Public Finance Review
Volume 34 Number 4
July 2006 1-26

10.1177/1091142106289330
http://pfr.sagepub.com
hosted at
http://online.sagepub.com

Out-of-Field Teachers and Student Achievement

Evidence from "Matched-Pairs" Comparisons

Thomas S. Dee
Swarthmore College and NBER
Sarah R. Cohodes
The Urban Institute

This study examines whether subject-specific teacher certification and academic degrees are related to teacher quality. The research design exploits contemporaneous, within-student comparisons made possible by a unique feature of the National Education Longitudinal Study of 1988 (NELS:88). Specifically, NELS:88 contains subject-specific outcomes for eight-grade students in two subjects as well as data on their teachers for those subjects. The analysis of these data indicates that assignment to a subject-certified teacher is associated with higher test scores. However, these gains appear to be concentrated in social studies and mathematics. Furthermore, the authors also find that subject-certified teachers are not more effective at promoting the intellectual engagement of their students but are more likely to have negative opinions of a given student's performance.

Keywords: *teacher quality; student achievement; certification*

There is a wide consensus among researchers that teacher quality is an important determinant of student achievement. The No Child Left Behind Act (NCLB) explicitly acknowledged this view by requiring that every elementary and secondary public school teacher be "highly qualified" by the end of the upcoming school year. NCLB's definition of "highly qualified" requires that teachers have bachelor degrees and state certification

Authors' Note: We would like to thank Jan Ondrich and the participants at the Teacher Quality and Teacher Retention Conference at Syracuse University for helpful suggestions.

that has not been waived on a temporary or emergency basis. However, NCLB also requires that highly qualified teachers separately demonstrate proficiency in the subjects that they teach. For middle and high school teachers, this "demonstration of competency" can be met by having a college major or graduate degree in the subject they teach, credits equivalent to a college major, passing a state-developed subject-matter test, or having advanced certification.[1]

These features of NCLB reflect the growing concern about "out-of-field" teaching. Out-of-field teaching is typically defined as a situation where a teacher does not have an academic major or certification in the subject they teach (e.g., U.S. Department of Education 2004). In the 1999-2000 school year, nearly two-thirds of the middle school teachers whose main assignment was mathematics (and roughly half of science teachers) did not have a major in their subject (U.S. Department of Education 2004, Table B-2). And roughly 20 percent of mathematics and science teachers did not have certification in their subject.

Critics (e.g., Ingersoll 1999; Jerald 2002) have argued that the prevalence of out-of-field teaching in such core academic subjects is "unacceptably high." In particular, the concern is that exposure to an out-of-field teacher compromises student achievement as measured by test scores as well as forms of achievement not well captured by standardized tests (e.g., student interest in the subject and critical thinking skills). Furthermore, out-of-field teaching may also contribute to achievement gaps since it occurs more frequently among poor and minority children as well as those in lower tracks.

However, the direct evidence that out-of-field teaching actually harms student achievement is surprisingly limited.[2] More specifically, it consists largely of a few cross-sectional studies that focus on mathematics achievement at the high school level. This evidence consistently indicates that student performance is higher when the teacher has a college major, additional coursework, or certification in their subject (Monk and King 1994; Goldhaber and Brewer 1997a, 1997b, 2000; Rowan, Chiang, and Miller 1997). However, a fundamental concern with this limited evidence is that it may be biased by the unobserved determinants of student achievement. More specifically, these results could overstate the benefits of "in-field" teachers in the likely circumstance that students with an unobserved propensity for achievement (or achievement growth) are more likely to be assigned to such teachers.

This study presents new empirical evidence on whether teachers with subject-specific certification and degrees are more successful than other teachers at improving students' educational outcomes. This study contributes to the extant literature on this topic in several distinct ways, by focusing on additional academic subjects, on middle school students, and on a diverse set of outcome measures. Perhaps the most important innovation is an identification strategy that can eliminate the bias that would occur in more basic cross-sectional evaluations when a student's unobserved propensity for achievement is correlated with the subject-specific qualifications of their teachers. More specifically, this study examines the effects of teacher qualifications in models that condition on student fixed effects. This type of panel analysis is possible because of a unique feature of the National Education Longitudinal Study of 1988 (NELS:88). Specifically, NELS:88 collected data from each of the sampled student's teachers in two of four distinct academic subjects (mathematics, science, social studies, English). This implies that NELS:88 contains contemporaneous data on student outcomes and teacher observables in two different academic subjects. The availability of these "matched-pairs" data makes it possible to estimate how the *same* student performed when assigned to teachers whose subject qualifications differed.

This study is organized as follows. The next section provides a brief overview of the relevant literature on teacher quality. Then we discuss the NELS:88 data and our econometric specifications. The next section presents our results, and the final section concludes with some discussion of how these results relate to the prior literature and what they mean for current policies.

Teacher Qualifications and Quality

Recent studies indicate that there is substantial variation in the quality of teaching within schools (Rockoff 2004; Hanushek, Kain, and Rivkin 1998; Hanushek et al. 2005).[3] For example, Hanushek, Kain, and Rivkin (1998) find that teacher quality accounts for at least 7 percent of the total variation in student achievement. Rockoff (2004) finds that a one–standard deviation increase in teacher quality raises student achievement in reading and math by 0.1 standard deviations. However, while the importance of teacher quality seems uncontested, the importance of specific, observed teacher characteristics is often highly controversial.

For example, teacher pay is often linked to the completion of postsecondary degrees. However, the available evidence on whether teachers with more advanced degrees are more effective is mixed.[4] Another area of particular controversy involves whether teachers who have obtained state certification are more effective than those who have not. In general, teacher certification requires completion of a teacher preparation program and some evidence of subject area knowledge (e.g., passing a subject test, a college major in the field to be taught). Proponents of teacher certification, such as Darling-Hammond (2002), argue that these procedures ensure that teachers have the professional skills and knowledge (e.g., classroom management, curriculum development, and pedagogical technique) that are particularly critical for the education of at-risk students. Darling-Hammond also discusses evidence that teacher education reduces attrition from the profession.

