PROFESSOR QUALITIES AND STUDENT ACHIEVEMENT

Florian Hoffmann and Philip Oreopoulos*

Abstract—This paper analyzes the importance of teacher quality at the college level. Instructors are matched to objective and subjective characteristics of teacher quality to estimate the impact of rank, salary, and perceived effectiveness on student performance and subject interest. Student and course fixed effects, time of day and week controls, and students' lack of knowledge about first-year instructors help minimize selection biases. Subjective teacher evaluations perform well in measuring instructor influences on students, while objective characteristics such as rank and salary do not. Overall, the importance of college instructor differences is small, but important outliers exist.

I. Introduction

UNIVERSITIES and colleges emphasize teaching as the most important determinant to a student's academic experience and successful transition into the labor force. Yet administrators often have difficulty identifying and cultivating specific characteristics related to teacher effectiveness. Most colleges rely on summary statistics of student evaluations to assess teacher quality. Faculty opinion on the reliability of these measures ranges from "reliable, valid, and useful" to "unreliable, invalid, and useless" (Wachtel, 1998). Administrators also wonder whether part-time or nontenured faculty that focus exclusively on teaching are as effective as or more effective than tenured faculty in fostering student performance, and whether teaching effectiveness improves with experience. This paper contributes to this literature by applying recent advances in the area of teacher quality research to the college level. Administrative data of instructors are matched to objective and subjective characteristics of teacher quality to analyze 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 its measurement 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 *does* 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, Laine, & Greenwald, 1994). Studies that examine the relationship between teacher quality and longer-run outcomes, such as earnings, find more consistent evidence that teacher quality matters (for example, Card & 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 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. Consistent with this hypothesis, Jacob and Lefgren (2005) find that 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 postsecondary level is virtually nonexistent. 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 Zhang (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 that this could be driven by the tendency for schools with a higher proportion of adjuncts to also be more likely to have students on the margin of dropping out. The most closely related research to this paper is by Bettinger and Long (2004, 2005), who use an administrative data set of public four-year universities in Ohio to estimate the effects of being taught by an adjunct professor on course selection and completion. Using year-to-year and class-to-class variation in first-year instructors they conclude 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 drop out 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

Received for publication October 6, 2006. Revision accepted for publication August 21, 2007.

* Department of Economics, University of Toronto.

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, to conference participants at the Society of Labor Economists Annual Meetings (Boston), the Higher Education Group of the National Bureau of Economic Research, and the Conference on Education, Training and the Evolving Workplace (Vancouver), and to two anonymous referees on an earlier draft. Oreopoulos thanks the Social Sciences and Research Council for financial support.

The Review of Economics and Statistics, February 2009, 91(1): 83–92

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, and 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 setup 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 the extent to which instructor differences matter at all and whether reasons for these differences are observed or unobserved. Similar to previous studies at the primary and secondary school level, we do this by estimating the variance in instructor fixed effects on academic achievement.

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 virtually 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 better in reflecting an instructor's influence on students compared with 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 1 standard deviation increase in an instructor's perceived effectiveness increases standardized test scores by about 5% of its standard deviation (compared with a course dropout rate of 9%). The same increase in perceived effectiveness is also associated with a 1.3 percentage point increase in the course completion rate and a small increase in the number of same-subject courses taken in later years. The effects are similar among males and females, science and nonscience 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 sessions of academic years starting between 1996 and 2004. We focus on the 32,666 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.

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 enroll are female, and high school grade averages are tightly distributed around the mean of about 85%. Annual fall registration status shows that about 10% of 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%. One-third of all students report a mother tongue other than English or French, and 10% are of Asian citizenship. Program at entry is almost evenly split between science, social science, and undeclared.

Course selection is concentrated among large first-year introductory classes. The 47 largest courses, with average annual enrollment sizes of 200 or more, make up 78% of a student's curriculum, on average. We focus our main analysis on these core courses to minimize variation in student background characteristics across classes, to help ensure that the main results are not driven by particularly small or upper-year courses, and to reduce computational time of the estimates. Since we include course and year fixed effects in our analysis (or, in an alternative specification, course by year fixed effects), these multisection courses are our major source of identification, even when using a sample with all courses. Column 2 of table 1 lists these courses and their characteristics. The main results remain virtually unchanged when including all classes.

