37.64)	1.000 (0.000)	70.95 (10.20)	0.660 (0.032)	4.901 (0.153)	61.34 (1.776)
1st year siminar in Humanities	525	15.63 (3.216)	4.686 (1.374)	76.80 (10.52)	0.835 (0.093)	6.250 (0.455)	89.81 (9.812)
Introduction to Film Study	1088	86.68 (15.50)	1.787 (0.410)	68.48 (12.91)	0.640 (0.081)	5.727 (0.559)	85.30 (5.975)
Total	103780	332.5 (319.9)	2.908 (2.166)	69.18 (13.97)	0.565 (0.173)	5.541 (0.571)	65.63 (15.21)

Note: Standard deviations in parantheses

Table A1: Summary statistics of big courses, Part 2

	number of obs	section size	number of sections	grade	response rate	Effective overall	Would take course again
Western Tradition	935	41.49	3.893	70.49	0.589	6.047	85.83
		(9.117)	(0.916)	(12.07)	(0.116)	(0.502)	(8.302)
Calculus and Linear Algebra for Commerce	4585	101.4	5.150	64.05	0.437	5.340	51.19
		(13.59)	(0.579)	(16.56)	(0.145)	(0.609)	(6.909)
Calculus fo Science I	8932	132.9	7.592	69.23	0.542	5.650	57.63
		(13.80)	(0.657)	(14.11)	(0.273)	(0.834)	(11.03)
Introduction to Rigorous Calculus	3117	74.00	5.901	63.40	0.469	5.580	46.38
		(22.43)	(1.697)	(19.46)	(0.169)	(0.641)	(10.80)
Linear Algebra I	920	51.00	2.605	62.04	0.389	4.976	42.06
		(20.37)	(0.898)	(19.46)	(0.111)	(0.622)	(9.315)
Financial Accounting I	4176	287.0	2.156	68.80	0.576	4.892	58.49
		(150.8)	(0.947)	(14.36)	(0.007)	(0.352)	(5.323)
Introduction to Management	157	157.0	1.000	69.42	0.473	5.200	64.50
		(0.000)	(0.000)	(12.41)	(0.000)	(0.000)	(0.000)
Management Accounting I	1423	163.5	2.330	65.16	0.451	6.064	54.63
		(27.93)	(0.470)	(14.03)	(0.004)	(0.778)	(4.083)
Introduction to Philosophy	2601	172.3	2.309	69.12	0.517	5.619	71.29
		(72.17)	(0.566)	(12.65)	(0.057)	(0.249)	(6.100)
History of Western Philosophy	1058	141.7	1.365	69.23	0.528	5.948	83.15
		(47.36)	(0.482)	(13.84)	(0.072)	(0.376)	(2.723)
Basic Physics	991	113.3	1.000	70.86	0.306	4.606	35.00
		(19.08)	(0.000)	(12.49)	(0.000)	(0.000)	(0.000)
Physics for Life Science I	6614	347.3	3.817	71.94	0.781	4.967	35.41
		(334.3)	(1.740)	(11.20)	(0.255)	(0.266)	(4.087)
Introduction to Canadian Politics	929	134.8	1.000	68.57	0.545	5.689	68.22
		(52.60)	(0.000)	(10.56)	(0.024)	(0.066)	(0.779)
Canada in Comparative Perspective	1037	135.3	1.000	67.80	0.477	5.690	67.09
		(28.83)	(0.000)	(13.03)	(0.064)	(0.094)	(3.590)
Global Networks	1784	350.0	1.000	68.65	0.268	6.367	89.222
		(111.3)	(0.000)	(12.25)	(0.000)	(0.000)	(0.000)
Cultures of Conflict	264	264.0	1.000	67.89	0.662	6.400	85.00
		(0.000)	(0.000)	(13.42)	(0.000)	(0.000)	(0.000)
Democracy and Dictatorship	252	252.0	1.000	66.92	0.528	5.700	76.00
		(0.000)	(0.000)	(14.44)	(0.000)	(0.000)	(0.000)
Introduction to Psychology	6248	748.2	1.000	65.33	0.607	5.891	76.39
		(189.7)	(0.000)	(13.68)	(0.034)	(0.424)	(3.594)
Introduction to Sociology	2192	462.9	1.000	67.71	0.540	5.291	70.35
		(88.92)	(0.000)	(10.28)	(0.091)	(0.247)	(2.056)
Spanish for Beginners	797	23.64	4.754	64.43	0.573	5.610	80.94
		(4.152)	(1.216)	(16.40)	(0.182)	(0.512)	(6.829)
1st year seminar on Social Science	162	16.35	1.790	77.68	0.766	5.998	83.45
		(1.612)	(0.859)	(13.07)	(0.067)	(0.405)	(9.759)
Introduction to Probability	1400	106.4	2.086	70.24	0.593	5.721	61.63
		(29.44)	(0.453)	(17.12)	(0.076)	(0.618)	(8.896)
Introduction to Semiotics and Communication	1064	199.3	1.000	78.42	0.310	6.220	95.48
		(100.8)	(0.000)	(10.45)	(0.048)	(0.495)	(1.588)
Total	103780	332.5	2.908	69.18	0.565	5.541	65.63
		(319.9)	(2.166)	(13.97)	(0.173)	(0.571)	(15.21)