However, critics charge that the evidence linking certification to teacher quality is "astonishingly deficient" and that the impediments created by the certification process discourage high-ability individuals from ever entering the teaching profession (Walsh 2001). For example, in his second Annual Report on Teacher Quality (U.S. Department of Education 2004), the former U.S. Secretary of Education wrote that "there is little compelling evidence that certification requirements, as currently structured in most states, are related to teacher effectiveness."

The mixed evidence on the effects of teacher degrees and certification may partly reflect the fact that prior studies did not identify whether the teacher credential was specific to the subject being taught. However, this is not entirely clear since there are relatively few studies that examine the effectiveness of teachers with subject-specific credentials. For example, in a recent review of the literature on teacher quality, Wayne and Youngs (2003) found only three studies of subject-specific teacher degrees and certification (Monk and King 1994; Goldhaber and Brewer 1997a, 1997b, 2000) whose research design met their criteria for being "compelling as opposed to merely suggestive."[5]

However, the studies cited by Wayne and Youngs (2003) suggest that "in-field" teachers are more effective than "out-of-field" teachers. For example, using data from the Longitudinal Survey of American Youth (LSAY), Monk and King (1994) found that high school students had higher gain scores in mathematics when assigned to teachers who had more coursework in mathematics. Similarly, using data from tenth and twelfth graders in NELS:88, Goldhaber and Brewer (1997a, 1997b, 2000) found that, conditional on prior achievement, students assigned to teachers with math certification or a mathematics degree had significantly higher math scores than students whose teachers lacked these traits. However, Goldhaber and Brewer (2000) found statistically insignificant results for models of science achievement. They also found that the performance of teachers with probationary and temporary certification in the subject was indistinguishable from that of those with regular certification, and they argued that this "casts doubt" on the claim that standard certification should be required of all teachers.[6] Rowan, Chiang, and Miller (1997) found that tenth graders in NELS:88 had higher scores in mathematics when their teacher had a college degree in math.[7]

This study attempts to address some of the gaps in this literature. First, prior evidence focuses exclusively on high school students, while this study focuses on outcomes among eight graders. Critics (Jerald 2002) allege that out-of-field teaching is a particular problem in both middle and high schools. Furthermore, the regulations that apply to subject-specific teacher credentials are often the same for high and middle school teachers (most notably, as in NCLB). However, it is quite possible that out-of-field teaching is less relevant for younger students since they are following a less advanced curriculum.

Second, the prior literature largely focuses on teacher qualifications and student outcomes in math and science. However, this study examines teacher qualifications and outcomes in four major academic areas (i.e., mathematics, science, social studies, and English). The prior emphasis on mathematics is in some sense justifiable. Specifically, mathematics may provide a relatively powerful test of the effects of out-of-field teaching because the correspondence between a college degree in mathematics and the material being taught is relatively strong. In contrast, a middle school science teacher who has a degree in biology would be classified as "in field" but may not be particularly effective in teaching material related to physics and chemistry. Similarly, a social studies teacher with a degree in sociology or economics may technically be "in field" but have relative little proficiency in history. However, given that regulations of teacher credentials do not typically make such fine-grained distinctions, additional evidence from fields other than mathematics is relevant for evaluating current policies.

Third, the prior studies focused exclusively on test scores as an outcome measure. Apart from concerns about the meaning of variation in low-stakes tests, there is the possibility that test scores may fail to capture some of the important educational consequences of out-of-field teaching. Specifically, Ingersoll (1999) suggests that out-of-field teachers may be less effective at promoting the intellectual engagement and enthusiasm of students. This study examines this question using student-reported attitudes toward each academic subject. This study also examines teacher perceptions of an individual student's performance as outcome measures. Although not commonly studied by economists, a teacher's subjective perception of a student's performance is a potentially important outcome because, even when it is inaccurate, it can influence the student's access to future opportunities.[8]

Fourth, the prior studies rely exclusively on cross-sectional comparisons—that is, regression-adjusted differences in achievement across students who were assigned to in-field teachers relative to those who were not. One concern with this approach is that the students who are assigned to teachers with "better" observables are more likely to be those with an unobserved propensity for high achievement. In fact, we found using the NELS:88 data described below that the within-school assignment to an in-field teacher was significantly and positively related to socioeconomic status. This suggests that prior studies may overstate the true benefits of a subject-qualified teacher. The approach to addressing this "omitted-variables" problem has been to estimate "value-added" specifications that condition on prior achievement. However, that procedure may not address the source of bias adequately. More specifically, the students assigned to subject-qualified teachers may also be those who have an unobserved propensity for achievement *growth* (e.g., high socioeconomic status students who may have less "summer learning melt"). In the next section, we introduce data and methods that address this concern in an alternative manner.

National Education Longitudinal Study of 1988

NELS:88 is a nationally representative, longitudinal study that began in 1988 with a sample of 24,599 eighth-grade students from 1,052 public and private schools (Ingels et al. 1990). NELS:88 had a two-stage sampling design. Schools, the primary sampling unit, were selected with probabilities proportional to their eighth-grade enrollment. Approximately 26 students were then randomly chosen within each participating school.