Instructor evaluations are taken near the end of a semester in each class. The form is anonymous and identical across all arts and science courses. The question that this university uses most often for tenure decisions and teaching reviews is the perceived effectiveness question: "All things considered, this instructor performs effectively as a university teacher." The effectiveness question is on a seven-point scale, ranging from 1 (extremely poor) to 7 (outstanding). We use the mean evaluation across all classes for each instructor.[1] Students linked to classes associated with dif-

[1] Taking the mean of the means for instructor evaluations across classes ensures that instructor quality measures differ only when instructors differ. Transitory measures of instructor quality likely include measurement error and classroom idiosyncrasies. A constant measure of instructor quality over the nine-year period is more reliable and interpretable. Using instructor quality averaged over previous classes does not alter the main results, but estimates are less precise.

TABLE 1.—DESCRIPTIVE MEANS AND STANDARD DEVIATIONS, FIRST-YEAR STUDENTS

Student Data 1996–2004	(1)
Age at entry	18.54
	(0.719)
Female	0.593
High school grade	85.85
	(5.673)
GPA year 1	2.529
	(0.914)
GPA year 2	2.591
	(0.863)
GPA year 3	2.728
	(0.805)
Total credits year 1	4.464
	(0.916)
Total credits year 2	3.980
	(1.476)
Total credits year 3	3.938
	(1.539)
Registered in fall, year 1	0.978
	(0.145)
Registered in fall, year 2	0.903
	(0.296)
Registered in fall, year 3	0.837
	(0.370)
Undergraduate degree: all observations	0.443
Undergraduate degree: (entered program before fall 2000)	0.787
Number of students	32,666

First-Year Class Data, 1996–2004	All Courses (2)	Large Courses (3)
Class characteristics		
Number of students in class	257.9	332.5
	(314.0)	(319.9)
Number of sections in course $\times$ year	2.589	2.908
	(2.077)	(2.166)
Class on Monday	0.393	0.390
Class on Friday	0.264	0.276
Class begins before 10 a.m.	0.084	0.100
Class begins after 4 p.m.	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	5.540
	(0.566)	(0.570)
Fraction of evaluations with values 1–2 (bad evaluations)	0.054	0.057
	(0.061)	(0.061)
Fraction of evaluations with value 7 (good evaluations)	0.259	0.256
	(0.170)	(0.171)
Instructor is lecturer (nontenure or nontenure-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	0.199
	(0.329)	(0.329)
Number of courses	1,029	47
Number of classes (course $\times$ section $\times$ year)	4,108	784
Number of different instructors	1,844	389
Fraction of students in large courses	0.78	

Student $\times$ First-Year Class Data, 1996–2004	Large Courses (4)
Dropped course	0.091
	(0.288)
Dropped course in first month	0.055
	(0.227)
Dropped course late	0.047
	(0.211)
Grade	69.184
	(13.974)
Additional number of same-subject courses	1.563
	(3.062)
Additional number of same-subject credits	0.785
	(1.534)
Number of observations	103,780

Note: Standard deviations in parentheses.

ferent instructor quality measures must have different instructors. Across 1,844 first-year instructors, the mean of the classroom mean for effective overall is 5.6, with a standard deviation of 0.60. The 25th percentile instructors and the 75th percentile instructors differ by almost exactly 2 standard deviations (5.2 versus 6.2). We also consider alternative measures of instructor quality, including the fraction of students giving an instructor very poor ratings (1 or 2) or very high ratings (7), the fraction of enrolled students that attended class on the day the evaluations were taken, and the fraction of students who agree with the statement, "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?"