Note: Standard deviations in parantheses

Table A2
Teacher quality regressions for different quality measures and outcome variables, all first year students and courses

	Mean and s.d. of teacher quality variable	(1) Dropped course	(2) Grade	(3) Subject Credits, year2	(4) Subject Credits, year3	(5) Subject Avg, year2	(6) Subject Avg, year3
Mean and S.d. of dependent variables		*0.109*	*69.34*	*0.361*	*0.262*	*69.44*	*71.06*
		(0.312)	*(13.944)*	*(0.609)*	*(0.631)*	*(13.229)*	*(12.639)*
effective overall (Q 11)	*5.57*	-0.014	0.904	0.015	0.015	0.010	0.066
	(0.566)	[0.006]**	[0.205]***	[0.013]	[0.011]	[0.150]	[0.147]
bad evaluations	*0.055*	0.104	-5.943	-0.089	-0.091	0.238	0.569
	(0.062)	[0.042]**	[1.880]***	[0.089]	[0.086]	[1.359]	[1.492]
good evaluations	*0.259*	-0.047	2.953	0.049	0.042	0.092	0.346
	(0.170)	[0.021]**	[0.622]***	[0.042]	[0.035]	[0.511]	[0.551]
take course again	*67.89*	-0.001	0.074	0.001	0.001	-0.02	-0.005
	(15.445)	[0.000]***	[0.013]***	[0.001]*	[0.001]**	[0.011]*	[0.013]
lecturer	*0.458*	-0.005	-0.514	-0.004	-0.008	-0.200	-0.056
	(0.458)	[0.006]	[0.338]	[0.012]	[0.013]	[0.224]	[0.287]
tenured	*0.396*	0.003	0.604	0.002	0.011	0.470	0.314
	(0.439)	[0.007]	[0.384]	[0.017]	[0.014]	[0.242]*	[0.276]
top salary	*0.188*	-0.007	0.455	0.016	0.009	0.182	-0.405
	(0.332)	[0.010]	[0.449]	[0.016]	[0.015]	[0.274]	[0.214]*
Number of observations		133746	119131	133746	133746	48910	28685