NELS:88 also fielded questionnaires to the teachers responsible for teaching each of the selected students in two of four academic subjects: mathematics, science, reading, and social studies. The surveyed teachers were chosen by randomly assigning each participating school to one of four subject area groupings: mathematics/reading, mathematics/social studies, science/reading, and science/social studies. Two completed teacher surveys are available for 21,324 of the eighth-grade students. This reduction in the sample reflects the fact that some students did not have a class in one or both of their designated academic subjects. This feature of the sample underscores a limitation of this research design. Specifically, the within-student comparisons examined here explicitly rely on the ability to compare contemporaneous outcomes across two academic subjects. Therefore, these comparisons are only relevant for the students who are observed taking classes in both academic subjects.[9] Eliminating students who attended private schools further reduces the sample to 16,901 students. However, because the unit of observation is each teacher-student pairing, the final data set consists of 33,802 observations (Table 1).

The students participating in NELS:88 completed multiple-choice tests in the subjects taught by these teachers.[10] For purposes of this analysis, the formula scores on these tests have been standardized by subject so that the changes in these scores (STEST) can be understood as effect sizes. The other outcome variables used in this study reflect the student's perceptions of an academic subject and the teacher's perceptions of the sampled student.

This analysis focuses on three variables reflecting the students' perception of the class and subject taught by the responding teacher. More specifically, students were asked whether they see the subject as useful for their future, whether they look forward to class in the subject, and whether they are afraid to ask questions in that subject. The students were given four options in response to these questions (strongly agree, agree, disagree, strongly disagree), which are coded as integers from 1 to 4. However, for ease of interpretation, the order of the responses to the "afraid" question was reversed. This implies that, for each of the three questions, higher values of the ordinal response imply a negative view of the subject. Furthermore, within each subject, the responses to each of these three questions were standardized (Table 1) to create the variables used in this analysis (i.e., AFASK, NOTLF, NOTUSE).

The remaining outcome variables used in this study are three pejorative teacher assessments: whether the student rarely completed homework (NOHWK) and whether the student was seen as consistently inattentive (INATT) or frequently disruptive (DISRUPT). The response options to these questions were simply yes or no, so these three variables are binary. One potential complication with these variables is that a student may become disruptive or inattentive simply because he or she has mastered the classroom material relative to their peers. However, the data do not support that hypothesis. More specifically, using these NELS:88 data, we found that, conditional on student and subject fixed effects, students performed significantly lower on subject tests when the teacher for that

Table 1
Descriptive Statistics Matched Student-Teacher Observations, Eighth-Grade Public School Students, National Education Longitudinal Study of 1988 (NELS:88)

Variable	Description	Mean	Standard Deviation	Sample Size
STEST	Test score in subject	0	1.0	32,646
NOTUSE	Subject not useful for my future	0	1.0	32,152
NOTLF	Do not look forward to subject	0	1.0	32,246
AFASK	Afraid to ask questions in subject class	0	1.0	32,197
NOHWK	Student rarely completes homework	0.224	0.417	33,022
INATT	Student is consistently inattentive	0.226	0.418	33,962
DISRUPT	Student is frequently disruptive	0.137	0.343	33,018
SCERTIFD	Certified by state in subject	0.884	0.320	33,479
MAJOR	Undergraduate/graduate major in subject	0.565	0.496	33,411
OTHSEX	Teacher of opposite gender	0.503	0.500	33,802
TFEMALE	Female teacher	0.541	0.498	33,802
OTHRACE	Teacher of opposite race/ethnicity	0.316	0.465	33,802
TBLACK	Black teacher	0.091	0.287	33,802
THISP	Hispanic teacher	0.025	0.155	33,802
TOTHER	Teacher of other race/ethnicity	0.025	0.157	33,802
TE1	Teacher experience missing	0.006	0.079	33,808
TE2	One to three years of teacher experience	0.099	0.299	33,802
TE3	Four to six years of teacher experience	0.084	0.277	33,802
TE4	Seven to nine years of teacher experience	0.090	0.285	33,802
TE5	Ten to twelve years of teacher experience	0.105	0.306	33,802
TE6	Thirteen to fifteen years of teacher experience	0.124	0.330	33,802

TE7	Sixteen to eighteen years of teacher experience	0.149	0.356	33,802
TE8	Nineteen to twenty-one years of teacher experience	0.109	0.311	33,802
TE9	Twenty-two to twenty-four years of teacher experience	0.082	0.274	33,802
TE10	Twenty-five or more years of teacher experience	0.151	0.359	33,802
CLSSIZE	Class size	24.5	5.87	33,162
PCTLEP	Percentage of class with limited English proficiency	0.014	0.072	31,362

subject viewed them negatively. The students viewed negatively by teachers were also substantially less likely than other students in their school to take any Advanced Placement courses over the subsequent two years and more likely to have dropped out of high school. However, a complication that does appear to be relevant to this analysis is that an in-field teacher may be more likely to view a student pejoratively simply because he or she has higher expectations (e.g., assigning more homework).

The two key independent variables used in this study are dummy variables, one (SCERTIFD) indicating whether the teachers are state certified in the subject they are teaching and another (MAJOR) indicating whether they have an undergraduate or graduate major in the subject they are teaching. However, a number of other controls for teacher and classroom observables are also included in the regression models discussed below. These include dummy variables for the gender and race-ethnicity of the teacher as well as two dummy variables that identify whether the student shares the teacher's gender and the teacher's race-ethnicity (Table 1).[11] Teacher experience is measured by ten categorical dummies (Table 1). This relatively unrestrictive approach to measuring teacher experience may be important given the evidence of nonlinear returns to teacher experience (Hanushek et al. 2005). The final controls capture two observable traits of the teacher's class, the number of students in the class, and the percentage of students in the class who are limited English proficient (LEP).