We use historical university course calendars to match an instructor's name to his or her corresponding rank. We use an instructor's most frequent position over the nine-year period to create an indicator variable for (i) whether an instructor is a lecturer, employed full-time primarily to teach (31%), (ii) whether an instructor is an assistant or associate professor, employed full-time and expected to do research (18%), (iii) whether an instructor is a full professor, with tenure (27%), or (iv) whether an instructor falls into another category (24%). 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 CDN$100,000. We use these data to create a variable (called "top salary") to indicate what years an instructor earned CDN$100,000 or more, calculated in 2006 real Canadian dollars using Statistics Canada's consumer price index.[2]

Column 4 of table 1 shows outcome data categorized by student and first-year class. This is our baseline data set used in the analysis. Most classes taken in the freshman year last two semesters and are worth 1 credit. About 15% of courses are fall semester courses and another 15% are winter courses, worth 0.5 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 classes or even to encourage better teacher evaluations. Fortunately, several courses use identical tests or assignments across classes. We include results using students enrolled in these courses with different instructors. We also focus on outcomes for whether a student drops a course early (between the first day and the end of the first month), late (between the end of the first month and the last day), and the total number of subsequent courses taken in the same subject as the first-year course.

III. Instructor Value Added

To asses the overall importance of instructor differences, we estimate and compare instructor fixed effects. Consider, for example, a standardized outcome of interest, y_{iktps}, with mean 0, for student *i,* in course *k,* in year *t,* with instructor (professor) *p,* in class *s.* We can decompose y_{iktps} by the following:

$$y_{iktps} = \delta_p + \delta_k + \delta_t + \delta_{iktps}, \tag{1}$$

where δ_k is a course-specific effect, δ_t is a year-specific effect, and δ_{iktps} reflects student-specific effects and effects other than instructor, course, or time.

The instructor fixed effect, δ_p, captures the expected increase or decrease in a student's outcome from attending a class with a particular instructor, relative to the course mean. This effect is also sometimes referred to as an instructor's value added. The value-added standard deviation indicates the extent to which *any* teacher differences matter in determining student performance, whether observed or not. A standard deviation of 0 implies that it makes no difference, on average, to a student's performance which teacher she is assigned to.

Measurement error that arises from estimating instructor fixed effects makes value-added comparisons difficult. A number of approaches have been adopted to address measurement error that focus on estimating value-added variance rather than value added for each instructor. We adopt an approach similar 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 at the primary and secondary school level. For each course, the covariance between two students with the same instructor but in different classes is

$$C(y_{i=1,k,t,p,s=1}, y_{i\neq 1,k,t,p,s\neq 1}) = V(\delta_p) + 2C(\delta_p, \delta_{ik\times tps}). \tag{2}$$

If students take classes independently of who teaches (so that $C(\delta_p, \delta_{iktps}) = 0$), the covariance across classes measures the permanent value-added variance, $V(\delta_p)$. Otherwise, the covariance is an upper-bound estimate of this variance. Matching students taught by the same instructors but in different classes helps avoid bias from selective classroom sorting due to friends wanting to be in same classes with other friends, or students in similar programs ending up in similar classes.[3]

[2] Interestingly, instructor type and salary are generally uncorrelated with perceived effectiveness. The mean 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.

[3] Hoffmann and Oreopoulos (forthcoming) find slightly higher covariance estimates when matching students from same classes.

TABLE 2.—ESTIMATES OF STANDARD DEVIATIONS OF INSTRUCTOR FIXED EFFECTS

Grade, all courses (with mean = 0, s.d. = 1)	0.0849***
	(0.0115)
Grade, courses with standardized tests (with mean = 0, s.d. = 1)	0.0549***
	(0.0185)
Dropped course (with mean = 0.09, s.d. = 0.29)	0.0137***
	(0.0029)
Additional number of same-subject courses (with mean = 0, s.d. = 1)	0.0840***
	(0.0085)
Course fixed effects	Yes
Year fixed effects	Yes
Individual fixed effects	No

Notes: The student-course outcome variable is regressed on course and year fixed effects. For courses with standardized tests, the regression includes course-by-year fixed effects. The residuals are used to estimate 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 1996 and 2004. Standard deviations for the grades are calculated from a subsample of courses in which examinations are standardized across sections in the same academic year.

