ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Economics of Education Review ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Economics of Education Review

www.elsevier.com/locate/econedurev

Does teacher testing raise teacher quality? Evidence from state certification requirements

Joshua D. Angrist[a,*], Jonathan Guryan[b]

[a]*Department of Economics, MIT, E52-353, 50 Memorial Drive, Cambridge, MA 02142-1347, USA and NBER*
[b]*University of Chicago Graduate School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA and NBER*

Received 14 February 2006; accepted 23 March 2007

Abstract

The education reform movement includes efforts to raise teacher quality through stricter certification and licensing provisions. Most US states now require public school teachers to pass a standardized test such as the Praxis. Although any barrier to entry is likely to raise wages in the affected occupation, the theoretical effects of such requirements on teacher quality are ambiguous. Teacher testing places a floor on whatever skills are measured by the required test, but testing is also costly for applicants. These costs shift teacher supply to the left and may be especially likely to deter high-quality applicants from teaching in public schools. Moreover, test requirements may disqualify some applicants that schools would otherwise want to hire. We use the Schools and Staffing Survey to estimate the effect of state teacher testing requirements on teacher wages and teacher quality as measured by educational background. The results suggest that state-mandated teacher testing is associated with increases in teacher wages, though we find no evidence of a corresponding increase in quality. Consistent with the fact that Hispanics have marked lower licensure scores than non-Hispanic Whites or Blacks, testing appears to reduce the fraction of new teachers who are Hispanic.

JEL classification: I20; I21; I28; J45

Keywords: Teachers; Testing; Certification; Occupational licensure; Teacher quality; Teacher salaries

1. Introduction

Economists, educators, and policymakers generally agree that better teachers are likely to lead to more effective schools. The question of how to attract better teachers remains open. A natural economic solution is to raise teacher pay. The evidence on the relationship between salaries and measures of teacher quality or performance has been mixed (Figlio, 2002; Hanushek, Kain, & Rivkin, 1999; Murnane, Singer, Willett, Kemple, & Olsen, 1991). In spite of occasional efforts to boost pay, teacher aptitude as measured by standardized test scores has fallen since around 1960 (see, e.g., Corcoran, Evans, & Schwab, 2004).

Beginning in the 1960s, some states began testing prospective teachers in a direct effort to ensure teachers meet minimum standards for basic skills

*Corresponding author. Tel.: +1 617 253 8909.
E-mail addresses: Angrist@mit.edu (J.D. Angrist), Jonathan.Guryan@chicagoGSB.edu (J. Guryan).

0272-7757/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.econedurev.2007.03.002

and subject knowledge.[1] By 1999, 41 states required applicants to pass some sort of standardized certification test such as the National Teacher Examination or Praxis examinations published by the Educational Testing Service (ETS).[2] Although there is some cross-state reciprocity in the form of probationary and provisional licensing, states relying on tests for certification typically require newly employed teachers to pass their own tests even if they are licensed in other states.

As a theoretical matter, the impact of teacher testing is ambiguous. Test requirements may establish a minimum achievement standard, as their proponents argue, but certification requirements may also deter high-quality applicants from teaching in the public schools. Moreover, stricter certification procedures raise barriers to entry that increase labor costs, and may be seen as especially costly by the most experienced teachers or teachers with attractive employment options in other fields.

In this paper, we estimate the impact of state-mandated certification tests on teacher quality and teacher salaries. Data for our study come from the Schools and Staffing Survey (SASS), a nationally representative repeated cross-section of teachers and schools, initially conducted during the 1987–1988 school year. This sample coverage is useful since testing requirements have grown most sharply in recent years. In addition to providing information on teacher salaries, the 1987–1988, 1993–1994, and 1999–2000 waves of the SASS include measures of teacher educational background that we take to be proxy measures of teacher quality.[3] Our analysis starts by showing that state provisions on teacher testing were binding in the sense that teachers were indeed more likely to be tested after states introduced a testing requirement. Consistent with the notion that certification requirements establish barriers to entry, we find that teacher testing is associated with increases in teacher salaries. Despite the corresponding increase in salaries, however, we find no evidence that teachers hired in state testing regimes are more likely to be drawn from more selective colleges or to teach material studied in college or graduate school. We also find that testing is associated with reduced Hispanic representation among new teachers, perhaps because Hispanics are less likely than other groups to pass the main licensure test.

2. Background and context

A 1986 report of the Carnegie Task Force on Education and a follow-up report released in 1996 called for the introduction of more centralized systems of certification for public school teachers. A policy of stricter and more centralized teacher licensing has also received support from the National Education Association and groups promoting education reform (Ballou & Podgursky, 2000). Proposed licensing systems typically involve the accreditation of education programs, longer apprenticeships, and teacher testing. Proponents of teacher licensing point to the spread of medical licensure in the early 20th century as evidence that licensing raises professional standards. On the other hand, economists have long warned that licensing and certification are potentially cost-raising barriers to entry (e.g., Friedman & Kuznets, 1945). Moreover, there is little hard evidence for any consumer benefits of mandatory occupational licensing, including in medicine. In this paper, we attempt to estimate the impact of what is perhaps the simplest component of teacher licensing provisions, a requirement that teachers pass a certification test that can be seen as analogous to medical boards and legal bar exams.

[1] Testing of teachers is not a new phenomenon. Teachers were tested in basic subjects in many states in the 19th century and at the beginning of the 20th century. However, in most states the tests were graded and certificates were issued at the county level. In the early part of the 20th century a number of states began more widespread use of testing for certification. Soon thereafter World War II led to a decrease in teacher supply and a subsequent increase in hiring of teachers with alternative certification. As a result, most states had discontinued the use of required teacher testing by the end of the war.

[2] Since 1998, the ETS National Teachers Examination, widely used to certify Education School graduates for work as teachers, has been known as the Praxis II, and is part of a series that includes Praxis I, also known as the Pre-Professional Skills test (PPST) which is used to screen applicants to Education Schools, and a series of classroom performance assessments known as Praxis III. Many states (e.g., Minnesota as of September 2001) require both Praxis I and Praxis II. As of this writing, sample Praxis content is available at http://www.ets.org/praxis/download.html. The Praxis examinations consist of dozens of subtests. Each state selects their own credentialing requirements. Some states, such as California, require a combination of Praxis tests and locally developed tests. Others, such as Massachusetts, rely on a locally developed exam only.

[3] Although imperfect, measures of the selectivity of teachers' undegraduate institution is often used in studies of teacher quality (see, e.g., Bacolod, 2007; Figlio, 1997; Hoxby & Leigh, 2004). Our companion paper (Angrist & Guryan, 2004) shows that institutional selectivity is a good predictor of individual teacher aptitude.

Please cite this article as: Angrist, J. D., & Guryan, J. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review* (2007), doi:10.1016/j.econedurev.2007.03.002

In a recent survey of research on occupational licensing, Kleiner (2000) observes that more American workers are affected by licensing requirements than belong to unions or are covered by the minimum wage. Yet there are remarkably few studies of the impact of licensing on wages or productivity. Standard economic arguments suggest licensing provisions are likely to affect economic outcomes through a number of channels. First, occupational licensing may provide a signal of worker quality and help to maintain quality standards when information about quality is imperfect. Indeed, this is the stated rationale for government-imposed licensing requirements. As Kleiner (2000) notes, however, the evidence of consumer benefits from most licensing requirements is thin or non-existent. In addition, *mandatory* licensing requirements impose a barrier to occupational entry that is likely to increase wages in the licensed occupation.[4]

One of the few previous attempts to estimate the effect of teacher licensing requirements is a study by Berger and Toma (1994), who find that SAT scores are lower in states that require teachers to have a Master's degree. Berger and Toma hypothesize that this negative relationship may be evidence of a supply response by prospective teachers who view the education requirement as costly, particularly so for talented teachers with better alternatives. Also consistent with this entry-barriers story, Hanushek and Pace (1995) find that state requirements for courses and tests significantly lower the probability prospective teachers complete training, again using cross-state variation.

A related study of licensing requirements, by Kleiner and Petree (1988), links state licensing requirements with average teacher pay, pupils' SAT and ACT scores, and high school graduation rates. Their results show no clear relationship between licensing and pupil achievement or teacher pay, though there is a robust negative association between licensing and pupil-teacher ratios. The authors attribute these ambiguous results on licensing to the weak licensing provisions in force during their sample period. The recent strengthening of state teacher licensing provisions may provide stronger evidence on licensing effects. Another related study is Goldhaber and Brewer (2000), who link student achievement with state teacher licensing and testing requirements. Their analysis does not exploit changes in state provisions over time, and the effects of testing enter only as interactions with other licensing provisions.

Most studies of the economic consequences of occupational licensing look at the medical and dental professions. In a study of dentistry, for example, Kleiner and Kudrle (2000) found that people in states that strictly regulated entry to dentistry had dental health no better than those in states with less regulation. Dental regulation, however, does appear to increase both the hourly earnings of dentists and consumers' cost of dental care.[5] Similarly, in a recent study of immigrant physicians in Israel, Kugler and Sauer (2003) found that immigrant physicians who obtained a license to practice medicine in Israel had sharply higher earnings than those who did not. At the same time, a comparison of OLS and instrumental variables (IV) estimates of the effect of licensing suggests that doctors who obtain licenses and end up practicing medicine have lower earnings potential than those who do not. It should be noted, however, that teachers differ from medical professionals in that they are more likely to work in the public sector. Regulation may more effectively reveal worker quality in the absence of the market forces at work in the private sector.