Specifications

The basic econometric specification presented here is a straightforward variation of those used to examine another type of matched-pairs data: information on the schooling and labor market outcomes of monozygotic twin pairs (Ashenfelter and Krueger 1994; Ashenfelter and Rouse 1998; Rouse 1999). More specifically, the initial specification allows the educational outcome of student i in class c of subject 1 (i.e., y_{1ic})to be a function of observed student traits X_i and the observed traits of the teacher and classroom Z_{1c}:

$$y_{1ic} = \alpha \mathbf{X}_i + \lambda \mathbf{Z_{1c}} + \mu_i + \varepsilon_{1ic}. \tag{1}$$

The terms μ_i and ε_{1ic} are, respectively, a student fixed effect and a mean-zero error term. The term Z_{1c} includes indicators for whether the teacher is "in field" (i.e., subject certification, SCERTIFD; undergraduate or graduate in-subject major, MAJOR) as well as fixed effects for the subject of the class and other observed attributes of the teacher and the classroom (Table 1).

Equation (1) refers to the student when observed in either math or science. We assume that a similar specification applies when the student is observed in the second subject (i.e., social studies or reading):

$$y_{2ic} = \alpha \mathbf{X}_i + \lambda \mathbf{Z_{2c}} + \mu_i + \varepsilon_{2ic}. \tag{2}$$

A methodological concern that motivated this approach was that a student's assignment to a subject-qualified teacher may be correlated with the unobserved student effects that influence educational outcomes. The existence of this sort of nonrandom assignment implies that the parameters of interest cannot be reliably identified by evaluating equations (1) or (2) in isolation. However, the matched-pairs nature of the NELS:88 data may make it possible to eliminate this potential source of bias. Specifically, differencing equations (1) and (2) leads to the following:

$$(y_{1ic} - y_{2ic}) = \lambda(\mathbf{Z_{1ic}} - \mathbf{Z_{2ic}}) + (\varepsilon_{1ic} - \varepsilon_{2ic}). \tag{3}$$

The first-difference (FD) estimates based on equation (3) provide a way to identify the effect of assignment to a subject-qualified teacher, conditional on μ. However, to examine whether student unobservables are actually related to the subject qualifications of their assigned teachers, we also present ordinary least squares (OLS) estimates of the effects of subject certification and in-subject major, based on stacked versions of equations (1) and (2) that condition on school fixed effects instead of student fixed effects.

Furthermore, as discussed earlier, the effects of subject certification and in-subject major may matter more in some subjects (e.g., mathematics) than in others. To examine this issue, we also present the results of models that interact subject certification and in-subject major with the subject fixed effects. We also examine the effects of these variables in subsamples defined by school and student traits of interest (e.g., urbanicity, race, and socioeconomic status). It should be noted that the error term in this equation could conceivably be heteroscedastic at the level of the student, the classroom, the teacher, or the school. We experimented with White standard errors clustered at these levels. We found that clustering at the school level led to the most conservative (i.e., the largest) standard errors, and we report those here. This approach is also an appropriate one in light of the fact that schools were the primary sampling unit in NELS:88's sampling design.

As suggested earlier, a key innovation of this research design is the ability to condition on student unobservables, which may be correlated with both a student's propensity to achieve and his or her likelihood of being assigned to a subject-qualified teacher. This approach can unambiguously address the concern that the prior links between "in-field" teachers and student achievement merely reflect the nonrandom sorting of students. However, the research design used here is not entirely immune to similar concerns about biases. But, in this study, the source of such a bias would take different forms. For example, consider a situation where a student is particularly capable in one subject relative to another. It may be that parents and administrators are more likely to assign such a student to a teacher who is qualified in that student's best subject. Under such a scenario, the econometric specification outlined here would overstate the importance of the teacher's subject qualifications for student outcomes. A similar bias would occur if subject-qualified teachers tended to be assigned classrooms with desirable traits that also influenced student outcomes (e.g., smaller classes). We discuss these important possibilities in light of the results we present below.

Results

In Table 2, we present the key results of specifications that examine the effects of subject certification and in-subject major on student test scores. The results in columns (1) and (2), which condition on school—not student—fixed effects suggest that assignment to a teacher with either of these traits increases achievement by quite large and significant amounts. However, in models that condition on student fixed effects, these point estimates are noticeably smaller. In particular, the point estimates reported in columns (1) and (2) are outside the 95 percent confidence intervals associated with the remaining FD estimates. These comparative results indicate that, within schools, students with an unobserved propensity for high achievement are more likely to be assigned to subject-qualified teachers.

The FD estimates consistently indicate that assignment to a teacher with an undergraduate or graduate degree in the subject being taught (i.e., MAJOR) has small and statistically insignificant effects on student achievement. However, these results also suggest that assignment to a subject-certified teacher (i.e., SCERTIFD) increases achievement by statistically significant 0.04 to 0.05 standard deviations.

An issue of interpretation that is worth underscoring is that these models do not indicate the nature of the structural relationship between

Table 2
Ordinary Least Squares (OLS) and First-Difference (FD) Estimates of the Effects of Subject-Specific Teacher Qualifications on Test Scores

Independent Variable	OLS	OLS	FD	FD	FD	FD	FD	FD
SCERTIFD	.148*** (.025)	.122*** (.026)	.044** (.019)	.042** (.021)	.038** (.019)	.036* (.021)	.050** (.020)	.049** (.021)
MAJOR	—	.055*** (.016)	—	.009 (.013)	—	.011 (.013)	—	.005 (.014)
R^2	.2938	.2945	.0012	.0013	.0073	.0073	.0085	.0084
Sample size	30,108	29,911	15,992	15,784	15,992	15,784	14,071	13,887
School fixed effects	Yes	Yes	No	No	No	No	No	No
Student fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Classroom controls	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	No	No	No	No	Yes	Yes

Note: Standard errors, adjusted for school-level clustering, are reported in parentheses. All models include subject fixed effects. For definitions of variables, see Table 1.
*Statistically significant at the 10 percent level. **Statistically significant at the 5 percent level. ***Statistically significant at the 1 percent level.