We follow the covariance estimation procedure used by Page and Solon (2003). Let y'_{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:

$$\hat{C}(y'_{i,ktps}, y'_{i',ktps}) = \left[\sum_{k=1}^{K}\sum_{t'=1}^{T}\sum_{p'=1}^{P}\sum_{s=1}^{S_{ktp}}\sum_{s'=1}^{S_{ktp}}\sum_{i=1}^{I_{ktps}}\sum_{i'=1}^{I_{ktps}} y'_{i,k't'p's}\, y'_{i'\neq i,k't'p's'\neq s}\right]/N', \tag{3}$$

where I_{ktps} is the number of individuals within class $ktps$, S_{ktp} is the number of classes in a course, K is the number of courses, T 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 2. All outcome variables, except course completion, are standardized to have mean 0 and standard deviation 1. The across-class covariance for first-year standardized course grade is 0.007. Column 1 shows the square root of this value, 0.085, which is an estimate for the standard deviation of instructor fixed effects on grades for a given course. A 2 standard deviation difference in instructor fixed effects between students would thus account for approximately a 2.4 percentage point difference in grade performance ($2 \times 0.085 \times 14$, with 14 being the standard deviation of a first-year student's grade). Some of this variation may simply be due to different grading standards across instructors, rather than value added. Even with grading standards at the discretion of the instructor, the portion of a course's grade distribution attributable to being taught by different instructors is small.

In an effort to remove grade discretion effects, results in the second row of table 2 are from the subset sample of courses for which students in different classes with different instructors take the same tests. These courses are first-year chemistry, physics, and mathematics. Student grades are regressed on course-by-year fixed effects so that residual differences cannot be due to grade discretion. The value-added standard deviation estimate is 0.055, about two-thirds the estimate from using the full sample with courses graded by each class instructor. This value-added variation at the college level is considerably smaller than estimates of value-added variation at the primary and secondary school level (less than half).[4]

The standard deviation of value added to a student's likelihood of dropping out is about 0.012. This suggests, for a class size of 200, an additional five students would drop the course if an instructor's value added was 2 standard deviations lower. We also measure value-added variation for getting students to take additional courses in following years in the same subject as the first-year course. A 2 standard deviation increase in the effect an instructor has on the number of additional same-subject courses leads to an average increase of about 16% of the standard deviation in the number of subject courses.

We interpret these findings to suggest that the expected impact from assignment to a different college instructor is likely to be small, but not 0. Primary and secondary school teachers appear to have more influence on achievement. Important outliers may exist, however, and perhaps observable instructor characteristics may help predict the likely influence instructors have on student achievement. We turn to this analysis below.

IV. Instructor Qualities

College administrators often use instructor evaluations for making salary and promotion decisions. To see whether these measures and other observable instructor characteristics predict classroom behavior and subsequent behavior, we use a more specific version of equation (1):

$$y_{iktps} = \beta Q_p + \gamma X_{kts} + \delta_i + \delta_k + \delta_{iktps}, \tag{4}$$

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 by instructor-course-groups.

[4] Rockoff (2004), for example, estimates a value-added standard deviation of 0.11 among elementary school students. Rivkin, Hanushek, and Kain (2005) arrive at similar estimates for Texas elementary students. Kane, Rockoff, and Staiger (2006) find 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 0.25 of its standard deviation. Aaronson, Barrow, and Sander (2002) estimate that one semester with a high school teacher rated 2 standard deviations higher in value added would increase standardized math score performance by 0.25 to 0.45 of a standard deviation.

TABLE 3.—REGRESSIONS OF CLASS DROPOUT, GRADE, AND NUMBER OF SAME-SUBJECT CLASSES TAKEN IN SUBSEQUENT YEARS ON YEAR-ONE INSTRUCTOR QUALITY