3. Theoretical framework

Although the theoretical impact of teacher testing on wages seems clear cut, the effect of testing on quality is less so. The policy objective that motivates teacher testing, as with other worker screening devices, is to identify and hire those most qualified to teach. In practice, however, effective testing strategies are hard to design since tests are noisy predictors of worker quality. Moreover, testing is costly for employers and employees. Teacher supply therefore shifts in the face of testing to reflect the time and effort job applicants expend in being tested. Finally, risk-averse workers should see employment opportunities that are contingent on

[4]The literature on occupational regulation distinguishes between mandatory licensing such as required of medical professionals and voluntary certification such as obtained by auto mechanics.

[5]Dentistry appears to be the most widely studied occupation in research on licensing. See Kleiner and Kudrle (2000) for references to earlier work on dentists. Kleiner (2000) also compares wages in licensed occupations with wages in unlicensed occupations requiring approximately the same level of education and training.

Please cite this article as: Angrist, J. D., & Guryan, J. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review* (2007), doi:10.1016/j.econedurev.2007.03.002

stochastic test results as less attractive than unconditional offers.

A large theoretical literature looks at the impact of worker screening mechanisms on wages and job assignments. We use basic elements of the Guasch and Weiss (1980, 1981) worker screening model to discuss the possible implications of standardized testing for teacher quality.[6] Because school districts are not necessarily profit maximizers or even cost minimizers, we focus on the impact of test-based hiring standards on teachers' labor supply, as opposed to the more complex question of how worker testing affects equilibria in competitive markets.

Suppose an applicant for a teaching job can earn an alternative wage, w_i, and teachers are paid a fixed wage w, which might be set by collective bargaining. Job applicants must be tested, a process which they view as costing an amount c. We can think of c as a monetary cost or as the cost of time and effort directed towards preparation and completion of the test. More generally, testing might involve a probationary period, in which case any wage reduction during the probation period is part of the testing cost.

Worker i passes the test with probability p_i. We presume the test has some screening value, so that p_i and w_i are positively correlated. In other words, higher quality applicants, as measured by outside earnings potential, are more likely to pass the test. Assuming teachers maximize expected utility with von Neumann–Morgenstern utility of income $U(X)$, applicant i must be offered a wage that satisfies

$$p_i U(w-c) + (1-p_i)U(w_i - c) \geqslant U(w_i), \tag{1}$$

if he or she is to find it worth applying for a test-contingent job. Clearly as c increases, the wages offered teachers, w, must be higher to obtain an applicant pool of the same quality. This is the entry-barrier effect on wages; positive c reduces the supply of applicants, holding fixed the underlying distribution of quality as measured by w_i. Note also that this deterrent effect is larger with risk averse than with risk-neutral applicants. Risk-neutral applicants require only that $p_i(w-w_i) \geqslant c$.

We highlight three mechanisms by which testing affects the average quality of newly hired teachers. Continuing to think of quality as indexed by the alternate wage, w_i, suppose that school boards, who do not observe w_i, would like to select applicants with $w_i \geqslant \bar{w}$. For simplicity, suppose also that applicants are risk neutral and that the certification test is a perfect screen of teacher ability. In other words, the test is designed so that $p_i = 1$ if $w_i \geqslant \bar{w}$ and is zero otherwise. Then average teacher quality in the testing regime is

$$E[w_i | w - c > w_i \geqslant \bar{w}]. \tag{2}$$

This average can be compared with the average teacher quality in a no-testing regime, $E[w_i | w > w_i]$. This comparison is depicted in Fig. 1 for uniformly distributed w_i. The shaded portion of the top panel of the figure represents the set of applicants who are eligible and choose to teach in a world without teacher testing requirements. The shaded portion of the middle panel represents the set of applicants who are eligible and who choose to teach in a world with a testing requirement and a test that measures ability perfectly. The imposition of a lower bound, $\bar{w}$, clearly increases quality. We label this mechanism the *information effect* because the test provides schools with information about applicants' ability that allows them to avoid hiring some low-ability teachers.[7]

If testing is viewed as costly, some applicants will choose not to teach to avoid having to study for and take the test. Because the cost is common to all individuals in this model, applicants on the margin between teaching and an alternative occupation are the highest-quality teachers. In other words, these applicants have the best outside options. As c rises, more of these marginal applicants are discouraged from teaching (the *discouragement effect* in the figure) and average quality may decline. This decline

[6] See also Lelande (1979). For more recent and more elaborate models along these lines, see, e.g., Wang (1997) and Wang and Weiss (1998). Goldhaber's (2004) discussion of the theoretical impact of licensure considers essentially the same forces as captured by our model, below. In particular, Goldhaber notes that licensure may dissuade some more academically talented teachers (which we read as suggested a higher outside wage) from teaching.

[7] A more elaborate model would allow for discrimination against teachers who come from more selective schools, as suggested by Ballou (1996). This could be captured here by allowing $\bar{w}$ to be a function of teachers' institution type. This mechanism seems to be in the spirit of our discussion since it would encourage teachers in a testing regime to attend lower-quality schools that focus on certification as opposed to more rigorous subject-oriented programs. In addition, it bears emphasizing that our empirical work looks at working teachers and not applicants; the quality of working teachers is a consequence of forces operating in both the selection and discrimination channels, at the certification and hiring stages.

Please cite this article as: Angrist, J. D., & Guryan, J. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review* (2007), doi:10.1016/j.econedurev.2007.03.002

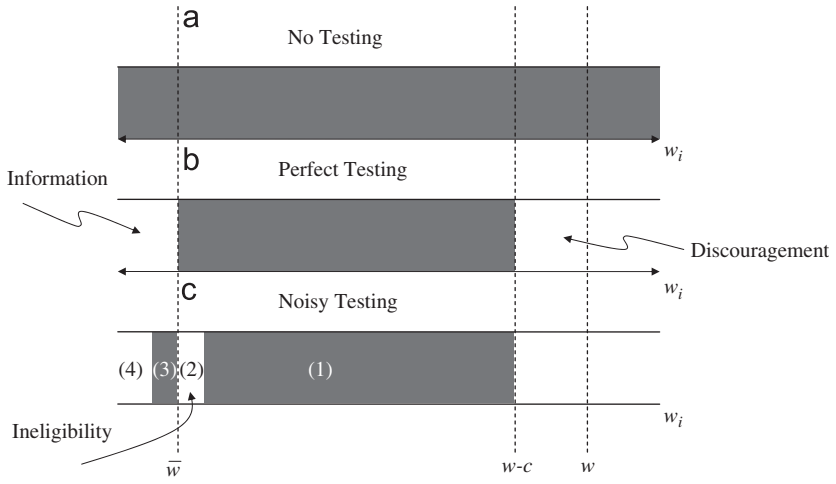

Fig. 1. The wage and ability distribution of teachers under alternative testing regimes. Note: the shaded region represents the applicants who choose and are hired to teach under three alternative testing regimes: (a) no testing requirement, (b) a required test that measures teachers' ability perfectly, and (c) a required test that measures teachers' ability imperfectly.

in average quality occurs in spite of the fact that the lower quantiles of the quality distribution will have increased if testing is effective. More generally, however, the discouragement effect is ambiguous since c may be related to opportunity costs. If, for example, c reflects time spent preparing for the test, costs are increasing in w_i, and the discouragement effect is clearly negative. If, however, costs are decreasing in ability, as might be true if test preparation is easier for the more able, the marginal applicant will have poor outside options.

The third mechanism through which testing may affect the average quality of newly hired teachers comes into play when, as seems likely, tests measure applicant ability imperfectly. Specifically, suppose applicants pass the test if $w_i + \eta_i > \bar{w}$, where η_i is a mean-zero random error uncorrelated with w_i. Applicants can now be classified in the groups listed below, numbered as in Fig. 1:

(1) $w_i > \bar{w}$ and $w_i + \eta_i > \bar{w}$ (pass),
(2) $w_i > \bar{w}$ and $w_i + \eta_i < \bar{w}$ (fail),
(3) $w_i < \bar{w}$ and $w_i + \eta_i > \bar{w}$ (pass),
(4) $w_i < \bar{w}$ and $w_i + \eta_i < \bar{w}$ (fail).

School districts prefer applicants in groups (1) and (2) but regulation requires them to hire applicants in groups (1) and (3). Since the average ability of applicants in group (3) is less than that in group (2), noise in testing reduces the average ability of teachers hired.[8] We call this an *ineligibility effect* since some qualified applicants are made ineligible for teaching jobs by the testing requirement. Angrist and Guryan (2004) report statistics showing that the Praxis II strongly favors applicants who attended Schools of Education, in spite of the fact that these students are academically weak. This finding suggests that teacher licensure tests may measure non-productive attributes of applicants, or in other words that the ineligibility effect may be substantial.[9]

4. Data and descriptive statistics

4.1. The Schools and Staffing Survey, teacher quality measures, and state testing laws

In what follows, we estimate the effects of testing requirements on both teacher salaries and measures

[8]This is assuming schools hire roughly the same number of teachers in both regimes. In the longer run, schools may adjust hiring in light of the new quality mix if this is revealed ex post.