Table 3
First-Difference (FD) Estimates of the Effect of Subject-Specific Teacher Certification on Test Scores by Sample Traits

Sample	SCERTIFD	Sample Size
Full sample	.050** (.020)	14,071
Urban schools	.107*** (.041)	3,014
Suburban schools	.026 (.033)	5,832
Rural schools	.049 (.033)	5,225
Black and Hispanic students	.040 (.029)	3,343
White non-Hispanic students	.040 (.025)	9,704
Female students	.046* (.026)	7,109
Male students	.056** (.023)	6,962
Low-SES students	.046** (.023)	7,463
High-SES students	.054* (.029)	6,604

Note: Standard errors, adjusted for school-level clustering, are reported in parentheses. All models include subject fixed effects and the classroom and teacher controls. SES = socioeconomic status.*Statistically significant at the 10 percent level.
**Statistically significant at the 5 percent level. ***Statistically significant at the 1 percent level.

subject-certified teachers and test scores. For example, it could be that the requirements of subject certification make teachers better than they otherwise would be. Alternatively, it could be that more able teachers undertake the costs of such certification (i.e., a signaling explanation).[12]

From the perspective of a marginal decision (e.g., whether to hire a teacher with particular credentials), this caveat is not necessarily important since the goal is to identify high-quality teachers. However, this distinction may be more relevant for considering the "general equilibrium" impact of policies such as the NCLB that ostensibly require subject-specific credentials of all teachers. For example, if subject certification were merely a sorting device that currently distinguishes teachers with native ability from those without, the effects of requiring subject certification of all teachers could be attenuated.

In Table 3, we examine how the results in Table 2 might differ by the urbanicity of the school or by the observed traits of individual students, including race-ethnicity, gender, and socioeconomic status (SES). This exercise is highly qualified since reductions in sample size lead to meaningful losses in statistical power in virtually all cases. However, one interesting finding is that the benefits of subject-certified teachers appear to be particularly large in urban schools (i.e., approximately 0.11 standard deviations). Interestingly, these estimated effects are also somewhat higher among male and high-SES students.

In Table 4, we present evidence on whether "in-field" teachers are more effective at promoting engagement among students. The results indicate that assignment to a teacher with subject-specific certification or a degree in the subject did not reduce the likelihood that the student would see the subject as not useful. Similarly, assignment to a subject-qualified teacher did not significantly influence the likelihood the student would not look forward to the subject or feel afraid to ask questions. Furthermore, assignment to a teacher with a subject-specific major led to a weakly significant *increase* in the likelihood that a student viewed that subject as not useful.

In Table 5, we present the key results from models where the teacher's perceptions of the sampled student are the dependent variables. It should be noted that these teacher perceptions appear to be educationally meaningful outcomes. For example, we found that, conditional on student and subject fixed effects, students performed significantly lower on subject tests when the teacher for that subject viewed them negatively. The students viewed negatively by teachers were also substantially less likely than other students in their school to take any Advanced Placement courses over the subsequent four years and more likely to have dropped out of high school. Interestingly, the results in Table 5 suggest that subject-certified teachers were *more* likely to have negative perceptions of their students. More specifically, subject-certified teachers were more likely to see a given student as rarely completing homework and frequently inattentive.

In Table 6, we present the results of models for each of the seven dependent variables where the effects of subject certification are interacted with the subject fixed effects. The results indicate that the test score effects of subject-certified teachers are concentrated in mathematics and social studies. Specifically, these results indicate that assignment to a subject-certified teacher increased achievement by 0.12 standard deviations in math and 0.08 standard deviations in social studies.[13] For this test score model, an F-test allows us to reject the null hypothesis that subject certification has the same effects across the four subjects ($p = .0168$). However, the results for the remaining six dependent variables are statistically imprecise. Specifically, the hypothesis that subject certification has similar effects across subjects cannot be rejected.

Taken together, the test score results suggest that subject-certified teachers increase student achievement and that these effects are concentrated in mathematics and social studies. However, as suggested earlier, the

(text continues on p. 19)

Table 4
First-Difference (FD) Estimates of the Effects of Subject-Specific Teacher Qualifications on Student Perceptions

	Dependent Variable					
Independent Variable	NOTUSE		NOTLF		AFASK	
SCERTIFD	–.017 (.028)	–.025 (.030)	–.014 (.038)	–.003 (.041)	.046 (.029)	.044 (.031)
MAJOR	—	.032* (.017)	—	.0002 (.024)	—	.012 (.018)
R^2	.0045	.0046	.0158	.0156	.0047	.0047

Note: Standard errors, adjusted for school-level clustering, are reported in parentheses. All models include subject fixed effects and the classroom and teacher controls. For definitions of variables, see Table 1.
*Statistically significant at the 10 percent level.

Table 5
First-Difference (FD) Estimates of the Effects of Subject-Specific Teacher Qualifications on Teacher Perceptions

	Dependent Variable					
Independent Variable	NOHWK		INATT		DISRUPT	
SCERTIFD	.045*** (.012)	.038*** (.013)	.041*** (.015)	.033** (.015)	.005 (.011)	.004 (.012)
MAJOR	—	.010 (.008)	—	.018** (.009)	—	.004 (.007)
R^2	.0066	.0068	.0061	.0069	.0059	.0060

Note: Standard errors, adjusted for school-level clustering, are reported in parentheses. All models include subject fixed effects and the classroom and teacher controls. For definitions of variables, see Table 1.
**Statistically significant at the 5 percent level. ***Statistically significant at the 1 percent level.