	Mean and s.d. of Instructor Quality Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Panel A: Dependent Variable: Dropped Course (mean = 0.091, s.e. = 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 professor	*0.180*					0		0.004
	(0.320)					[0.011]		[0.014]
Part-time instructor	*(0.242)*					0.006		0.007
	(0.385)					[0.010]		[0.013]
Top salary	*0.20*						0.011	0.011
	(0.332)						[0.010]	[0.012]
		Panel B: Dependent Variable: Additional Number of Same-Subject Courses (mean = 1.563, s.e. = 3.062)						
Instructor effectiveness	*5.54*	0.05	0.049	0.045				0.05
	(0.572)	[0.056]	[0.040]	[0.040]				[0.042]
Lecturer	*0.310*				−0.009	−0.058		−0.066
	(0.434)				[0.071]	[0.069]		[0.087]
Assistant or associate professor	*0.180*					−0.144		−0.141
	(0.320)					[0.093]		[0.112]
Part-time instructor	*(0.242)*					−0.057		−0.043
	(0.385)					[0.082]		[0.089]
Top salary	*0.20*						0.056	0.007
	(0.332)						[0.084]	[0.103]
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		103,780	103,780	103,780	103,780	103,780	103,780	103,780

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 who earns more than $100,000 in 2005 Canadian dollars. Standard errors are clustered by instructor-course-groups and are in square brackets. ***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 1996 and 2004.

Course fixed effects account for course-specific outcome differences, so that we are identifying off of the within-course variation. The key identification strategy for estimating β is to use instructor quality variation across each student's set of first-year classes. The within-course analysis provides an intuitive counterfactual estimate of how different a student's subsequent achievement would be if she enrolled in the same course but with a different type of instructor. Student fixed effects absorb tendencies for some types of individuals to enroll 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 behavior typical across all courses.

Table 3 presents our main results for course completion and subject interest outcomes. 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 (effectiveness is averaged over all student evaluations recorded during academic years starting between 1996 and 2004). The estimated standard errors account for clustering of residuals by instructor-course-groups. The sample includes first-year students entering university between 1996 and 2004 who are 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 with 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 consistent with the possibility that few first-year students likely choose courses based on 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

TABLE 4.—REGRESSIONS OF COURSE GRADE ON TEACHER QUALITY

		All Courses		Courses with Standardized Tests	
		(1)	(2)	(3)	(4)
Mean and s.d. of dependent variables		*69.18 (13.974)*		*69.38 (14.624)*	
Instructor effectiveness	*5.54*	1.078	0.741	0.988	0.88
	(0.572)	[0.182]***	[0.114]***	[0.242]***	[0.137]***
Lecturer	*0.310*	−1.171	−0.022	−1.415	−0.128
	(0.434)	[0.420]***	[0.304]	[0.826]*	[0.383]
Assistant or associate professor	*0.180*	−1.1	−0.208	−3.427	−0.774
	(0.320)	[0.538]**	[0.377]	[1.055]***	[0.528]
Part-time instructor	*(0.242)*	−0.415	0.273	−1.158	0.204
	(0.385)	[0.389]	[0.265]	[0.776]	[0.411]
Top salary	*0.20*	−0.097	0.034	−0.082	0.493
	(0.332)	[0.384]	[0.255]	[0.831]	[0.294]*
Student fixed effects		Yes	Yes	Yes	Yes
Course fixed effects		Yes	No	Yes	No
Course-year fixed effects		No	Yes	No	Yes
Number of observations		94,301		36,204	

Notes: The table shows results from regressing course grades on the teacher quality variable and a set of teacher rank measures. All regressions include individual fixed effects and 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 who earns more than $100,000 in 2005 Canadian dollars. Standard errors are clustered by instructor-course-groups and are in square brackets. *** 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 1996 and 2004. “Standardized tests” indicates a subsample including courses with standardized tests across sections in the same academic year.

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 B of table 3 uses the number of additional same-subject courses taken in subsequent years at the university. This variable indicates subject interest and will be higher for students that specialize in the same area of study as the first-year introductory course. The results suggest that subjective and objective instructor qualities have minimal influence on subject specialization. A 2 standard deviation increase in perceived instructor effectiveness increases the number of courses taken in second year in the same subject by 0.06 courses—6% of the outcome variable’s standard deviation. All estimated instructor quality effects combined in column 7 of panel B are insignificantly different from 0.[5]

Table 4 shows results for grade performance outcomes. In some courses, grades may be adjusted by the instructor to normalize across classes or even to encourage better teacher evaluations. We therefore contrast the estimates from our full sample with those from a subsample of courses for which students in different classes with different instructors take the same tests. Accordingly, we also provide results when replacing course and year fixed effects with course-by-year fixed effects in columns 2 and 4. This helps to isolate instructor effects from differences in instructors across classes in the same year.

