# Occupational Licensing and Quality:

# Distributional and Heterogeneous Effects in the Teaching

# Profession

Bradley Larsen

Stanford University and NBER

May 14, 2015

## Abstract

This paper examines a common form of entry restriction: occupational licensing. The paper studies two questions: first, how occupational licensing laws affect the distribution of quality, and second, how the effects of licensing on quality vary across regions of differing income levels. The paper uses variation in state licensing requirements for teachers and two national datasets on teacher qualifications and student outcomes from 1983–2008. Two measures of quality are used: the qualifications of candidates entering the occupation (input quality) and the quality of service provided (output quality). Results show that more restrictive licensing laws—in the form of certification tests required for initial licensure—may lead some first-year teachers of high input quality to opt out of the occupation. In the sample of teachers who remain in the occupation multiple years, stricter licensing appears to increase input quality at most quantiles of the teacher quality distribution. Output quality, as measured by student test scores, also changes with stricter occupational licensing, revealing a widening of the distribution. For most forms of licensing studied, input and output quality improvements due to stricter licensing requirements occur in high-income rather than low-income school districts.

**Keywords**: Occupational licensing, entry restrictions, quality, teacher certification

**JEL Classification**: J4, L5, K2, I2, L8

# 1   Introduction

Occupational licensing affects nearly 30% of the US labor force (Kleiner and Krueger 2010), a larger proportion of workers than are in unions or covered by minimum wage laws, and over 800 occupations are licensed in at least one state ( Kleiner 2006). Doctors, lawyers, teachers, barbers, and, in some states, even interior designers, auctioneers, frog farmers, fortune tellers, and florists, are required to have a license to legally practice. Proponents of licensing argue that these laws decrease informational asymmetries by preventing unqualified candidates from practicing, thus raising the lower tail of the quality distribution, providing a minimum guaranteed level of quality and safety to consumers. [1]  Opponents argue that licensing merely secures higher rents for those in the occupation, raising prices and, in particular, harming low-income consumers who may not be able to afford their preferred level of service. These debates date back centuries and continue today.[2]

While studies have shown that licensing raises wages, raises prices for consumers, and deters entry, little is known about the effects of licensing laws on the distribution of quality or about how licensing laws differentially affect consumers of low vs high income. Section 2 discusses the existing empirical literature of the effects of occupational licensing on quality, which have focused on *average* quality for the *average* consumer, finding, for the most part, that licensing has no significant positive effect. However, from a public interest standpoint, occupational licensing is intended to eliminate the *worst* candidates from entering into an occupation (regulating input quality) or to prevent the worst outcomes from occurring (regulating output quality).[3] Therefore, in this paper I argue that effects on *average* quality may mask important effects of licensing. In particular, licensing may have little effect on average quality but still move the tails of the quality distribution. Moreover, price increases from occupational licensing may more negatively impact low-income consumers than high-income consumers. Section 2 discusses theoretical and

---

[1]Many proponents present consumer safety as an argument for licensing. Ballou and Podgursky (1998) explain, "The [ professional teacher organization] frequently resorts to argument by analogy, comparing teaching to medicine...Doctors put in years of training in medical school and residencies before achieving full professional standing. They must pass rigorous licensing examinations. Why should we expect less of teachers?" Such arguments are frequently used in other professions as well, such as in the recent Florida case, where interior designers, lobbying against a bill that would have de-regulated their occupation by no longer requiring licenses, argued that the actions of unlicensed designers would lead to furniture in jail cells being used as weapons, hospital fabrics spreading disease, and flammable rugs spreading fires, "contributing to 88,000 deaths every year" (Campo-Flores 2011). The interior designer lobby won the debate and the profession remains licensed in Florida. Other proposals in Florida at this same time to deregulate other occupations, such as auctioneers, failed as well.

[2]Adam Smith (1776) criticized the occupational licensing regimes of his time, and a recent *Wall Street Journal* article highlighted that the Obama administration has budgeted $15 million to study the costs and benefits of occupational licensing (Litan 2015).