[9]The theoretical discussion in this section assumes that the cost of testing is common across individuals. The cost of testing could be either negatively or positively correlated with quality. The former might be the case if high-ability teachers suffered less psychological disutility from test-taking. The latter might be the case if the test was focused on specialized material less likely to be studied by high-ability individuals, as suggested by the results in Angrist and Guryan (2004). If high-quality applicants experience lower testing costs, the deterrent effect on average quality is of course muted.

Please cite this article as: Angrist, J. D., & Guryan, J. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review* (2007), doi:10.1016/j.econedurev.2007.03.002

of teacher quality and characteristics. Our unit of analysis is the school district, because salaries are typically set at the district level. The data for this study come chiefly from two components of the SASS. The first round of the SASS was conducted in the 1987–1988 school year, followed by rounds in 1990–1991, 1993–1994 and 1999–2000. District information on salaries and the use of testing come from a survey of school district administrators. Characteristics of teachers and measures of teacher quality are derived from a survey of teachers sampled from these districts.[10]

Our first two measures of teacher quality are features of teachers' undergraduate educational institution or college. In particular, we use information on the average 1983 SAT score of entering freshmen at teachers' undergraduate institutions, as compiled by the Higher Education Research Institute (Astin, Green, Korn, & Maier, 1983). Carnegie classifications are also used to create an indicator for whether the teacher attended a research university or a liberal arts college.[11] Although these quality measures are not as detailed as we would like, the average SAT score at teachers' undergraduate institutions is a frequently used measure of new teacher quality (see, e.g., Figlio, 2002). Following the discussion of quality effects, we discuss results from a match of individual SAT and Praxis scores which shed some light on the question of whether institution-based quality measures are informative.

The other outcomes examined here include an indicator for whether the teacher majored in the subject she teaches, an indicator for whether the teacher has an alternative (i.e., non-standard) certification status on her current job, and the demographic characteristics of teachers. Subject major is relevant because school district administrators often claim to be interested in attracting math and science teachers who are trained in their subjects. The alternative certification variable may also be taken as a measure of teacher qualifications. In addition, this variable provides an indicator of concurrent changes in hiring practices.

For the purposes of our statistical analysis, individual teacher information was aggregated to the district level. Because testing should affect new teachers the most, we created aggregates for two sets of teachers in addition to the full sample, those hired in the past year (first-year teachers) and those with 3 years or less teaching experience (inexperienced teachers). Examples of the resulting aggregated outcome variables are the fraction of first-year teachers in each district who were Hispanic, and the fraction of inexperienced teachers in each district who had an alternative certification. Outcomes for first-year and inexperienced teachers were computed for subsamples of districts with first-year or inexperienced teachers. The samples used here exclude districts with less than 50 pupils (i.e., below about the first percentile in the district size distribution). The samples include public schools only and omit charter schools. Finally, information on state testing requirements was obtained from published summaries. For additional details on the construction of variables and our extract, and a list of references on state testing regulations, see Appendix A.

4.2. Descriptive statistics

Table 1 reports descriptive statistics by survey year. Each round of the survey contributes almost 5000 districts to the total sample. The typical district has about 3000 pupils and 160 teachers. The table also shows the proportion of districts with inexperienced teachers (hired in the 3 years preceding the survey) and the proportion of districts with teachers hired in the past year. Over 40% of districts have inexperienced teachers and almost 20% have teachers hired in the past year.

The first outcome variable used to measure the impact of teacher testing is wages. Although the theoretical discussion suggests the effect of testing on the distribution of teachers' alternate wages (i.e., their quality) is ambiguous, the effect on teachers' own pay is likely to be positive since testing restricts supply (note that $w-c$ has to exceed the quality threshold). The SASS reports the wages paid to teachers in each district by schooling and experience level. In particular, wages are reported separately for teachers with a Bachelor's degree (B.A.), with a Master's degree (M.A.), and with a Master's degree plus 20 or more years of experience. Table 1 shows

[10]To account for the stratified sampling frame used by the SASS, throughout we weight responses from the teacher survey using the teacher weight, and responses from the district survey using district weights. The variation of interest in our analysis is at the state-by-year level, and properly weighted, each wave of the SASS is designed to be representative at the state level.

[11]The three Carnegie classifications covered by this definition (Research I and II, and Liberal Arts I) are the three most selective non-professional classes.

Please cite this article as: Angrist, J. D., & Guryan, J. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review* (2007), doi:10.1016/j.econedurev.2007.03.002

Table 1
Means of selected variables

	All districts				Sample with inexperienced teachers				Sample with first-year teachers			
	87–88	90–91	93–94	99–00	87–88	90–91	93–94	99–00	87–88	90–91	93–94	99–00
Unweighted count	4790	4831	4920	4644	2073	2277	2390	2374	930	1068	1166	1138
District characteristics												
Enrollment	2751	2826	2976	3402	4069	3943	4372	5257	6257	5405	6365	7855
Full-time equivalent teachers	158.1	159.3	159.0	211.0	227.6	218.1	227.8	320.2	343.3	293.5	327.6	470.1
Frac. *w*/inexperienced teachers	0.401	0.426	0.427	0.446	1	1	1	1	0	0	0	0
Frac. *w*/first-year teachers	0.164	0.194	0.186	0.192	0	0	0	0	1	1	1	1
Teacher wage and quality measures												
Salary: B.A.	25,344	25,481	25,320	25,898	25,071	25,009	24,883	26,074	25,076	24,680	24,724	26,232
Salary: M.A.	27,683	27,765	27,649	28,303	27,327	27,265	27,150	28,489	27,335	26,965	27,020	28,673
Salary: M.A. + 20 years experience	41,939	42,529	42,950	44,108	40,992	41,250	41,478	43,948	41,145	40,415	41,213	43,777
Average SAT	907.4	–	909.9	905.2	905.7	–	906.5	907.5	912.6	–	908.8	910.5
Attended research univ. or liberal arts coll.	0.218	–	0.227	0.210	0.229	–	0.220	0.214	0.264	–	0.256	0.213
Majored in teaching subject	0.067	0.065	0.074	0.077	0.075	0.079	0.095	0.109	0.075	0.079	0.092	0.123
Frac. *w*/alternative certification	0.104	0.075	0.084	0.116	0.306	0.316	0.287	0.402	0.369	0.382	0.377	0.516
Other teacher characteristics												
Fraction female	0.693	0.677	0.662	0.695	0.720	0.666	0.645	0.666	0.700	0.647	0.634	0.676
Fraction black	0.026	0.029	0.025	0.028	0.024	0.018	0.027	0.036	0.024	0.022	0.047	0.042
Fraction Hispanic	0.015	0.018	0.020	0.029	0.025	0.027	0.044	0.043	0.021	0.048	0.056	0.050

Note: The table reports district weighted means. District-weighted means are reported. Inexperienced teachers are defined as teachers with less than 4 years teaching experience. All salaries are reported in 1999 dollars. Average SAT, Fraction of Teachers who Attended Carnegie I Schools, and Fraction of Teachers with Alternative Certification are measured for all teachers, inexperienced teachers or first-year teachers. For all other variables, district means are estimated using all schools or using the sample of schools that employ either inexperienced or first-year teachers.

teacher wages for those with a B.A. were equal to 25,000–26,000 (in 1999 dollars) in the sample period. Wages went up between 1993–1994 and 1999–2000. Wages were also about 10% higher for those with an M.A., and much higher for experienced teachers with a Master's degree.

The quality analysis begins by looking at average SAT scores in teachers' undergraduate institution and whether the institution was coded as a research university or liberal-arts college. Note that the SAT and Carnegie variables cannot be linked to the 1990–1991 SASS because this round did not identify teachers' undergraduate institutions. The college-based quality variables generally show fairly stable quality over the sample period. In contrast, there was an increase in the proportion of inexperienced and new teachers with alternative certification, and an increase in the proportion of teachers who have a degree in the subject they teach. Finally, the table provides descriptive information for two additional quality variables, the proportion of teachers with a degree in the subject they teach and the proportion with alternative state certification.

4.3. Test prevalence and requirements

The proportion of districts subject to state-mandated Basic Skills testing increased from just over 40% in 1987–1988 to 70% in 1993–1994. This can be seen in the first row of Table 2, which reports the prevalence of State testing requirements based on our match of information for each state to the SASS. Although the number of districts requiring Basic Skills testing fell slightly between 1993–1994 and 1999–2000, the number of states with a testing mandate continued to increase and reached two-thirds.[12] Fewer states required a Subject test than required a test of Basic Skills in 1987, but this

[12]A few large states reversed their Basic Skills testing requirements so the proportion of districts requiring testing dipped between 1993 and 1999.