Column 1 indicates that students with instructors who tend to receive better evaluations also tend to receive significantly higher grades. A 2 standard deviation improvement in perceived instructor effectiveness is associated with a 1.2 percentage point increase in the classroom average grade (an increase of 0.088 standard deviations). When conditioning on course-by-year fixed effects instead of course and year fixed effects in column 2, the point estimate is slightly lower. Columns 3 and 4 focus on the subset of courses where students take the same examinations across classes to rule out the possibility that instructors that grade easier may tend to receive better evaluations. The point estimates are about the same as the ones using the full sample, suggesting real gains in aptitude from better-evaluated instructors. Students from the full sample taking classes with lecturers and younger professors receive a final grade about 1.1 percentage points lower than students taking classes with full professors. This estimate may reflect lecturers and younger professors tending to grade students worse, since the relationship does not hold when focusing on courses where students take the same exams.[6]

Students evaluate instructors across a variety of traits. Table 5 shows estimates of the effects of alternative measures of subjective quality on student achievement.[7] 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. The choice

[5] Results using students and instructors from second-year courses are very similar and shown in Hoffmann and Oreopoulos (forthcoming). Subjective instructor effectiveness predicts second-year course dropout outcomes and grade performance. Instructor rank, part-time or full-time status, or salary do not.

[6] The correlation between estimated instructor fixed effects and subjective effectiveness is about −0.12 for course dropout outcomes, and 0.17 for grade outcomes.

[7] We also examined the effects from young experience by restricting our sample to young faculty whose first year of hire is identified in the data. Since few junior professors teach first-year students, we also included data from second-year courses. The results were imprecise, but overall, suggested experience among young professors increases course completion rate and subject interest, but has no effect on grades.

TABLE 5.—REGRESSIONS OF STUDENT-CLASS OUTCOMES ON SUBJECTIVE QUALITY

	Mean and s.d. of Instructor Quality Variable	Dropped Course (1)	Grade, All Courses (2)	Grade, Courses with Standardized Tests (3)	Additional Number of Same-Subject Courses (4)
Mean and s.d. of dependent variables		*0.091*	*69.18*	*69.38*	*1.563*
		(0.288)	*(13.974)*	*(14.624)*	*(3.062)*
Average instructor's perceived	*5.54*	−0.015	0.939	0.775	0.045
effectiveness	*(0.572)*	[0.005]***	[0.209]***	[0.112]***	[0.040]
Fraction that give instructor highest	*0.256*	−0.042	3.242	2.364	0.112
rating	*(0.171)*	[0.017]**	[0.558]***	[0.258]***	[0.128]
Fraction that give instructor lowest	*0.057*	0.141	−6.465	−6.178	−0.339
rating	*(0.061)*	[0.046]***	[2.354]***	[1.350]***	[0.331]
Provides helpful comments and	*4.909*	−0.025	1.645	1.238	0.058
feedback	*(0.463)*	[0.009]***	[0.352]***	[0.153]***	[0.060]
Available to meet	*5.231*	−0.013	1.06	0.998	0.002
	(0.488)	[0.006]**	[0.352]***	[0.188]***	[0.055]
Answers questions clearly and	*5.316*	−0.018	0.974	0.883	0.065
effectively	*(0.495)*	[0.006]***	[0.293]***	[0.185]***	[0.045]
Communicates enthusiasm and	*5.525*	−0.014	0.657	0.694	0.054
interest	*(0.639)*	[0.005]***	[0.205]***	[0.106]***	[0.036]
Explains concepts clearly	*5.393*	−0.015	0.911	0.825	0.054
	(0.563)	[0.005]***	[0.194]***	[0.098]***	[0.041]
Presents material in organized	*5.404*	−0.013	0.848	0.723	0.027
manner	*(0.557)*	[0.005]**	[0.183]***	[0.123]***	[0.042]
Provides fair evaluation of student	*4.980*	−0.024	2.21	1.383	0.095
learning	*(0.421)*	[0.009]***	[0.357]***	[0.164]***	[0.068]
Would take course again given	*65.637*	−0.001	0.08	0.055	0.006
experience	*(15.175)*	[0.000]***	[0.015]***	[0.008]***	[0.003]**
Fraction of students who filled out	*0.410*	−0.023	−0.029	2.287	−0.067
the evaluation	*(0.444)*	[0.018]	[1.134]	[0.330]***	[0.117]
Number of observations		103,780	94,301	36,204	103,780