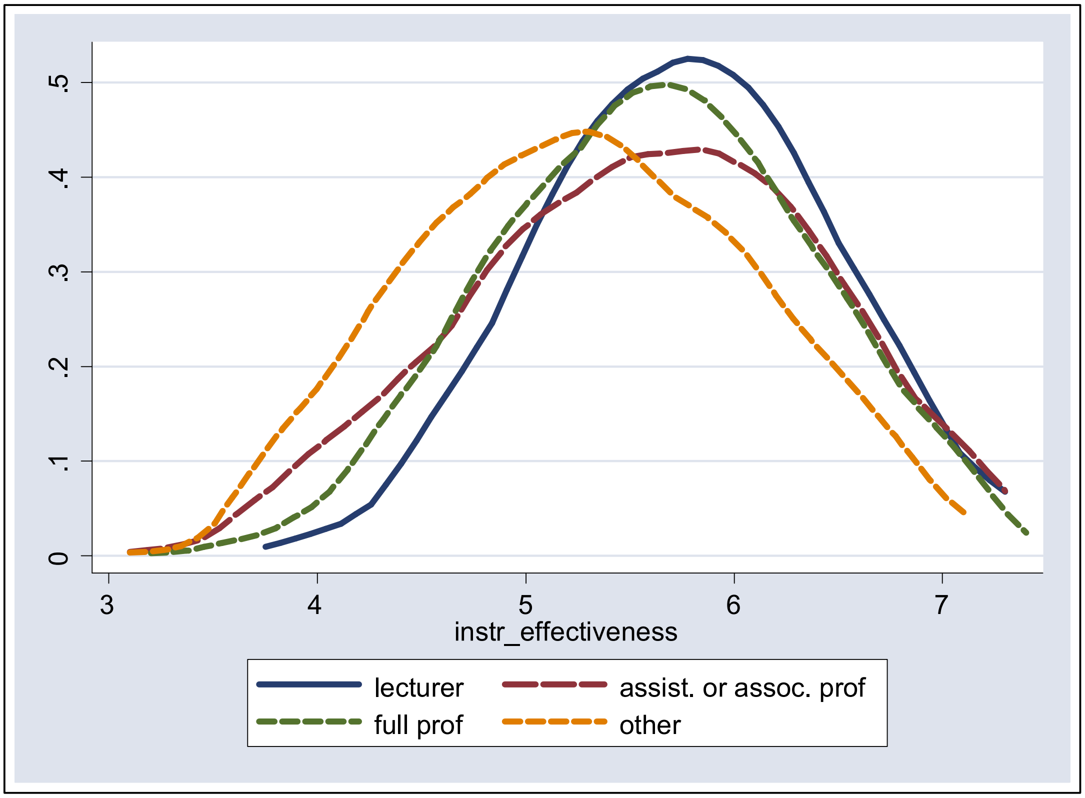
Notes: Each column reports estimates from regressing the student outcome variable on measures of instructor quality plus course and student fixed effects, time of day and week controls The rank coefficients, lecturer, assistant or associate professor, and other are relative to the omitted rank variable, full professor. The top salary variable indicates an instructor earns more than $100,000 in 2005 Canadian dollars. Standard Errors in parantheses. ***

Table A3
Regressions of Student-Class Outcomes on Instructor Quality Measures, Second Year Students and Courses

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Mean and s.d. of instructor quality variable	Dropped Class	Grade	Same-subject Courses, year2	Same-subject Credits, year2	Same-subject Avg, year2	Same-subject Courses, year3	Same-subject Credits, year3	Same-subject Avg, year3
Mean and S.d. of dependent variables		*0.131*	*69.87*	*1.964*	*0.980*	*69.679*	*1.308*	*0.66*	*71.61*
		(0.337)	*(13.233)*	*(1.433)*	*(0.714)*	*(12.664)*	*(1.838)*	*(0.921)*	*(11.246)*
instructor effectiveness	*5.35*	-0.021	1.144	-0.011	-0.005	0.545	-0.023	-0.011	0.038
	(0.619)	[0.005]***	[0.202]***	[0.012]	[0.006]	[0.127]***	[0.021]	[0.011]	[0.104]
lecturer	*0.253*	-0.016	-0.477	-0.007	-0.003	-0.192	0.04	0.02	-0.31
	(0.407)	[0.012]	[0.385]	[0.028]	[0.014]	[0.268]	[0.041]	[0.021]	[0.237]
assistant or associate professor	*0.219*	-0.008	0.199	0.026	0.013	-0.096	0.026	0.011	-0.187
	(0.379)	[0.012]	[0.459]	[0.025]	[0.013]	[0.332]	[0.039]	[0.020]	[0.227]
Emeritus, missing or other instructor	*(0.300)*	-0.017	-0.254	0.028	0.014	-0.248	0.012	0.005	-0.087
	(0.405)	[0.010]	[0.367]	[0.024]	[0.012]	[0.243]	[0.033]	[0.016]	[0.188]
top salary	*0.15*	-0.021	-0.392	0.034	0.016	-0.189	0.009	0.006	0.037
	(0.331)	[0.012]*	[0.389]	[0.027]	[0.014]	[0.246]	[0.038]	[0.019]	[0.196]
Number of observations		73817	63830	73817	73817	67563	73817	73817	32490