Please cite this article as: Angrist, J. D., & Guryan, J. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review* (2007), doi:10.1016/j.econedurev.2007.03.002

Table 2
Testing requirements and prevalence

	Proportion of districts				Proportion of states			
	87–88	90–91	93–94	99–00	87–88	90–91	93–94	99–00
State requirements								
Requires basic skills test	0.429	0.622	0.697	0.648	0.431	0.588	0.627	0.667
Requires subject test	0.336	0.365	0.538	0.674	0.373	0.373	0.529	0.608
Requires any test	0.540	0.693	0.736	0.820	0.529	0.647	0.667	0.803
District response in the SASS								
Requires basic skills test	0.361	0.425	0.493	0.646				
Requires subject test	0.243	0.341	0.394	0.552				
Requires any test	0.379	0.452	0.514	0.669				
District response in the SASS with state requirements imposed								
Requires basic skills test	0.554	0.726	0.778	0.827				
Requires subject test	0.494	0.613	0.703	0.799				
Requires any test	0.612	0.744	0.802	0.880				

Note: Left columns of the table report weighted fractions of districts. The top panel reports fraction (of states or districts) that require new teachers to pass basic skills and/or subject tests to be licensed. The middle panel reports the fraction of districts that report in the SASS that they require teaching candidates to have passed basic skills and/or subject tests. The bottom panel reports the fraction of districts that either report in the SASS that they require teaching candidates to have passed basic skills and/or subject tests or are in a state that requires that they do so.

requirement also saw a marked and steady increase, so that the proportion requiring Subject and Basic Skills tests were about equal by 1999. By 1999, over 80% of districts faced some kind of state-mandated test.

In addition to using published *state* testing requirements, we took information from SASS variables containing district administrators' reports on the use of testing in each *district*. This information comes from the following two questions:

> Do you require or use information on whether an applicant passed a STATE test of basic skills?
> Do you require or use information on whether an applicant passed a STATE test of subject knowledge?

SASS respondents (i.e., the officials who completed the SASS on behalf of their district) answered with: Require; Use, but not require; Do not use. Rows 4, 5, and 6, in Table 2 show the proportion of districts that report they require State tests. Surprisingly, this proportion is below the proportion of districts who were subject to a state-mandated test requirement. There are, however, a number of districts that report test requirements that exceed the state requirements. These districts have about 2% higher per-capita income and about 11% smaller district enrollment, but have about one percentage point larger minority enrollment.

The difference between state testing requirements and districts' reported testing practices seems most likely to be due to inaccurate responses and misunderstandings of the SASS questions related to local procedures for applicant screening. We substantiated this hypothesis by surveying a sample of districts ourselves. In particular, we administered the applicant-qualifications portion of the 1999–2000 SASS to 211 "dissonant districts", defined as those where SASS responses to questions on testing conflicted with state requirements.[13] In response to our survey, only 13% of districts reported they neither use nor require a state test of basic skills while only 17% reported they neither use nor require a state test of subject knowledge. On further inquiry with some districts, we discovered that where tests are required, districts may waive this in a tight labor market, but typically still hope

[13]A district was determined to be dissonant if the response to both the basic skills and subject test questions indicated no test requirement and no test use while the state required testing. We sampled up to 10 dissonant districts in any state with dissonant districts. Of the 322 districts sampled, we obtained responses from 211 districts for a response rate of 66%. The original sample contained seven vocational districts and one charter district, so these factors cannot account for reporting conflicts.

Please cite this article as: Angrist, J. D., & Guryan, J. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review* (2007), doi:10.1016/j.econedurev.2007.03.002

to "use the test". An occasional source of confusion, however, had to do with the definition of a "state requirement" or a "state test". For example, the state may require ETS's Praxis exam, which is not strictly speaking a state test along the lines of, say, the test independently developed and used by districts in the state of Massachusetts. In any case, the overwhelming majority of districts in our sample appeared to be trying to follow state testing mandates.

The last three rows of Table 2 show the proportion of districts using tests based on a variable constructed by recoding the response to SASS test-use questions to be consistent with state requirements (e.g., districts who report not using a subject test in a state that requires subject testing were recoded as using a subject test). Not surprisingly, these recoded variables have higher means than the raw SASS responses in the second three rows. They also show a consistent pattern of increasing test use over time. The impact of state requirements on testing is gauged below on the basis of these recoded variables.

4.4. Effects of state testing requirements on testing

The impact of state testing requirements on test use is summarized by regressing dummies for test use on dummies for state mandates, along with state and year main effects, dummies for urban, suburban, and rural districts, district enrollment, district fraction minority enrollment, and a quadratic in the state unemployment rate. In particular, Table 3 reports estimates of the coefficients α_1 and α_2 on basic skills and subject test mandate dummies, b_{jt} and s_{jt}, in the equation

$$y_{djt} = \mu_j + \delta_t + X'_{djt}\beta + \alpha_1 b_{jt} + \alpha_2 s_{jt} + \varepsilon_{djt}, \quad (3)$$

where y_{djt} is an indicator for test requirements in district d in state j in year t, μ_j and δ_t are state and year effects, and X_{djt} is the vector of other covariates. Some of the models combine the separate basic skills and subject dummies into a single dummy for "any test".

Estimates of Eq. (3) can be seen as a calibration exercise telling us the difference between rates of test use with and without state requirements. Table 3

Table 3
First-stage estimates with state and year fixed effects

	District requires basic skills test				District requires subject test			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Full sample								
Basic skills test law	0.516 (0.050)		0.487 (0.053)		0.335 (0.061)		0.143 (0.044)	
Subject test law		0.297 (0.070)	0.081 (0.049)			0.588 (0.056)	0.525 (0.060)	
Any test law				0.521 (0.060)				0.549 (0.054)
R^2	0.736	0.638	0.738	0.719	0.568	0.645	0.653	0.634
N	18,288							
Inexperienced teachers								
Basic skills test law	0.476 (0.048)		0.454 (0.050)		0.275 (0.066)		0.078 (0.047)	
Subject test law		0.258 (0.073)	0.063 (0.055)			0.604 (0.064)	0.570 (0.070)	
Any test law				0.482 (0.066)				0.533 (0.069)
R^2	0.719	0.632	0.721	0.698	0.549	0.648	0.650	0.617
N	6476							
First-year teachers								
Basic skills test law	0.449 (0.046)		0.433 (0.048)		0.205 (0.062)		0.048 (0.042)	
Subject test law		0.221 (0.075)	0.057 (0.058)			0.566 (0.060)	0.548 (0.063)	
Any test law				0.452 (0.065)				0.466 (0.074)
R^2	0.706	0.622	0.707	0.682	0.539	0.631	0.631	0.595
N	3008							

Note: Inexperienced teachers are defined as teachers with less than 4 years teaching experience. Dependent variable is an indicator for whether the districts either report in the SASS that they require teaching candidates to have passed a basic skills or subject test or are in a state that requires that they do so. Reported coefficients are estimated from an OLS regression on state testing requirement dummy variables and a set of controls. Controls include state and year fixed effects, city, suburb and rural fixed effects, a quadratic in the state unemployment rate, district enrollment, and district fraction minority enrollment. All regressions are weighted using district sampling weights. Standard errors corrected for state-by-year correlation in the error term are reported in parentheses.

Please cite this article as: Angrist, J. D., & Guryan, J. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review* (2007), doi:10.1016/j.econedurev.2007.03.002

shows that state-required testing of teachers' basic skills increases the likelihood of basic skills testing in school districts by about 50%. As can be seen in column 2, subject test requirements are also correlated with the use of basic skills tests, but column 3 shows that when both dummies are included, the basic skills requirement dominates. The reverse pattern appears in columns 5–7 for models with the use of subject tests on the left hand side. The imposition of any test requirement also increases the likelihood of testing by about 50%. Moreover, as the lower two-thirds of Table 3 shows, these effects are similar when the sample of districts is limited to those that have new or inexperienced teachers.

5. Results

5.1. Effects on wages

State testing requirements are associated with slightly higher wages. This can be seen in Table 4a, which reports estimates of Eq. (3) for models with the log of teacher salaries on the left hand side. Many of the estimated salary effects are significant. For example, column 1 shows that the salaries of teachers with a B.A. degree are about 2.4% higher when states require a test of basic skills, an effect estimated with a standard error of 0.9%.[14] Subject test requirements also appear to be associated with higher wages, though the estimated effects of testing requirements are not significant when both the subject and basic skill testing variables are entered at the same time.

Most new teachers have a B.A. As the estimates in columns 5–8 and 9–12 show, however, state testing requirements are also associated with higher wages for teachers with an M.A. and for experienced teachers with an M.A. Since teachers with more advanced degrees and more experience are less likely to have been hired recently, these effects may reflect the maintenance of relative wages by shifting the entire pay scale in response to testing requirements.

An alternative interpretation of the increase in wages for more educated or more experienced teachers is that these effects reflect some sort of omitted variables bias. The possibility of omitted variables bias is also raised by the fact that the estimated wage effects are not markedly larger when the sample is limited to districts with inexperienced and new-teachers. Because testing requirements are time-varying state-level variables, the most likely source of bias is some sort of state-specific trend in teacher wages in states that adopt testing requirements.[15] Therefore, as a check on this, we re-estimated the wage equations using a model that adds state-specific linear trends to specification (3). This controls for the fact that teacher wages may be increasing due to secular trends that contribute to the demand for higher entry barriers. One possibility, for example, is that unions raise entry barriers in good times. On the other hand, our survey suggests districts and therefore perhaps also states want to weaken formal requirements when teachers are hard to find.[16]

The results of estimating Eq. (3) with state-specific linear trends, reported in Table 4b, show even stronger wage effects than appear in Table 4a. For example, the imposition of a state test of basic skills is associated with roughly 3.4% higher wages for teachers with a B.A., a precisely estimated effect. Moreover, the effects of requiring tests of basic skills and subject matter remain significant when entered jointly. The fact that estimates with state trends are larger than those without is consistent with the view that testing provisions are weakened in a strong economy. Perhaps most importantly, the pattern of effects is now more consistent with a causal interpretation that attributes higher wages to the impact of state testing regulations.