Table 6
First-Difference (FD) Estimates of the Effects of Subject-Specific Teacher Qualifications by Subject and Outcome

	Dependent Variable						
Independent Variable	STEST	NOTUSE	NOTLF	AFASK	NOHWK	INATT	DISRUPT
SCERTIFD × Mathematics	.116*** (.031)	–.007 (.048)	.088* (.053)	.088** (.044)	.041** (.020)	.042* (.023)	.006 (.018)
SCERTIFD × Science	–.016 (.033)	–.082 (.052)	–.047 (.072)	–.033 (.050)	.017 (.021)	.032 (.034)	.003 (.022)
SCERTIFD × English	.024 (.031)	–.020 (.050)	–.077 (.067)	.090* (.048)	.040** (.020)	.024 (.024)	–.0003 (.021)
SCERTIFD × Social Studies	.081** (.040)	.001 (.043)	–.040 (.064)	–.020 (.056)	.065*** (.026)	.063*** (.023)	.002 (.021)
R^2	.0096	.0044	.0164	.0053	.0069	.0067	.0057
p-value (H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4$)	.0168	.5714	.1320	.1139	.5129	.6876	.9926

Note: Standard errors, adjusted for school-level clustering, are reported in parentheses. All models include subject fixed effects and the classroom and teacher controls. The p-values refer to tests of the null hypothesis that the four coefficients for a teacher qualification are equal. For definitions of variables, see Table 1.
*Statistically significant at the 10 percent level. **Statistically significant at the 5 percent level. ***Statistically significant at the 1 percent level.

apparent benefit of subject-certified teachers could reflect the bias that would occur if students with a relative propensity to achieve in a particular subject are more likely to be assigned to a teacher who is certified in that subject. Similarly, the presence of unobserved teacher or classroom traits could bias these key inferences. We cannot definitively address these important concerns. However, one weakly suggestive indication that this sort of bias is not problematic is that subject-certified teachers in subjects other than mathematics and social studies are not associated with increased achievement. If a problematic pattern of nonrandom, within-student assignments did occur, we might reasonably expect it to occur in all four subjects.

Furthermore, the robustness of this study's results to conditioning on observed teacher and classroom traits is also consistent with the absence of bias. More specifically, in Table 2, the introduction of controls for other teacher observables (e.g., teacher experience, race, and gender) reduced the estimated effect of subject certification by only a fraction of a standard error. And introducing controls for class size and the percentage of peers who are LEP actually increased the estimated effect of subject certification by a relatively modest amount. To the extent that this pattern of "selection on observables" mirrors that of the selection on unobservables, these results imply that the estimated effects of subject certification are not confounded by unobserved teacher or classroom traits.

As additional evidence regarding another potential source of bias, we examined a simple falsification test that focused on the mathematics and social studies results. Again, the concern with these results is that it may merely reflect the fact that students with a particular propensity for achievement in those subjects are more likely to be assigned to a teacher who is certified in that subject. We examined this possibility by estimating the effect of teachers certified in math and social studies on *science* and *reading* scores, respectively. Specifically, we substituted the standardized science score for the test score observed when a student was with a mathematics teacher. And we substituted the standardized reading score for the social studies score observed when a student was with a social studies teacher. We then evaluated a specification such as that in column (1) of Table 6 using the full data set.

The basis for this test is our conjecture that students with a relative propensity for achievement in mathematics are likely to have a similar propensity for success in science given the complementarities between these two subjects. Similarly, achievement in social studies is likely to be related to achievement in reading.[14] Given this assumption, the existence of a particularly large effect of a math-certified teacher on science scores would suggest the existence of a confounding bias in our main results. Similarly, a large estimated effect of a subject-certified social studies teacher on reading scores would also suggest the existence of bias in the main FD results.

The results with regards to mathematics are not entirely dispositive. The first-difference estimate of the effect of a math-certified teacher on science scores is 0.07 ($p = 0.013$). This estimate is 40 percent smaller than the estimated effect of such teachers on math scores (Table 6). The comparative reduction in this point estimate is consistent with the hypothesis that subject-certified math teachers are genuinely effective at raising achievement. However, there is also some ambiguity since the estimated effect of such teachers on science scores is still large and statistically significant. This result could reflect bias (i.e., more able students being more likely to be assigned to subject-certified math teachers) as well as the spillover benefits from such teachers. The results with respect to social studies are more definitive. This falsification test indicates that a teacher certified in social studies actually has negative (but statistically insignificant) effects on reading achievement. This suggests that the achievement gains associated with assignment to a teacher certified in social studies do not reflect the nonrandom assignment of students with subject-specific propensities for achievement.

In general, these results are consistent with the hypothesis that, at least in social studies and mathematics, eighth-grade students benefit from subject-qualified teachers. However, another interesting dimension of the results reported here is that a subject-qualified teacher is more likely to have a poor opinion of a particular student. Of course, these results may be internally consistent because the manner in which subject-qualified teachers increase student achievement could involve their high expectations for homework and attentiveness. Nonetheless, the pejorative perceptions that a teacher maintains could also harm a student's relative educational opportunities in more informal ways (e.g., recommendations for future placement and the nature of classroom interactions).

However, this pattern of results may suggest another, more important qualification. The discussion and evidence on teacher quality tends to assume that a "highly qualified" teacher is effective for every type of student. However, this assumption may not be valid. In particular, it may be that the students most at risk of academic failure are harmed by dramatic mismatches between their likely outcomes and the relatively high expectations of "in-field" teachers.

To examine this issue further, we assess whether subject-qualified teachers have different effects at different points in the test score distribution.

Specifically, we constructed seven dummy variables for whether a student's standardized test score was greater than or equal to certain values (i.e., –1.5, –1.0, –0.5, 0, 0.5, 1.0, 1.5). We then estimated specifications where these dummy variables were the dependent variables. The results of this application are reported in Table 7.

These results suggest that, at most points in the test score distribution, assignment to a mathematics teacher who is certified in mathematics increases the probability of having a high test score. The sole exception is in the far-left tail of the test score distribution. There, assignment to a subject-certified mathematics teacher significantly reduces the likelihood that a student's test score will be greater than –1.5. We do not observe a similar pattern of response heterogeneity with respect to the social studies results.