[3]This terminology distinguishing between input and output quality follows Shapiro (1986).

qualitative studies which have touched on these issues—Leland (1979) and Arrow (1963) presented models where licensing affects the distribution of quality, and Shapiro (1986) and Friedman and Friedman (1962) highlighted that licensing may differentially affect low vs. high-income consumers—but these issues have received little empirical attention. This paper is a first attempt to answer these questions empirically.

I focus on licensing in the market for teachers, using manually collected data on 26 years of state requirements for teachers (1983–2008) and two national datasets on teacher qualifications and student outcomes. Section 3 describes this data in detail, along with historical background about teacher certification tests. The measures of licensing stringency I employ are indicators for whether a given state in a given year required teachers to pass certain certification tests (basic skills tests, subject tests, and professional knowledge tests) prior to initial licensure.[4] Teacher qualifications are measured by the strength of teacher's undergraduate institution using data on 60,000 teachers in the Schools and Staffing Survey (SASS), serving as a measure of input quality, or the quality of candidates entering the occupation. Student outcomes are measured by eighth grade math scores from over 300,000 students administered the National Assessment for Education Progress (NAEP) exam, and serve as a measure of output quality, or the quality of the service provided.

The teaching profession provides a useful setting for studying the effects of licensing for several reasons. First, as described in Section 3, teacher licensing laws have been a major focus of public policy debates in recent years (see recent *Wall Street Journal* article by Mehta 2013; see also Nadler and Peterson 2009 and Weddle 2014) and in the past (Darling-Hammond 1997; Ballou and Podgursky 1998; Gardner 1983). Second, I observe measures of quality, which is rare in other occupations (Kleiner 2006; Gross 1986). Third, the use of licensing test laws varied greatly across states and across time from 1983–2008, allowing for identification of the effects of this particular form of licensing. Fourth, teacher certification tests represent a real obstacle to some candidates.[5] Also, these certification tests must be passed prior to initial licensure for teachers seeking standard certification and, in most states, for teachers seeking alternative certification, making it simple to identify which subgroup of teachers were affected by law

---

[4]In much of the occupational licensing literature (Kleiner 2006), the term "licensing" refers to mandatory requirements that must be met in order to legally practice in the profession, while "certification" refers to a qualification that a professional may obtain but which is not required in order to legally practice. Given that in the teaching profession licensing exams are referred to as "certification" exams, I use the both terms to refer mandatory requirements.

[5]The nationwide pass rate was 76% for whites and 42% for non-whites in 1984 (Rudner 1987), 93% in 2000 (Paige, Sroup, and Andrade 2002), and 96% in 2006 (Duncan and Ochoa 2002). Note that the test could be a costly obstacle to some even if the reported pass rate were 100%, as this pass rate is an equilibrium outcome of test difficulty and candidates' preparation effort. Further evidence that certification tests are a challenge to at least some candidates is found in the fact that test preparation companies offer numerous study products to aid students in preparing for the Praxis exams, the set of exams used by most states today.

changes.[6] Finally, certification test laws are easily quantifiable. Coursework or degree requirements, while potentially more costly to candidates than certification tests, are difficult to translate into a measure of licensing stringency.

Section 4 presents the empirical approach. To estimate the effects of certification test laws on the distribution of quality, I adopt the grouped quantile regression approach of Chetverikov, Larsen, and Palmer (2015), a quantile extension of Hausman and Taylor (1981). This approach allows me to focus on the effects of licensing on quantiles of the quality distribution rather than simply on average quality. To estimate how occupational licensing laws differentially affects average quality in high vs. low-income areas, I interact measures of licensing stringency with the percent of students qualifying for free lunch. I compute these effects for measures of input and output quality, and do so both for first-year teachers alone as well as for a pooled sample of new and experienced teachers, allowing me to form a synthetic panel that exploits additional variation in state licensing requirements.