[14]The regression estimates reported here and elsewhere in the paper were weighted using district sampling weights. Standard errors are corrected for state-year clustering. Standard errors clustered by state only are similar.

[15]The timing of the adoption of state testing requirements seems to be correlated with state-level socio-economic status and educational outcomes. States with lower NAEP scores, fewer high school graduates in 1928, a lower Putnam Social Capital Index, and more poorly ranked health care were more likely to require teacher testing by 1988. An analysis of trends in dropout rates, however, indicates that for this variable at least, selection into licensing status is based on permanent differences across states. These differences should therefore be accounted for by state fixed effects. We thank Doug Staiger for pointing out these particular correlations.

[16]Angus (2001) and Ravitch (2002) suggest teacher certification requirements have been at the heart of battles over entry to the teaching profession since the mid-19th century. Barriers to entry commonly increased in times of relatively abundant teacher supply and testing was a relatively common component of teacher certification in the early part of the 20th century. Barriers have typically been relaxed in times of shortage and/or high demand (e.g., WWII). We note, however, that endogeneity of this sort would tend to bias our wage effects downwards.

Please cite this article as: Angrist, J. D., & Guryan, J. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review* (2007), doi:10.1016/j.econedurev.2007.03.002

Table 4

	Log (salary with B.A.)				Log (salary with M.A.)				Log (salary with M.A. + 20 years)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Wage estimates controlling for state and year fixed effects[a]												
Full sample												
Basic skills test law	.024		.018		.025		.022		.026		.026	
	(.009)		(.010)		.025		(.009)		(.009)		(.009)	
Subject test law		.023	.015			.019	.010			.012	.000	
		(.012)	(.013)			(.013)	(.014)			(.014)	(.014)	
Any test law				.019				.021				.022
				(.009)				(.009)				(.010)
R^2	.831	.831	.831	.831	.803	.802	.803	.802	.767	.767	.767	.767
N							18,060					
Inexperienced teachers												
Basic skills test law	.022		.015		.024		.017		.033		.027	
	(.011)		(.012)		(.011)		(.012)		(.012)		(.012)	
Subject test law		.028	.022			.027	.020			.031	.019	
		(.014)	(.015)			(.015)	(.016)			(.019)	(.019)	
Any test law				.019				.020				.042
				(.013)				(.013)				(.015)
R^2	.858	.858	.859	.857	.832	.832	.833	.832	.789	.789	.790	.790
N							6404					
First-year teachers												
Basic skills test law	.024		.019		.022		.018		.036		.035	
	(.014)		(.014)		(.014)		(.014)		(.014)		(.014)	
Subject test law		.026	.019			.023	.016			.017	.004	
		(.019)	(.019)			(.020)	(.020)			(.023)	(.022)	
Any test law				.016				.012				.029
				(.018)				(.017)				(.019)
R^2	.866	.866	.866	.865	.839	.839	.840	.839	.786	.785	.787	.786
N							2979					
Wage estimates controlling for state and year fixed effects and state-specific linear trends[b]												
Full sample												
Basic skills test law Test Law	.034		.027		.033		.026		.020		.015	
	(.010)		(.011)		(.010)		(.010)		(.009)		(.010)	
Subject test law		.033	.022			.031	.020			.021	.014	
		(.011)	(.011)			(.011)	(.012)			(.014)	(.015)	
Any test law				.030				.031				.019
				(.009)				(.009)				(.011)

Please cite this article as: Angrist, J. D., & Guryan, J. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review* (2007), doi:10.1016/j.econedurev.2007.03.002

Table 4 (*continued*)

	Log (salary with B.A.)				Log (salary with M.A.)				Log (salary with M.A. + 20 years)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
R^2	.844	.844	.844	.843	.816	.815	.816	.815	.775	.775	.775	.775
N							18,060					
Inexperienced teachers												
Basic skills test law	.046		.033		.047		.032		.032		.013	
	(.014)		(.016)		(.013)		(.014)		(.016)		(.016)	
Subject test law		.051	.038			.052	.039			.055	.049	
		(.013)	(.014)			(.012)	(.013)			(.018)	(.019)	
Any test law				.041				.045				.050
				(.012)				(.012)				(.018)
R^2	.873	.873	.874	.873	.847	.848	.848	.847	.800	.801	.801	.800
N							6404					
First-year teachers												
Basic skills test law	.047		.035		.045		.031		.025		.011	
	(.015)		(.017)		(.015)		(.016)		(.018)		(.017)	
Subject test law		.051	.038			.059	.047			.049	.045	
		(.018)	(.018)			(.017)	(.018)			(.028)	(.029)	
Any test law				.037				.042				.029
				(.013)				(.013)				(.021)
R^2	.886	.886	.887	0.885	.860	.860	.861	.859	.800	.800	.800	.799
N							2979					

[a]Inexperienced teachers are defined as teachers with less than 4 years teaching experience. Reported coefficients are estimated from an OLS regression of log salary on state testing requirement dummy variables and a set of controls. Controls include state and year fixed effects, city, suburb and rural fixed effects, a quadratic in the state unemployment rate, district enrollment, and district fraction minority enrollment. All regressions are weighted using district sampling weights. Standard errors corrected for state-by-year correlation in the error term are reported in parentheses.

[b]Inexperienced teachers are defined as teachers with less than 4 years teaching experience. Reported coefficients are estimated from an OLS regression of log salary on state testing requirement dummy variables and a set of controls. Controls include state and year fixed effects, state-specific linear trends, city, suburb and rural fixed effects, a quadratic in the state unemployment rate, district enrollment, and district fraction minority enrollment. All regressions are weighted using district sampling weights. Standard errors corrected for state-by-year correlation in the error term are reported in parentheses.

Please cite this article as: Angrist, J. D., & Guryan, J. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review* (2007), doi:10.1016/j.econedurev.2007.03.002

A comparison of estimates with and without state-specific trends rules out the hypothesis that the results in Table 4a are driven by correlated trends in testing and teacher salaries.

A simple falsification exercise supports this interpretation. Districts that imposed testing between 1987–1988 and 1990–1991 had no greater salary growth between 1990–1991 and 1999–2000 than those that did not impose testing during this earlier period. Omitted trends correlated with treatment status would lead to continued growth in treatment states even in the absence of certification. As a further check to ensure that certification increases preceded salary increases and not vice versa, we also calculated a Granger-style causality test by adding a dummy for the introduction of certification requirements *next year* into our basic model. When added to the basic specification without state-specific trends, this variable typically comes in with a small and insignificant coefficient. For example, the *p*-value that corresponds to the correlation between a future basic skills test requirement and the starting salary for teachers with a B.A. is 0.419. The corresponding *p*-values for salaries paid to starting teachers with an M.A. and to teachers with an M.A. plus 20 years of experience are 0.471 and 0.804, respectively. Of the 45 coefficients reported in Table 4a, six of the corresponding Granger coefficients have *p*-values below 0.1.

Furthermore, the effects of testing controlling for state-specific trends are generally larger for teachers with a B.A. than for those with more education or experience, consistent with the notion that entry wages should change the most in response to barriers. Similarly, the effects are larger in the sample of districts that have new or inexperienced teachers than in the full sample of districts. Finally, it is worth repeating in this context that the first-stage estimates of the effect of testing requirements in Table 3 are on the order of 50% points. This implies that two-stage least squares (2SLS) estimates of the effect of *district* testing on teacher wages—using state testing regulations as instruments—are about twice as large as the reduced-form effects of testing regulations reported in Tables 4a and b. The 2SLS interpretation, however, turns on whether state regulations satisfy an exclusion restriction. In practice, it seems likely that mandatory testing could affect the wage distribution in districts that tested even in the absence of a state requirement.

An important consideration in this context is the role of collective bargaining in the teacher labor market. In particular, district unionization status may be an omitted factor in models for the effects of licensure. Motivated by this possibility, we estimated specifications that add controls for unionization to the wage regressions reported in Tables 4a and b. To do so, it was necessary to reduce the sample sample because the unionization variable appears only in the 1993–1994 and 1999–2000 waves of the SASS. These results, which can be found in Angrist and Guryan (2007), are similar to those in Tables 4a and b.

Finally, though we have focused here on testing requirements, it is worth noting that there may be other important changes in state licensing rules.[17] One of the most important of these is the institution of alternative certification procedures that waive standard licensing requirements. If alternative certification provisions are changed at the same time that tests are introduced, the wage effects reported here may be only partially due to testing. The results reported below, however, show no relation between testing and the use of alternative certification, suggesting it is reasonable to look at testing in isolation. As a caveat, we note that some of the wage effects reported in Tables 4a and b may be due to non-testing teacher licensure requirements. Even if so, however, this can be seen as consistent with our basic argument that the costs of screening constitute a barrier to entry.