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, and time of day and week controls. In column 6, course-year fixed effects are used instead of course and year fixed effects separately. 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 seven-point scale, with 1 meaning strongly disagree and 7 meaning strongly agree. Standard errors are clustered by instructor-course-groups and are in square brackets. *** 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 1996 and 2004. "Standardized tests" indicates a subsample including courses with standardized tests across sections in the same academic year.

of the instructor evaluation measure does not substantially affect the implied effects, perhaps not surprisingly, since the evaluation measures are all highly correlated.[8] The retake rate predicts the largest differences in achievement. A 2 standard deviation increase in this measure is associated with a 3 percentage point reduction in the course dropout rate[9] and a 1.6 percentage point increase in the classroom standardized grade average. Interestingly, the fraction of students enrolled that fill out an instructor's evaluation indicates that the level of class attendance also predicts student achievement.

Table 6 explores how the main results differ by gender, mother tongue, high school grade, and program of study. The estimated effects from perceived instructor effectiveness are quite similar regardless of student 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 subgroups (results not shown in table). The table also suggests that instructor quality impacts course completion and subject interest outcomes differently depending on students' past performance. From columns 5 and 6, the subjective instructor quality helps predict course completion outcomes for university students with poorer high school grades only. Among students in the lowest high school grade quartile, a 2 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 0. The table also indicates that lower-performing students are more likely to take additional courses in the same subject when taught by an instructor with higher perceived effectiveness. Grade effects, however, are about the same for students with different ability backgrounds.

The last two columns in table 6 explore whether instructor effects are larger among students majoring in the same subject as the course. We sort our sample by whether students' program of study matches with the subject in the course (for example, chemistry majors with science and math classes, commerce majors with business and economics classes). The interaction between students and instructors may be larger for students taking

[8] Multicollinearity across instructor quality variables leads to imprecise coefficient estimates when these variables are used together in the same regression.

[9] Interestingly, subjective instructor quality predicts both dropping out early (within one month of the course starting date) and dropping out late (after one month). These results are available upon request.

TABLE 6.—TEACHER QUALITY REGRESSIONS, BY STUDENT CHARACTERISTIC

	Male (1)	Female (2)	Mother Tongue English (3)	Mother Tongue Non-English (4)	Lowest HS Quartile (5)	Highest HS Quartile (6)	Science Majors (7)	Arts and Social Science Majors (8)	Nonmajor Course Sample (9)	Major Course Sample (10)
Dropped course	−0.013	−0.012	−0.014	−0.01	−0.049	0.002	−0.01	−0.018	−0.03	−0.01
	[0.006]**	[0.006]**	[0.006]**	[0.006]*	[0.012]***	[0.004]	[0.005]*	[0.007]**	(0.007)***	[0.004]**
Grade, all courses	0.974	1.094	1.06	1.084	1.298	1.035	1.049	0.996	0.843	1.147
	[0.195]***	[0.206]***	[0.182]***	[0.207]***	[0.317]***	[0.186]***	[0.233]***	[0.224]***	[0.253]***	[0.216]***
Grade, courses with standardized within-year variation	0.749	0.929	0.936	0.691	0.698	0.909	0.767	1.237	0.299	0.755
	[0.196]***	[0.140]***	[0.140]***	[0.175]***	[0.400]*	[0.145]***	[0.118]***	[0.338]***	[0.641]	[0.116]***
Additional number of same-subject courses	0.003	0.077	0.076	0.002	0.226	0.017	−0.044	0.084	0.188	0.023
	[0.044]	[0.049]	[0.043]*	[0.052]	[0.071]***	[0.039]	[0.048]	[0.059]	[0.087]**	[0.041]
Number of observations	42,198	61,582	65,820	37,960	25,945	25,945	41,768	62,012	21,149	54,175