In Section 5 I present the findings of the paper: an analysis of how certification test requirements for teachers affect the distributions of input and output quality, and how the effects of these laws differ for high vs. low-income areas. When examining the *input* quality distribution for first-year teachers, I find that subject area certification tests for initial licensure lowered the upper tail of the *input* quality distribution, suggestive that these licensing requirements may have driven away from teaching some candidates coming from highly selective undergraduate institutions. When instead examining long-run effects (i.e., looking at experienced teachers), I find that the effects of subject test laws are positive and relatively constant across the distribution of quality. In examining *output* quality, I find that, for first-year teachers, certain certification tests are associated with a decrease in the lower tail and an increase in the upper tail of the distribution of output quality, suggesting that the student test score distribution widens under stricter licensing requirements.

In examining how licensing requirements differentially affect areas of high vs. low-income, I find that any positive effects of teacher certification tests on quality accrue primarily to higher-income school districts. This is true for both input and output quality effects for most types of teacher certification tests. I also find some weak evidence that when certification tests are in place, lower-income school districts may substitute away from hiring licensed professionals by instead hiring emergency-certified teachers or increasing class sizes. Finally, I perform an event study that suggests that the primary impact of subject

---

[6]This is not the case with other licensing requirements, such as specific coursework requirements or a requirement to have a master's degree within a certain number of years, which differ by state and by whether or not the teacher was regularly or alternatively certified; in some states, a change in such laws would affect a potential teacher after completing a bachelor's degree, while in others, a change in requirements would affect the candidate midway through an undergraduate program.

test laws on quality occurs two or more years after the change in certification requirements, consistent with the existence of phase-in periods for these regulatory changes or with undergraduate major choices being influences by these changes.

Finally, Section 6 concludes and offers several suggestions for future research in occupational licensing to continue moving beyond studying average quality.

# 2    Theoretical and empirical work on occupational licensing

This section provides a brief discussion of previous theoretical and empirical work on occupational licensing, in particular in relation to the effects of licensing on the quality distribution and the differential effects of licensing for high and low-income consumers. Leland (1979) presented a model in the spirit of the lemons model of Akerlof (1970) in which licensing can weed out low-quality candidates from the occupation. Arrow (1963) provides a similar qualitative discussion. Shapiro (1986) presented a model in which licensing can encourage professionals to invest in producing higher quality services, as licensing laws provide an assurance to professionals that they will be able to reap the benefits of such investment. On the other hand, arguments such as those in Stigler (1971) and Peltzman (1976) suggest that licensing may decrease competition, which could lead to lower quality. Leland (1979) also pointed out that licensing standards set by industry organizations, as many are, will likely be set too high in an effort to increase market power. In addition, Ballou and Podgursky (1998) and other opponents of licensing argued that increasing licensing stringency may lead to a decrease in the upper tail of quality if the costs (monetary or time costs) of licensure induce candidates with higher outside opportunities to choose a different career.

The preceding literature is entirely theoretical; previous *empirical* research on occupational licensing has largely ignored the effects of licensing on the distribution of quality, instead focusing on *average* quality. Kleiner and Petree (1988), Berger and Toma (1994), Hanushek and Pace (1995), Goldhaber and Brewer (2000), Rockoff, Jacob, Kane, and Staiger (2008), Kane, Rockoff, and Staiger (2008), and Angrist and Guryan (2008, 2004), all focused on average effects and found no effect or a negative effect of increased licensing stringency on standardized test scores for students, the selectivity of teachers' undergraduate institutions, or other measures of teacher quality. Studies have also found non-positive effects of licensing on quality for electricians (Carroll and Gaston 1981), contractors (Maurizi 1980a), dentists (Kleiner and Kudrle 2000; Carroll and Gaston 1981), and physicians (Kugler and Sauer 2005).