5.2. Effects on quality of undergraduate institution

Although state testing requirements are associated with an increase in the use of teacher tests and with higher teacher wages, there is little evidence that this translates into better teachers, at least along the quality dimensions we can measure. These results can be seen in Table 5a, which reports quality estimates for the samples of new and inexperienced teachers since these are the samples where we expect effects to be largest. For example, columns 1–4 of Table 5a show no evidence of an association between testing requirements and the quality of teachers' undergraduate institution as measured by average SAT scores. While the subject

[17]See e.g. Goldhaber (2004) for a discussion of other teacher licensure requirements.

Please cite this article as: Angrist, J. D., & Guryan, J. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review* (2007), doi:10.1016/j.econedurev.2007.03.002

Table 5

	College selectivity							
	Average SAT				Research univ. or liberal arts coll.			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Teacher quality estimates[a]								
Inexperienced teachers								
Basic skills test law	.77		−2.10		−.002		−.007	
	(2.96)		(3.07)		(.018)		(.020)	
Subject test law		7.41	8.31			0.011	.014	
		(4.05)	(4.41)			(.020)	(.023)	
Any test law				1.44				−.006
				(3.72)				(.021)
R^2	.273	.274	.274	.273	.062	.062	.062	.062
N			6476				6716	
First-year teachers								
Basic skills test law	−5.52		−5.07		.031		.036	
	(4.56)		(4.62)		(.026)		(.030)	
Subject test law		−3.48	−1.56			.000	−.013	
		(6.23)	(6.39)			(.030)	(.033)	
Any test law				−6.42				−.011
				(6.57)				(.026)
R^2	.239	.239	.239	.239	.072	.072	.072	.072
N			3008				3160	
	Other qualifications							
	Majored in teaching subject				Alternative certification			
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Inexperienced teachers								
Basic skills test law	.008		−.001		−.017	−.019		
	(.008)		(.009)		(.025)	(.027)		
Subject test law		.027	.027			−.003	.006	
		(.011)	(.012)			(.036)	(.038)	
Any test law				.006				−.019
				(.010)				(.024)
R^2	.018	.019	.019	.018	.183	.183	.184	.183
N			9114				9114	
First-year teachers								
Basic skills test law	.004		.001		−.024		−.020	
	(.019)		(.017)		(.038)		(.038)	

Subject test law			.012		.011				−.023		−.015	
			(.019)		(.019)				(.048)		(.049)	
Any test law						−.005						−.041
						(.019)						(.040)
R^2	.027		.027		.027	.027		.183	.183		.183	.183
N				4302						4302		

	Fraction Black				Fraction Hispanic				Fraction female			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Teacher characteristics estimates[b]												
Inexperienced teachers												
Basic skills test law	−.005		−0.006		−.009		−.007		−.001		.004	
	(.005)		(.005)		(.004)		(.005)		(.016)		(.017)	
Subject test law		.001	.003			−.008	−.005			−.016	−.018	
		(.007)	(.007)			(.008)	(.009)			(.019)	(.020)	
Any test law				−.003				−.010				−.003
				(.007)				(.005)				(.016)
R^2	.141	.141	.141	.141	.081	.081	.081	.081	.032	.032	.032	.032
N			9114				9114				9114	
First-year teachers												
Basic skills test law	−.002		−.005		−.020		−.016		−.012		−.009	
	(.007)		(.007)		(.008)		(.009)		(.026)		(.025)	
Subject test law		.008	.010			−.019	−.013			−.012	−.008	
		(.011)	(.011)			(.013)	(.014)			(.029)	(.027)	
Any test law				.006				−.021				−.023
				(.009)				(.009)				(.027)
R^2	.162	.162	.162	.162	.126	.126	.126	.126	.042	.042	.042	.042
N			4302				4302				4302	

[a]Inexperienced teachers are defined as teachers with less than 4 years teaching experience. Dependent variables are defined as follows: Average SAT is the average SAT score of matriculating freshmen in 1983 at the undergraduate college attended by the teacher; Research University or Liberal Arts College is the fraction of teachers that attended undergraduate colleges in one of the following categories as defined by the Carnegie Commission on Higher Education: Research Universities I and II, and Baccalaureate (Liberal Arts) Colleges I; Alternative Certification is an indicator for whether the teacher was hired without regular state certification (alternatives are temporary certification, provisional certification and emergency certification). Majored in Teaching Subject is the fraction of teachers whose primary teaching assignment is in the same subject as their B.A., M.A., Ph.D. or Education Specialist degree major. Reported coefficients are estimated from an OLS regression on state testing requirement dummy variables and a set of controls. Controls include state and year fixed effects, city, suburb and rural fixed effects, a quadratic in the state unemployment rate, district enrollment, and district fraction minority enrollment. All regressions are weighted using district sampling weights. Standard errors corrected for state-by-year correlation in the error term are reported in parentheses. There are fewer observations for the Average SAT and Carnegie Type I specifications because the 1990–91 SASS does not report the teacher's undergraduate college.

[b]Inexperienced teachers are defined as teachers with less than 4 years teaching experience. Dependent variables are the fraction of inexperienced or first-year teachers in the district who fall into the respective category. Reported coefficients are estimated from an OLS regression on state testing requirement dummy variables and a set of controls. Controls include state and year fixed effects, city, suburb and rural fixed effects, a quadratic in the state unemployment rate, district enrollment, and district fraction minority enrollment. All regressions are weighted using district sampling weights. Standard errors corrected for state-by-year correlation in the error term are reported in parentheses.

Please cite this article as: Angrist, J. D., & Guryan, J. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review* (2007), doi:10.1016/j.econedurev.2007.03.002

test requirement is associated with a marginally significant increase in test scores when both testing requirements are entered jointly and the sample includes only districts with inexperienced teachers (column 3), the corresponding effect is smaller and insignificant for all other specifications in this sample, and negative and insignificant in the sample of districts with new teachers. Similarly the estimates in columns 5–8 of Table 5a point to the absence of an association between testing and the quality of teachers' undergraduate institutions as measured by the institutions' Carnegie classification as a research university or a liberal arts college.[18]

As noted in the theoretical section, the extent to which teacher testing can raise teacher quality is determined partly by the power of the test as a quality screen. The screen may be weak since pass rates for the widely-used Praxis II test are very high; about 87% of 1997 applicants obtaining a passing score on the composite known as "General Praxis II" (Gitomer, Latham, & Ziomek, 1999). The fact that most applicants pass could well explain weak effects of state test requirements on the quality distribution. At the same time, the implied Praxis failure rate of 13% is taken from a population that is already (for the most part) subject to mandatory testing and is therefore preparing for the test. The failure rate in this population is probably lower than the failure rate that arises when testing is introduced in an applicant population not previously subject to testing. (And, of course, many states use tests other than the Praxis. Failure rates on the Massachusetts Teachers Exam, for example, have been much higher).

To get a sense of what the imposition of stricter testing standards might mean for the applicant quality distribution as we measure it, we used a unique data set provided by the ETS linking the SAT scores to the Praxis results of prospective teachers, most recently for 1997. These data, documented in Gitomer et al. (1999), allow us to determine the relationship between alternate testing standards on the Praxis II and the SAT scores of teacher applicants.[19] Moreover, we can look at the extent to which the effect of passing Praxis on SAT scores is diluted by replacing individual applicants' SAT scores with their college average score.

Our analysis here uses information for 1997, the sample year in which testing was most widespread. We discarded applicants who took the test voluntarily in the hope that this makes the results more representative. The analysis begins by exploring the relationship between alternative standards for passing Praxis and SAT scores. The regression-adjusted SAT score differential between the 87% who passed the General Praxis Standard in 1997 and those who failed is 224 points.[20] Imposing a somewhat higher standard which defines passers as applicants who succeed on all three core batteries, the pass rate is 75% and the SAT gap is 235 points.

The next step is to estimate the same score differential as measured by undergraduate institutional average—rather than individual—SAT scores. Those meeting the general Praxis II standard had a 55 point higher institutional average SAT score as measured using the ETS's 1997 sample, and a 46 point higher average as measured using the 1983 average SAT variable used in Table 5a. Likewise, those meeting the "three core battery" standard had a 79 point higher institutional average SAT score as measured using the 1997 ETS sample, and a 55 point higher average as measured using the 1983 average SAT variable used in Table 5a. This comparison of pass/fail differentials for individual and institutional-average SAT scores is consistent with the fact that teachers' undergraduate institution explains 20–30% of the total variance in their individual SAT scores.

Finally, we turn to the mechanical impact of a state-wide testing requirement on teacher quality, as measured by the average SAT score of teachers' undergraduate institutions. As discussed above, the differential between those passing and failing the test is on the order of 50 points for most standards. In the absence of any discouragement effect on highly-able teachers, turning away the 13% of applicants who failed the General Praxis II section would lead to a six point (46 point gap for passers $\times$ 0.13 failure rate) increase in the 1983

[18]In future work, we plan to look at student outcomes as well as teacher characteristics. It is worth noting, however, that because districts may reduce their demand for other productive inputs (e.g., small classes) in response to the requirement to purchase more of the skills measured by teachers' test scores, it seems likely that achievement effects will be smaller than effects on teacher characteristics. We thank Doug Staiger for pointing this out.

[19]These data are not publicly available, and were graciously provided for the purposes of this project by ETS, with permission of the College Board.

[20]Regressions adjust for teachers' year of birth and state of residence.