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, and time of day and week controls. "Within-year variation" indicates that course-year fixed effects are used instead of course and year fixed effects separately. The outcome variable varies across rows; the subsample considered varies across columns. Standard errors are clustered by instructor-course-groups and are in square brackets. ***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 1996 and 2004. "Standardized tests" indicates a subsample including courses with standardized tests across sections in the same academic year.

required courses than for those taking electives. In fact, differences in instructor quality appear to affect dropout and course selection behavior more for students taking electives than for students taking required courses. Grade effects, though, are about the same for classmates from different programs of study.

V. Conclusion

This paper is among the first to focus on the importance of teacher quality at the college level. We use a new administrative data set 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 classes within the same course. The within-course analysis provides an intuitive counterfactual estimate of how different a student's subsequent achievement would 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 to minimize remaining selection issues.

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). 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 seven-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, standardized grades would rise by about 8% of a standard deviation, and the number of related courses taken in second year would increase by about 4%. 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 overall college instructor influence on student achievement is smaller than the overall influence suggested in earlier 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 with class grade distributions across elementary and secondary school instructors. Perhaps by the time students enter college, cognitive ability and motivation are less malleable than in early childhood and, consequently, teachers have less impact. Two caveats are that the effects of hiring instructors outside the quality range examined in this paper (bounded by who is allowed to teach) may matter more and that students may respond differently if exposed to large changes in the teaching environment. Instructor effects on student experience, which are not estimated here, may also be valued.

Often universities are ranked by the fraction of full-time faculty teaching undergraduates. Perceptions also exist that research-based faculty tend to teach worse because they are too preoccupied. Our results suggest there is not a strong correlation between research-focused and teaching-focused college instructors—both have effective and noneffective teachers within each group. At the margin, instructors do not make a large difference to student achievement but to the extent that they do, instructor evaluations can be used to evaluate these effects.

REFERENCES

Aaronson, Daniel, Lisa Barrow, and William Sander, "Teachers and Student Achievement in the Chicago Public High Schools," Federal Reserve Bank of Chicago working paper series WP-02-28 (2002).

Bettinger, Eric, and Bridget Terry Long, "Does It Matter Who Teaches the Course? The Impact of Adjuncts and Graduate Instructors," Case Western working paper (2004).

——— "Help or Hinder? Adjunct Professors and Student Outcomes," mimeograph (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* 100 (February 1992), 1–40.

——— "School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina," NBER working paper no. 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, DC: U.S. Government Printing Office, 1966).

Ehrenberg, Ronald G., and Liang Zhang, "Do Tenured and Tenure-Track Faculty Matter?" *Journal of Human Resources* XL (2005), 647–659.

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* 23:3 (April 1994), 5–14.

Hoffmann, Florian, and Philip Oreopoulos, "A Professor Like Me: The Influence of Instructor Gender on College Achievement," *Journal of Human Resources* (forthcoming).

Jacob, Brian A., and Lars Lefgren, "Principals as Agents: Subjective Performance Measurement in Education," Harvard University 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 no. W12155 (April 2006).

Page, Marianne, and Gary Solon, "Correlations between Brothers and Neighboring Boys in Their Adult Earnings: The Importance of Being Urban," *Journal of Labor Economics* 21 (October 2003), 831–855.

Rivkin, Steven G., Erik Hanushek, and John Kain, "Teachers, Schools and Academic Achievement," *Econometrica* 73:2 (2005).

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

Wachtel, Howard K., "Student Evaluation of College Teaching Effectiveness: A Brief Review," *Assessment & Evaluation in Higher Education* 23:2 (1998), 191–212.