Figure 1 illustrates effects that would be missed by examining only average quality. In panel (a), licensing leads to an increase in quality in the lower tail of the distribution. In (b) quality increases at

every quantile of the distribution. Focusing solely on average quality would not enable one to disentangle whether licensing serves only to drive out low-quality services or instead improves service across the board. Similarly, average quality decreases in both panels (c) and (d), but in (c) the decrease occurs primarily in the upper tail, driving out high-quality services and in (d) quality decreases at all quantiles. The results in Wiswall (2007) and Sass (2011) for teachers and in Ramseyer and Rasmusen (2015) for lawyers can be interpreted as evidence consistent with panel (c). Perhaps the most interesting case is (e), where licensing increases the lower tail, as suggested is possible in the Leland (1979) model, for example, and simultaneously drives out high-quality, high-outside-option candidates, as suggested by Ballou and Podgursky (1998), leading to a zero effect on average quality. Panel (f) provides a similar illustration with a mean-preserving spreading of the distribution rather than condensing.

Understanding which portion of the quality distribution is affected is key to understanding several policy issues: first, it can aid in understanding whether occupational licensing has any impact at all on quality (such as increasing the lower tail, even without moving the average); and, second, if licensing does have an effect, such analysis can aid in determining whether the benefits of licensing outweigh its costs, or whether alternative forms of quality improvement would be preferred, such as reputation systems or non-mandatory certification systems, such as those suggested by Shapiro (1986) and Kleiner (2006).

The differential effects of occupational licensing on consumers who differ by income or preference for quality were highlighted in the theoretical model of Shapiro (1986), who suggested that any benefits of licensing will accrue primarily to those consumers with a greater preference toward quality, benefiting high-income consumers at the expense of low-income consumers. Friedman and Friedman (1962) made a similar point, suggesting that licensing laws may prevent low-income consumers from gaining access to the service due to a lack of low-quality or lower-priced services. Currie and Hotz (2004) found empirical evidence of this effect, demonstrating that tighter educational requirements for child care professionals led to higher quality for children who received care, but also led to price increases resulting in less children being served. Also, Kleiner (2006) found that more restrictive licensing requirements in dentistry tended to benefit high income states and had no effect on low-income states. However, the majority of the empirical occupational licensing literature has not examined these heterogeneous effects for high vs. low-income consumers. It may be that the zero average effects documented in previous studies (described in the above paragraphs) mask underlying *positive* effects for high-income consumers and *negative* effects for low-income consumers, as theory suggests. This concept is illustrated in panel (a) of Figure 2, where the dashed line represents the average effect of increasing licensing stringency and the solid line represents the effect at different income levels (or different levels of the percent of students qualifying for free lunch). Panel (b) illustrates the opposite phenomenon, where the effect of licensing is larger for low-income

(high percent free lunch) areas. The interaction effects studied below provide a test of whether licensing differentially affects these different areas.[7]

# 3  Background and data on teacher certification laws and quality measures

## 3.1  Teacher certification laws

Competency tests for teachers were used in many states in the early twentieth century (Rudner 1987). Over time, testing requirements were replaced by educational requirements for prospective teachers, such as a bachelor's degree or specific education coursework. In the late 1970s and early 1980s, states developed a renewed interest in teacher testing, sparked by professional organizations, such as the American Association of Colleges for Teacher Education (AACTE), and by the influential report, "A Nation at Risk," (Gardner 1983) commissioned by the Department of Education, which advocated for educational reform, arguing that education professionals too often came from the bottom quartile of high school graduates.

As certification tests became more widespread from the 1980s to today, they evolved into three categories: basic skills tests (testing listening, reading, writing, communication, and mathematics), professional knowledge tests (testing pedagogical skills and knowledge), and subject matter tests (testing specific subjects).[8] Figure 3 displays the proportion of states requiring each type of certification test during the years 1983 to 2010.[9] I manually compiled data on teacher certification test laws from the following teacher certification manuals, which list historical details of certification requirements for each state: AACTE (1990, 1991, 1993–1996), Boydston (1995-1999, 2001–2010), Coley and Goertz (1990), Goddard (1983–1993), Goertz (1986, 1988), NASDTEC (1991, 1994, 1996, 1996, 2000–2002), NASDTEC Knowledgebase (2003–2010), Neuweiler (1988), Rudner (1987).