Please cite this article as: Angrist, J. D., & Guryan, J. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review* (2007), doi:10.1016/j.econedurev.2007.03.002

average SAT variable used in Table 4a. The corresponding figure for a stricter standard relying on the three core battery tests is 13.75 points (55×0.25). The standard errors in Table 4a are such that effects of this magnitude on the score distribution would, in principle, be detectable at least in the full sample of teachers.[21]

The upshot of this discussion is that there appears to be enough of a link between Praxis standards and SAT scores for the use of Praxis to have, at least in principle, increased our imputed SAT measure by about 6–14 points, a hypothesis inconsistent with most of the findings in Table 5a. The absence of quality effects in Table 5a may be a consequence of the fact that tests like the Praxis screen out relatively few teachers, or a reflection of our theoretical prediction that testing deters some relatively high-SAT applicants from teaching in public schools.

We wish to highlight the other obvious possibility that undergraduate college selectivity is not a powerful measure of "teacher quality". Though there is evidence that the variation in teacher quality is an important determinant of student achievement (e.g., Aaronson, Barrow, & Sander, 2002; Hanushek, 1971; Rockoff (2004)), it is hard to find an observable characteristic of teachers that consistently predicts gains in student outcomes. One reason for the lack of positive effects in Table 5a may be that our measure of teacher quality is uncorrelated with important alternative dimensions of teacher quality. In the next section, we turn to an analysis of additional measures of teacher quality and teacher characteristics.

5.3. *Effects on alternative measures of quality and teacher demographics*

Estimated effects of testing on two alternative measures of teacher qualifications—the fraction of teachers that majored in their teaching subject and the fraction of teachers with alternative certification—are reported in the bottom panel of Table 5a.[22] In contrast with the institutional average SAT and Carnegie score, these outcomes vary within institutions. The results in columns 9–12, which show effects on the probability teachers majored in the subject they teach, shed some light on the hypothesis that testing causes schools to hire more desirable teachers from the same colleges. In fact, the probability that a teacher majored in his or her teaching subject appears to rise in states that impose a subject test. On the other hand, this effect is not very robust. When estimated in the sample of inexperienced teachers, the imposition of a subject test increases the probability teachers teach in their major by about 2.7%, with a standard error of 1.1%, but the corresponding estimate is about half as large and insignificant in the sample of new teachers.

As a measure of teacher quality, the use of alternative certification methods can be seen as a plus or a minus, depending on the value of traditional certification methods as a quality screen. An important question for our purposes, however, is whether the introduction of tests is confounded with other sorts of licensing reforms. It is particularly important to establish that districts do not avoid testing requirements by hiring more teachers without standard certification. As it turns out, alternative certification is uncorrelated with testing requirements (columns 13–16 of Table 5a), suggesting that our estimates of testing effects are not confounded with other changes in certification policy.

The last set of estimates looks at the relationship between state testing requirements and the demographic make-up of the teaching labor force. This inquiry is motivated partly by the fact that standardized tests are sometimes thought to be more of a barrier for minorities. The first four columns of Table 5b show no relationship between state testing requirements and the percent of new or inexperienced teachers who are black. On the other hand, there is some evidence of a negative association between testing requirements, especially for basic skills, and the number of new teachers who are Hispanic. Columns 5–8 suggest that testing requirements reduce the proportion of new teachers who are Hispanic by about 2% points, a large effect given that only 5% of new teachers were Hispanic in 1999–2000. Finally, there is no relation between

[21]The effects may be further diluted by the fact that about half of districts in states with no testing requirement nevertheless used tests. In practice, however, the testing regime in such states was less binding.

[22]Only math, science and English teachers were coded as having majored in the same subject they teach. At the same time, the sample is not limited to math, science, and English teachers because this may be an outcome.

(*footnote continued*)

Please cite this article as: Angrist, J. D., & Guryan, J. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review* (2007), doi:10.1016/j.econedurev.2007.03.002

mandatory testing and the fraction of new teachers who are female.

6. Summary and conclusions

Recent years have seen accelerating use of standardized tests to certify new teachers. Proponents hope these measures will increase quality, but economists have long been skeptical of entry barriers that shift supply and discourage otherwise-qualified individuals from applying for jobs. Our investigation of the impact of the use of tests to certify teachers for employment in public schools suggests state requirements increase the use of tests by about 50 percentage points. Testing requirements are also associated with higher teacher wages, consistent with a supply-shift story. Taking estimates from models that control for state-specific linear trends as representative, the reduced form effect of testing on wages is 3–5%. The implied 2SLS effect of the use of tests is twice as large. But there is little evidence of an impact of testing on teacher quality, at least as measured here. It bears emphasizing that our quality measures are proxies at best. Still, similar outcomes are widely used in the literature on determinants of teacher quality.

On balance, our results are reasonably consistent with the view that testing has acted more as a barrier to entry than a quality screen. One seemingly contradictory result, however is the finding that the imposition of a testing requirement is associated with an increase in the probability of new teachers teaching a subject in which they majored. This result points towards testing as having potential value as a device to select teachers with stronger subject-specific skills. On the other hand, this particular finding is not especially strong, and is sensitive to specification changes.

Another interesting finding is the negative association between teacher testing and the probability new teachers are Hispanic. The use of standardized tests often raises concerns about adverse impacts on members of minority groups, who often have lower test scores. In the case of college admissions, for example, lower SAT scores by non-whites have led some schools to put less weight on the SAT for admissions. Given that Hispanics have markedly lower Praxis scores than non-Hispanic Whites or Blacks (Gitomer et al., 1999), it is perhaps not surprising that teacher testing has a negative impact on Hispanic representation, especially among new teachers.

Concerns about testing notwithstanding, the question of how to increase and maintain the quality of public school teachers remains. Ballou's (1996) results indicate that teachers' employers pay surprisingly little attention to the selectiveness of applicants' undergraduate institution. Along these lines, Manski (1987) suggested that a floor for teachers' SAT scores could provide a useful screening mechanism. A reliance on SATs would appear to avoid some of the problems outlined in our theoretical discussion since this avoids the establishment of a unique barrier to teaching, and may also force school districts to focus more on college quality. This naturally raises the question of whether teachers with higher SAT scores are indeed better teachers, a subject for future research.

Acknowledgments

The authors thank Caroline Hoxby, Kevin Murphy and Doug Staiger for their comments and suggestions, and Nate Baum-Snow, Naveed Khawaja, Andrew Kolesnikov, Chris Smith, and Nirupama Rao for excellent research assistance. We also thank participants at the 2003 NBER Summer Institute, Columbia, Rice, Texas A&M, and the University of Houston for comments. The confidential data used in this paper were provided by the National Center for Education Statistics, US Department of Education and the College Board. Special thanks go to Drew Gitomer for facilitating our access to the College Board data. This work is supported by the Centel Foundation/Robert P. Reuss Faculty Research Fund at the Graduate School of Business, the University of Chicago. Angrist gratefully acknowledges financial support under NICHD grant R01HD043809-01A1.

Appendix A

The extract used here was drawn from the Public School TDSS component of the SASS. The TDSS is administered to a stratified random sample of school districts in the US. The data used in the analysis are from the restricted-use files of the 1987–1988, 1990–1991, 1993–1994 and 1999–2000 waves of the SASS. Individual teacher-level information is extracted from the Teacher Questionnaire of the SASS. Characteristics of colleges attended by teachers are then merged by college FICE codes to the teacher-level data. These data are then weighted by sampling means, aggregated to the district level,

Please cite this article as: Angrist, J. D., & Guryan, J. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review* (2007), doi:10.1016/j.econedurev.2007.03.002

and merged to the district-level TDSS. Finally, state-by-year economic measures are merged to the data set. Districts with fewer than 50 students are dropped from the analysis, as are charter districts in the 1999–2000 wave. Throughout the analysis, first-year teachers are defined as teachers who report their first year of teaching to be the year of the survey. Inexperienced teachers are defined as teachers who report their first year of teaching to be less than four years before the year of the survey.

The following definitions were used to create outcome variables extracted from the SASS:

Salary: *B.A.*: Data come from district-level responses. The base salary paid to a teacher in the district with a Bachelor of Arts degree, no teaching experience, and no other relevant credentials. Responses are inflated to 1999 dollars using the CPI-U.

Salary: *M.A.*: Data come from district-level responses. The base salary paid to a teacher in the district with a Master's degree, no teaching experience, and no other relevant credentials. Responses are inflated to 1999 dollars using the CPI-U.

Salary: *M.A.*+20 *years*: Data come from district-level responses. The base salary paid to a teacher in the district with a Master's degree, at least 20 years teaching experience, and no other relevant credentials. Responses are inflated to 1999 dollars using the CPI-U.

Majored in Teaching Subject: Data come from individual teacher responses. A dummy is created that equals one if one of the following three criteria are met: (1) the teacher's primary teaching assignment is Mathematics and he completed either a B.A, M.A. Ph.D. or Education Specialist degree with a major in either Mathematics, Engineering or Economics; (2) the teacher's primary teaching assignment is English and he completed either a B.A, M.A. Ph.D. or Education Specialist degree with a major in English Literature, Letters, Speech, Classics or Composition; (3) the teacher's primary teaching assignment is either Biology, Chemistry, Geology/Earth Science, Physics or General Science and he completed either a B.A, M.A. Ph.D. or Education Specialist degree with a major in either Biology, Chemistry, Geology/Earth Science, Physics, or another Physical Science. This dummy variable is then aggregated using sampling weights to compute the fraction of first-year teachers and inexperienced teachers for which the dummy is equal to one.