Discussion

Recent efforts to ensure that every public school teacher is "highly qualified" have focused on teacher proficiency (i.e., certification and a postsecondary major) in the subject that they teach. But are teachers with these subject-specific qualifications really more effective? The results of our study suggest that, at least at the middle school level, the answer to this question is decidedly mixed.

For example, we found that assignment to a social studies teacher with qualifications in that subject increased test scores by 0.08 standard deviations. And a subject-qualified mathematics teacher with qualifications increased student test scores by 0.12 standard deviations. Increases of these magnitudes are by no means trivial. One particularly relevant point of comparison is the minority achievement gap. Improving the relative academic performance of minority students is one of the most important educational goals in the United States. In the 1999 National Assessment of Educational Progress (NAEP) mathematics exam, the gap between white and Hispanic thirteen-year-olds was approximately 0.74 standard deviations; between white and black students, this gap was 0.98 standard deviations (U.S. Department of Education 2000a, 2000b). This implies that just one year with a subject-certified mathematics teacher in a predominantly minority school would close the achievement gap in that subject by at least 12 percent.

However, our results also suggest that the educational returns to a subject-qualified teacher in areas other than mathematics and social studies are

Table 7
First-Difference (FD) Estimates of the Effects of Subject-Specific Teacher Qualifications at Different Points in the Test Score Distribution and by Subject

	Dependent Variable: Dummy Variable for STEST =						
Independent Variable	–1.5	–1.0	–0.5	0	0.5	1.0	1.5
SCERTIFD × Mathematics	–.027** (.010)	.029* (.015)	.070*** (.020)	.048** (.019)	.040*** (.015)	.025** (.011)	.021*** (.008)
SCERTIFD × Science	.010 (.015)	–.001 (.016)	–.007 (.019)	–.013 (.018)	–.001 (.018)	–.001 (.014)	.008 (.012)
SCERTIFD × English	–.012 (.011)	.026 (.018)	.035* (.018)	.014 (.020)	–.017 (.016)	.001 (.013)	.017* (.010)
SCERTIFD × Social Studies	.006 (.016)	–.001 (.018)	–.0004 (.021)	.028 (.023)	.048** (.023)	.025 (.018)	.014 (.012)
R^2	.0086	.0053	.0061	.0044	.0067	.0042	.0037
Dependent mean	.96	.80	.61	.44	.29	.18	.10
p-value (H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4$	.0922	.3262	.0140	.0939	.0099	.3254	.8028

Note: Standard errors, adjusted for school-level clustering, are reported in parentheses. All models include subject fixed effects and the classroom and teacher controls. The p-values refer to tests of the null hypothesis that the four coefficients for a teacher qualification are equal. For definitions of variables, see Table 1.

*Statistically significant at the 10 percent level. **Statistically significant at the 5 percent level. ***Statistically significant at the 1 percent level.

smaller and statistically indistinguishable from zero. These findings could reflect the possibility that subject proficiency in these areas is less relevant at the middle school level. Alternatively, it could be that subject proficiency does matter in these subjects but that certification and academic majors, as currently regulated, fail to ensure that proficiency. This could occur, for example, if a middle school science teacher had a college degree in only one part of the science curriculum. Regardless, these results raise some doubt about how policy makers have chosen to identify high-quality teachers as well as about the academic consequences of out-of-field teaching at the middle school level.

Our results with respect to educational outcomes other than test scores compound these concerns. Specifically, we found that subject-qualified teachers were not significantly more effective than other teachers at promoting students' engagement and comfort with their subject. Furthermore, we found that subject-qualified teachers were more likely to view their students pejoratively (i.e., as inattentive and not completing homework). Of course, those teacher perceptions may reflect the relatively high expectations of subject-qualified teachers, and those expectations may, on average, promote student achievement. However, our results also suggest that, at least in mathematics, "in-field" teachers actually reduce the achievement of the very weakest students. These results imply that researchers and policy makers should also be aware of the possibility that what makes a teacher effective for a particular type of student may make him or her ineffective for others.

Notes

1. Veteran teachers can establish mastery of their subject matter by meeting their state's "high, objective, and uniform standard of evaluation" (HOUSSE). In most states, teachers can do this by earning a set number of points from a menu of approved activities. Walsh and Snyder (2004) criticize this approach, noting that many of the activities are only loosely related to mastery of subject matter. They also note that eleven states have argued that their existing systems of certification ensure that teachers are proficient in their subjects.

2. For example, Allen (2003) concludes that the research on whether subject knowledge contributes to teacher quality is "spotty and focuses largely on the teaching of mathematics."

3. These studies define teacher quality unrestrictively through teacher fixed effects or value added in models of student achievement.

4. See Wayne and Youngs (2003) and Goldhaber and Anthony (2003) for discussions of this literature. For more recent studies, see Jepsen (2005) and Hanushek et al. (2005).

5. These criteria were that the study control for prior student achievement and socioeconomic background.

6. However, Darling-Hammond, Berry, and Thoreson (2001) criticize this inference, noting, among other things, that there are relatively few teachers with probationary or emergency credentials in National Education Longitudinal Study of 1988 (NELS:88) data. This study does not address the distinction between regular and alternative certification.

7. They also found that students had higher mathematics achievement when with a teacher who answered a specific algebra question correctly.

8. For example, we find that students who are perceived pejoratively by their teachers are less likely to enroll in Advanced Placement courses in later grades.

9. Not surprisingly, the students who do not have classes in both academic subjects appear to have lower levels of achievement. Specifically, a regression indicates that students without two teacher surveys have significantly lower test scores, conditional on school and subject fixed effects.

10. For details on the cognitive tests, see U.S. Department of Education (1991). For several reasons, test scores are unavailable for roughly 4 percent of the 24,599 students who completed questionnaires. For example, some students were absent on the survey day and were only administered the questionnaire during a makeup session. Several participating schools also refused the test component of the study, and test sections were not scored if a student answered fewer than five questions.