---

[7]Other aspects of occupational licensing examined in the empirical literature, and not focused on in this paper are effects on minority participation in the profession (Federman, Harrington, and Krynski 2006; Law and Marks 2009) and wages (Maurizi 1980b; Kleiner and Petree 1988; Kleiner 2000; Kleiner and Kudrle 2000; Tenn 2001; Kugler and Sauer 2005; Angrist and Guryan 2008; the large majority of these studies found positive effects on wages).

[8]For elementary school teachers, subject matter exams test content specific to teaching younger children. Also, some tests are referred to as general knowledge tests, and can include general knowledge of social studies, math, literature, fine arts, and science. However, in the historical certification manuals I use, both general knowledge tests and basic skills tests are often referred to as basic skills tests, and are thus indistinguishable. Therefore, I treat both as basic skills tests.

[9]The main empirical analysis of the paper uses only years 1983–2008 because quality measures are not available for 2009–2010.

As shown in Figure 3, by the late 1980s most states required some form of teacher testing. In the late 1990s, professional teacher organizations, such as the National Commission on Teaching and America's Future (NCTAF), the National Council of Accreditation of Teacher Education (NCATE), and the National Board for Professional Teaching Standards (NBPTS), argued for even stricter certification requirements (NCTAF 1996). Opponents of this movement argued that increasing the restrictiveness of licensing requirements would only drive away more qualified applicants with higher opportunity costs, inducing them to turn down teaching for alternative careers (Ballou and Podgursky 1998).

In the most recent decade, as seen in Figure 3, the use of basic skills tests and professional knowledge tests decreased, and the use of subject matter tests increased, in part due to the No Child Left Behind Act of 2001. A portion of this legislation required that, by the end of the 2005-2006 school year, each state demonstrate that its teachers met the definition of a Highly Qualified Teacher (HQT), which included demonstrating subject matter competency (Kuenzi 2009). Table 1 displays the number of states (including Washington, D.C.) that experienced a change from having no subject test law to having a subject test law, or vice-verse, and similarly for the other types of tests. Notice that very few states (only two) experienced a change from having a subject test law to having none. I exploit this fact in the analysis below to perform an event study of the impact of licensing changes.

## 3.2 Input and output quality measures

I measure input quality (i.e. a candidate's qualifications) using the selectivity of the teacher's undergraduate institution. I begin with data on public school teachers from the Schools and Staffing Survey (SASS) from the US Department of Education's National Center for Education Statistics (NCES).[10] The SASS survey was administered in years 1988, 1994, 2000, 2004, and 2008 to a nationally representative sample of teachers.[11] The dataset provides information about the school district in which a teacher is employed, such as the percent of students eligible for free school lunch, which serves as a measure of the income level of the district. The data also contains information about individual teachers, such as their years of teaching experience, the state in which they teach, and their undergraduate institution. The undergraduate institution is recorded as a Federal Interagency Committee on Education (FICE) code,

---

[10]Public versions of SASS are available at nces.gov. I use the restricted-use SASS sample, which allows me to link teachers to their undergraduate institution. The restricted-use data is available in a secure fashion to researchers who apply for access through nces.ed.gov/pubsearch/licenses.asp. In compliance with NCES disclosure requirements, all sample sizes from raw NCES data reported herein are rounded to the nearest 10.

[11]These numbers refer to the year the school year ended. I follow the same convention throughout the paper. An additional SASS survey was administered in 1991 but did not contain information on teachers' undergraduate institutions, and thus is omitted from the analysis except in the regressions in Section 5.5 examining substitution effects, where I use other dependent variables (salary and emergency certification) found in the SASS survey.

which I merge with data on the average SAT score of entering freshmen for the undergraduate institution of the teacher, from a survey conducted by the Higher Education Research Institute (see Astin, Green, Korn, and Maier (1983) and the description in Angrist and Guryan (2008)). A single observation in the SASS data represents a particular teacher and the selectivity of that teacher's undergraduate institution, which I then link to state certification test laws for initial licensure for the specific year in which the teacher was certified. I then treat this observation as an independent draw from the teacher input quality distribution in a given year and state.[12]