Alternative Certification: Data come from individual teacher responses. Teachers are asked what type of state certification they hold in their main assignment field. A dummy is created that equals zero if the teacher describes his certification as either regular, standard or advanced, and one otherwise. This dummy variable is then aggregated using sampling weights to compute the fraction of first-year teachers and inexperienced teachers for which the dummy is equal to one.

The following definitions are used to define hiring-practices variables extracted from the SASS:

Requires Basic Skills Test: Data are drawn from the TDSS survey of school districts. Districts are asked whether they require teaching applicants to have passed a test of basic skills. A dummy is created which is equal to one if the district requires a state test of basic skills, a district test of basic skills or the National Teachers Exam/Praxis. In some of the analysis, this variable is automatically switched to one if the district is in a state that is mandated by law to require new teachers to pass a standardized test of basic skills.

Requires Subject Test: Data are drawn from the TDSS survey of school districts. Districts are asked whether they require teaching applicants to have passed a test of basic skills. A dummy is created which is equal to one if the district requires a state subject test, a district subject test or the National Teachers Exam/Praxis. In some of the analysis, this variable is automatically switched to one if the district is in a state that is mandated by law to require new teachers to pass a standardized subject test.

The following definitions are used to define quality measures of undergraduate institutions attended by teachers:

Average SAT Score: Data come from a survey conducted by the Higher Education Research Institute (Astin et al., 1983). The average combined Math and Verbal SAT score of entering freshman in the fall of 1983 is collected for colleges and universities from college guides and from surveys of college representatives. For schools that do not require students to take the SAT, ACT averages are translated into SAT averages using the following methodology. Samples of students who took both the SAT and ACT, or who took either test and a third common test (the National Merit Scholarship Qualifying Test) are compared. These overlapping samples are used to compute the equivalent percentiles in each test's distribution. ACT scores

Please cite this article as: Angrist, J. D., & Guryan, J. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review* (2007), doi:10.1016/j.econedurev.2007.03.002

are then replaced with the corresponding SAT scores at the equivalent point in the distribution.

Attended Research University or Liberal Arts College: A dummy variable is created that equals one if the college attended by the teacher is in one of the following categories of the Carnegie Classification of Institutions of Higher Education (1994 definitions): Research University I, Research University II, or Baccalaureate (Liberal Arts) Colleges I. The three categories included in the dummy are the three non-specialized categories with average SAT scores greater than 1000.

Information on state testing laws was drawn from the following sources:

Teacher Education Policy in the States: A 50-State Survey of Legislative and Administrative Actions, American Association of Colleges for Teacher Education (Washington, DC: December 1994).

Teacher Education Policy in the States: A 50-State Survey of Legislative and Administrative Actions, American Association of Colleges for Teacher Education (Washington, DC: Spring 1994).

Goertz, Margaret E., *State Educational Standards in the 50 States*: *An Update*, Educational Testing Service (Princeton, NJ: March 1988).

Coley, Richard J. and Goertz, Margaret E., *Educational Standards in the 50 States*: *1990*, Educational Testing Service (Princeton, NJ: June 1990).

Rudner, Lawrence M., What's Happening in Teacher Testing: An Analysis of State Teacher Testing Practices, US Department of Education (Washington, DC: August 1987).

The NASDTEC Manual on the Preparation and Certification of Educational Personnel, Kendall/Hunt Publishing Company (Dubuque, IA: 1999).

National Council for the Accreditation of Teacher Education (1997), *Standards*, *Procedures*, *and Policies for the Accreditation of Professional Education Units*, NCATE: Washington, DC.

References

Aaronson, D., Barrow, L., & Sander, W. (2002). *Teachers and student achievement in the Chicago public high schools*. Federal Reserve Bank of Chicago, Working Paper Series, WP-02-28.

Angrist, J. D., & Guryan, J. (2004). Teacher testing, teacher education, and teacher characteristics. *American Economic Review*, *94*(2), 241–246.

Angrist, J. D., & Guryan, J. (2007). *Does teacher testing raise teacher quality? Evidence from state certification requirements*. Princeton University Education Research Section Working Paper #21.

Angus, D. L. (2001). In J. Mirel (Ed.), *Professionalism and the public good: A brief history of teacher certification*. Washington, DC: The Fordham Foundation.

Astin, A. W., Green, K. C., Korn, W. S., & Maier, M. J. (1983). *The American freshman*: *National norms for fall 1983*. Cooperative Institutional Research Program of the American Council on Education, Higher Education Research Institute, Graduate School of Education, UCLA.

Ballou, D. (1996). Do public schools hire the best applicants? *Quarterly Journal of Economics*, *111*(1), 97–133.

Ballou, D., & Podgursky, M. (2000). Gaining control of professional licensing and advancement. In T. Loveless (Ed.), *Conflicting missions? Teachers unions and educational reform*. Washington, DC: The Brookings Institution.

Bacolod, M. P. (2007). Do alternative opportunities matter? The role of female labor markets in the decline of teacher quality, 1960–1990. *Review of Economics and Statistics, 89*(4), forthcoming.

Berger, M. C., & Toma, E. T. (1994). Variation in state education policies and effects on student performance. *Journal of Policy Analysis and Management*, *13*(3), 477–491.

Corcoran, S. P., Evans, W. N., & Schwab, R. M. (2004). Changing labor-market opportunities for women and the quality of teachers, 1957–2000. *American Economic Review*, *94*(2), 230–235.

Figlio, D. (1997). Teacher salaries and teacher quality. *Economics Letters*, *55*(2), 267–271.

Figlio, D. (2002). Can public schools buy better-qualified teachers? *Industrial and Labor Relations Review*, *55*, 686–699.

Friedman, M. F., & Kuznets, S. (1945). *Income from independent professional practice*. New York: National Bureau of Economic Research.

Gitomer, D., Latham, A. S., & Ziomek, R. (1999). *The academic quality of prospective teachers*: *The impact of admissions and licensure testing*. Educational Testing Service, Teaching and Learning Research Report Services. Princeton, NJ: ETS.

Goldhaber, D. D. (2004). Why do we license teachers? In F. M. Hess, A. J. Rotherham, & K. Walsh (Eds.), *A qualified teacher in every classroom?*. Cambridge: Harvard Education Press.

Goldhaber, D., & Brewer, D. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, *22*(2), 129–145.

Guasch, J. L., & Weiss, A. (1980). Wages as sorting mechanisms in competitive markets with asymmetric information. *Review of Economic Studies*, *47*, 653–664.

Guasch, J. L., & Weiss, A. (1981). Self-selection in the labor market. *American Economic Review*, *71*, 275–284.

Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: estimation using micro data. *American Economic Review*, *61*(2), 280–288.

Hanushek, E. A., & Pace, R. A. (1995). Who chooses to teach and why? *Economics of Education Review*, *14*, 101–117.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (1999). *Do higher salaries buy better teachers*? NBER Working Paper 7082.

Hoxby, C. M., & Leigh, A. (2004). Pulled away or pushed out? Explaining the decline of teacher aptitude in the United States. *American Economic Review*, *94*(2), 236–240.

Please cite this article as: Angrist, J. D., & Guryan, J. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review* (2007), doi:10.1016/j.econedurev.2007.03.002

Kleiner, M. M. (2000). Occupational licensing. *Journal of Economic Perspectives*, *14*, 189–202.

Kleiner, M. M., & Kudrle, R. T. (2000). Does regulation affect economic outcomes? The case of dentistry. *Journal of Law and Economics*, *43*, 547–582.

Kleiner, M. M., & Petree, D. L. (1988). Unionism and licensing of public school teachers: Impact on wages and educational output. In R. B. Freeman, & C. Ichniowski (Eds.), *When public sector workers unionize*. Chicago: University of Chicago Press [Chapter 11].

Kugler, A., & Sauer, R. (2003). *Doctors without borders: The returns to professional licensing for immigrant physicians in Israel*. CEPR Discussion Paper 3683.

Lelande, H. E. (1979). Quacks, lemons, and licensing: A theory of minimum quality standards. *Journal of Political Economy*, *87*, 1328–1346.

Manski, C. F. (1987). Academic ability, earnings, and the decision to become a teacher: Evidence from the National Longitudinal Study of the High School Class of 1972. In D. Wise (Ed.), *Public sector payrolls*. Chicago: University of Chicago Press.

Murnane, R. J., Singer, J. D., Willett, J. B., Kemple, J., & Olsen, R. (1991). *Who will teach? Policies that matter*. Cambridge: Havard University Press.

Ravitch, D. (2002). A brief history of teacher professionalism. In *White house conference on preparing tomorrow's teachers*, Washington, DC.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, *94*(2), 247–252.

Wang, R. (1997). Competition, wage commitments, and application fees. *Journal of Labor Economics*, *15*, 124–142.

Wang, R., & Weiss, A. (1998). Probation, layoffs, and wage-tenure profiles: A sorting explanation. *Labour Economics*, *5*, 359–383.

Please cite this article as: Angrist, J. D., & Guryan, J. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review* (2007), doi:10.1016/j.econedurev.2007.03.002