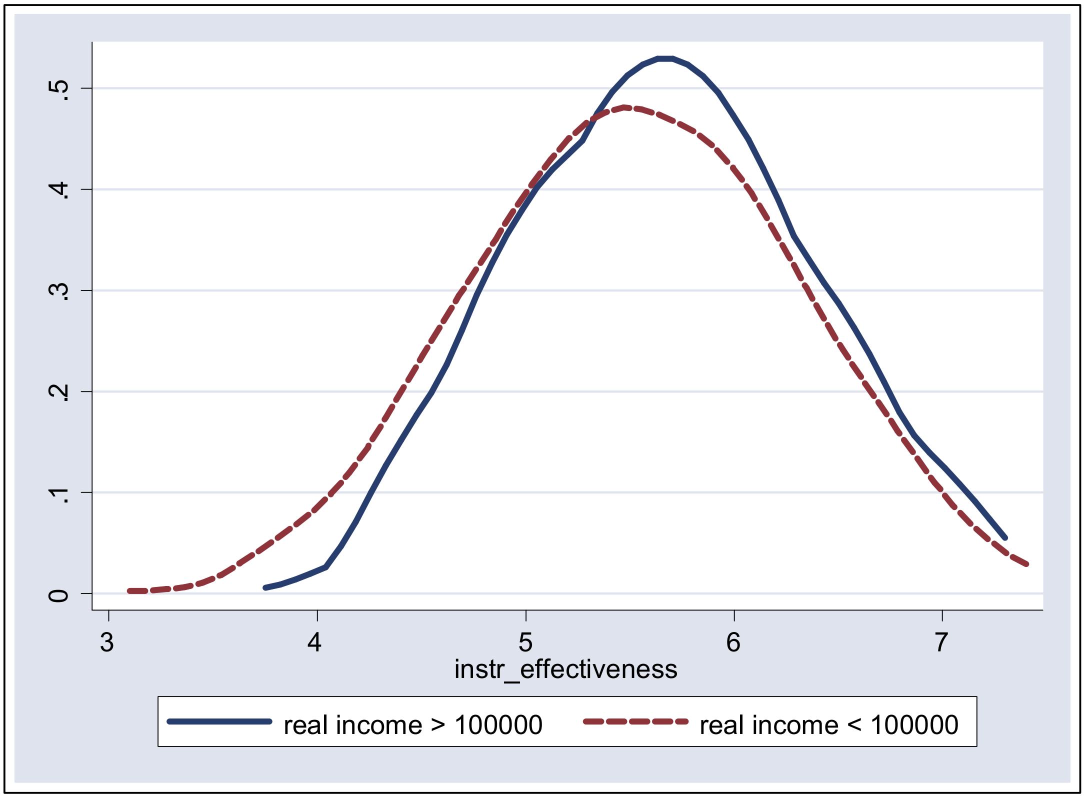
Notes: Each column reports estimates from regressing the student outcome variable on measures of instructor quality plus course and student fixed effects, time of day and week controls The rank coefficients, lecturer, assistant or associate professor, and other are relative to the omitted rank variable, full professor. The top salary variable indicates an instructor earns more than $100,000 in 2005 Canadian dollars. Standard Errors in parantheses. *** significant on 1%-level; ** significant on 5%-level; * significant on 10%-level. Sample includes all second year students in courses with average enrollment more than 50 students between 1995 and 2004.

Figure 1A
Kernel Densities of Perceived Instructor Effectiveness by Instructor Rank



Notes: Perceived effectiveness is evaluated on a 7 point scale, and averaged across all first year students who took the same instructor. The figure shows the distribution of mean perceived instructor effectiveness for the set of instructors who are lecturers (paid primarily to teach), assistant or associate professors, full professors, or other (mostly part-time sessional instructors).

Figure 1B
Kernel Densities of Perceived Instructor Effectiveness by Top Salary Status



Notes: Perceived effectiveness is evaluated on a 7 point scale, and averaged across all first year students who took the same instructor. The figure shows the distribution of mean perceived instructor effectiveness for the set of instructors who are paid more than $100,000 in 2005 Canadian dollars and those who are not.