Summary statistics for the main variables of interest from the SASS dataset are displayed in Table 2. The mean and standard deviation of the teacher input quality measure are 916 and 107 SAT points, respectively. In the analysis performed below, I standardize this quality measure so it has mean zero and variance one within the sample, meaning all quality results are reported in terms of standard deviations. I limit the sample to teachers with less than 12 years of experience. I refer to the years in which the SASS survey was administered as survey years. In the analysis below, I group together teachers whom I observe in the same state and year and of the same experience level, and I refer to this group as a cohort. For example, a teacher who began teaching in the 1992-1993 school year and who was surveyed in 1993-1994 school year would be in the cohort of teachers with two years of experience. I drop cohorts with fewer than ten teachers recorded. Table 2 shows that the number of teachers in a state-year-experience level cohort, described below, ranges from 10 to 113, with a mean of 33. The percent of students in the school district eligible for free lunch is 39% on average. Summary statistics for other covariates from the SASS dataset are reported in Appendix Table A.1.

To measure output quality I use public school student test scores from the eighth grade math National Assessment of Educational Progress (NAEP) exam from the NCES.[13] The test was administered in 1990, 1992, 1996, 2000, 2003, 2005, and 2007. The NAEP data contains information on student test scores and other covariates, as well as information on each student's teacher.[14] A single observation in the NAEP

---

[12]Measuring quality in any occupation is difficult, and a source of debate among researchers (Gross 1986). Similar measures of quality to that which I use have been used in Bacolod (2007), Kane and Staiger (2008), Hoxby and Leigh (2004), Figlio (1997), and Angrist and Guryan (2008, 2004). This input quality measure is not meant to capture everything about an individual teacher's qualifications; instead, it is meant to serve as some measure of the impressiveness of a candidate's resume to employers when the candidate enters the job market. It is reasonable to assume that the selectivity of a teacher's undergraduate institution is correlated with the prestige of the teacher's resume.

[13]Public versions of NAEP are available at nces.gov. I use the restricted-use NAEP sample, which allows me to link each student to the number of years of experience of the student's teacher. The restricted-use data is available in a secure fashion to researchers who apply for access through nces.ed.gov/pubsearch/licenses.asp.

[14]The data does not record raw test scores for students, but rather reports five plausible scores calculated using student responses and item response theory (IRT). As explained in Kolstad (2006), these five scores in the NAEP data represent independent draws from the posterior distribution of scores for students with similar item response patterns and similar

data represents a student test score, which I link to state certification test laws for initial licensure for the specific year in which the student's teacher was certified. I then treat this observation as an independent draw from the output quality distribution in a given year and state.[15]

Summary statistics are displayed in the lower half of Table 2. Observe that student scores on the NAEP eighth grade math exam are 262 on average, with a standard deviation of 39. As with input quality, in the analysis below, I standardize these test scores to have mean zero and variance one within the sample, so all results are reported in standard deviation units. As with the SASS survey, I refer to years in which the NAEP test was administered as survey years, and to students being taught by teachers of a given experience level as a cohort. For example, students who were tested in 1994 and who were taught by a teacher who began teaching in the school year ending in 1993 would be in the cohort of students taught by teachers with two years of experience. I drop cohorts with fewer than ten students recorded. The number of students in each state-year-experience cohort ranges from 10 to 1,439, with a mean of 203. The NAEP data also contains the percent of students eligible for free lunch, reported in bins rather than as continuous variables. A school is classified as having either 0–5%, 6–10%, 11–25%, 26–50%, 51–75%, or 76–100% of its students qualifying for free lunch, and the distribution (not shown) is roughly uniform across quartiles. Summary statistics for other variables from the NAEP dataset are found in Appendix Table A.1.

It is impossible to link the SASS and NAEP datasets. This is less problematic in this study than in other contexts, however, because I do not wish to derive a link between input and output quality or to estimate an education production function (Hanushek 1979). Instead, the focus of this study is simply to determine whether or not teacher certification tests have any detectable effect on the distribution of quality—either for input or output quality—and whether average quality effects differ for high vs. low income areas.