11. Prior research suggests that a demographically similar teacher may influence student outcomes through phenomena such as role model effects, stereotype threat, and teacher biases (e.g., Dee 2004, 2005; Hanushek et al. 2005).

12. One way to discriminate between these explanations is to evaluate models that include teacher fixed effects. Such an approach is possible since responding teachers taught some subjects for which they were in field and others for which they were not. However, as a practical matter, there are too few of such teachers in the data to generate much statistical power.

13. Interestingly, the mathematics estimate is almost exactly equal to the effect size that Goldhaber and Brewer (2000, 139) found for twelfth-grade mathematics scores. This suggests that their value-added specification did effectively control for unobserved student determinants of achievement.

14. The correlation coefficient for the math and science scores is 0.74, while the correlation coefficient for the reading and social studies scores is 0.73.

References

Allen, Michael. 2003. *Eight questions on teacher preparation: What does the research say?* Denver, CO: Education Commission of the States.

Ashenfelter, Orley, and Alan Krueger. 1994. Estimating the returns to schooling using a new sample of twins. *American Economic Review* 84:1157-73.

Ashenfelter, Orley, and Cecilia Rouse. 1998. Income, schooling and ability: Evidence from a new sample of identical twins. *Quarterly Journal of Economics* 113:253-84.

Darling-Hammond, Linda. 2002. Research and rhetoric on teacher certification: A response to Teacher Certification Reconsidered. *Education Policy Analysis Archives* 10 (36). Retrieved April 26, 2005, from http://epaa.asu.edu/epaa/v10n36.html

Darling-Hammond, Linda, Barnett Berry, and Amy Thoreson. 2001. Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis* 23:79-86.

Dee, Thomas S. 2004. Teachers, race and student achievement in a randomized experiment. *Review of Economics and Statistics* 86 (1): 195-210.

———. 2005. A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review* 95 (2): 158-65.

Goldhaber, Dan D., and Emily Anthony. 2003. *Indicators of teacher quality*. New York: ERIC Clearinghouse on Urban Education.

Goldhaber, Dan D., and Dominic J. Brewer. 1997a. Evaluating the effect of teacher degree level on educational performance. In *Developments in school finance 1996*, edited by William Fowler, 197-210. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics.

———. 1997b. Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human Resources* 32:505-23.

———. 2000. Does teacher certification matter? High school certification status and student achievement. *Educational Evaluation and Policy Analysis* 22:129-46.

Hanushek, Eric A., John F. Kain, Daniel M. O'Brien, and Steven G. Rivkin. 2005. The market for teacher quality. NBER Working Paper No. 11154.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 1998. Teachers, schools, and academic achievement. NBER Working Paper No. 6691.

Ingels, Steven J., Sameer Y. Abraham, Rosemary Karr, Bruce D. Spencer, and Martin R. Frankel. 1990. *National Education Longitudinal Study of 1988 base year: Student component data file user's manual.* Washington, D.C.: U.S. Department of Education, National Center for Education Statistics.

Ingersoll, Richard M. 1999. The problem of underqualified teachers in American secondary schools. *Educational Researcher* 28 (2): 26-37.

Jepsen, Christopher. 2005. Teacher characteristics and student achievement: Evidence from teacher surveys. *Journal of Urban Economics* 57:302-19.

Jerald, Craig D. 2002. *All talk, no action: Putting an end to out-of-field teaching*. Washington, D.C.: The Education Trust.

Monk, David H., and Jennifer King. 1994. Multi-level teacher resource effects on pupil performance in secondary mathematics and science: The role of teacher subject matter preparation. In *Contemporary policy issues: Choices and consequences in education*, edited by Ronald Ehrenberg. Ithaca, NY: ILR Press.

Rockoff, Jonah E. 2004. The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review* 94 (2): 247-52.

Rouse, Cecilia. 1999. Further estimates of the economic return to schooling from a new sample of twins. *Economics of Education Review* 18:149-57.

Rowan, B., F. S. Chiang, and R. J. Miller. 1997. Using research on employee's performance to study the effects of teachers on students' achievement. *Sociology of Education* 70: 256-84.

U.S. Department of Education. 1991. *Psychometric report for the NELS:88 base year test battery*. NCES 91-468. Washington, D.C.: U.S. Department of Education.

———. 2000a. *NAEP 1999 trends in academic progress: Three decades of student performance*. NCES 2000–469. Washington, D.C.: U.S. Department of Education.

———. 2000b. *Meeting the highly qualified teachers challenge: The secretary's second annual report on teacher quality*. Washington, D.C.: U.S. Department of Education.

U.S. Department of Education, National Center for Education Statistics. 2004. *Qualifications of the public school teacher workforce: Prevalence of out-of-field teaching, 1987–88 to 1999–2000*. NCES 2002–603 Revised. Washington, D.C.: U.S. Department of Education.

Walsh, Kate. 2001. *Teaching certification reconsidered: Stumbling for quality*. Baltimore: The Abell Foundation.

Walsh, Kate, and Emma Snyder. 2004. *Searching the attic: How states are responding to the nation's goal of placing a highly qualified teacher in every classroom*. Washington, D.C.: National Council on Teacher Quality.

Wayne, Andrew J., and Peter Youngs. 2003. Teacher characteristics and student achievement gains: A review. *Review of Educational Research* 73 (1): 89-12.

Thomas S. Dee is an associate professor in the Department of Economics at Swarthmore College and a Faculty Research Fellow with the National Bureau of Economic Research (NBER).

Sarah R. Cohodes is a research assistant in the Education Policy Center at the Urban Institute. She received her B.A. from Swarthmore College in 2005 with a major in economics and a minor in educational studies.