I use one additional data source in the regressions in Section 5.5 studying the effects of teacher certification laws on pupil to teacher ratios. This dataset comes from the NCES Common Core of Data

---

observable characteristics. I take the average of the five IRT draws, as in Fryer (2011), and treat this as an output quality draw. See also Jacob (2007) for a discussion.

[15]To measure output quality, it would be desirable to translate student test scores into a measure of teacher value-added by aggregating to the teacher level, as in Kane and Staiger (2008), for example. However, NAEP survey designers claim that the survey is not representative at the teacher level and warn against aggregating to the teacher level (Rogers and Stoeckel 2008). Also, in many cases very few students assigned to a particular teacher appear in the dataset. Finally, as the data is in the form of a repeated cross-section rather than a panel, no baseline measure of student ability is available, which Kane and Staiger (2008) show greatly increases the explanatory power of teacher value-added measures. Therefore, I focus on the distribution of student test scores rather than teacher value-added, and treat each realization of an NAEP test score as a draw from the output quality distribution.

(CCD) from 1987–2010 and is available publicly on the NCES website. Summary statistics are found in Appendix Table A.1.

# 4   Estimating distributional and heterogeneous effects

In this section I present several approaches to identifying the effect of teacher certification tests on quality. To clarify the estimation approach, I focus first on the input quality distribution for first-year teachers and then expand the approach to incorporate other levels of experience. I then explain the approach for estimating the effects on the output quality distribution and then the approach for identifying heterogeneous effects of licensing requirements on quality differing by income level.

## 4.1   Estimating distributional effects of teacher licensing requirements

I focus first on the case of first-year teachers. Recall that the SASS and NAEP were only administered in certain years, and hence records for first-year teachers include only those teachers who began their career during a survey year. Recall also that teacher input quality corresponds to the selectivity of the teacher's undergraduate institution (as measured by the averaged SAT score of entering freshmen). For a fixed $\tau \in (0,1)$, let $q_{st}^{\tau}$ be the $\tau^{th}$ quantile of teacher input quality within state $s$ for teachers who began teaching in year $t$, which I model as

$$q_{st}^{\tau} = \gamma_s(\tau) + \lambda_t(\tau) + Law_{st}'\delta(\tau) + \varepsilon_{st}(\tau) \tag{1}$$

where $\gamma_s(\tau)$ is a state effect and $\lambda_t(\tau)$ is a year effect. State effects capture characteristics that are unchanging over time within a state (e.g., some states may have higher quality teachers than others over the entire sample period). Year effects capture factors that affect every state in a given year (e.g., in 2005, No Child Left Behind legislation affected all states by mandating that each state specify a method for ensuring competence in teachers' subject areas). $Law_{st}$ is a three-element vector containing dummies equal to 1 if a subject test, basic skills test, or professional knowledge test was required in state $s$ in year $t$. $\delta(\tau)$, the parameter vector of interest, represents the effect of teacher certification test laws on the teacher input quality distribution. A positive value of $\delta(\tau)$ at $\tau = 0.1$, for example, would indicate that these occupational licensing laws increase the lower tail of input quality, barring some less-qualified candidates from the occupation. The final term, $\varepsilon_{st}(\tau)$, represents unobserved factors at the state-year level which can affect the $\tau^{th}$ quantile of quality in state $s$ and year $t$.

The identifying assumption of this model is that unobserved state-year specific factors $\varepsilon_{st}(\tau)$ are uncorrelated with the incidence of state certification laws. This assumption would be violated, for example,

if changes in state certification requirements were imposed at the same time as other state-specific policies which also affected the teacher quality distribution. While it is impossible to rule out all possible such unobservables, Angrist and Guryan (2008), examining average effects rather than quantile effects, argue that this exogeneity assumption is reasonable in this setting given that actual licensing requirements and hiring decisions are made at the school district level, not the state level, and districts often enforce a stricter standard than the state requirements, such as requiring teacher testing even when not mandated by the state.[16]

The quantile-based model in (1) is a special case of the general grouped instrumental variables quantile model presented in Chetverikov, Larsen, and Palmer (2015), and I apply here the estimator proposed therein. The grouped quantile estimator represents an extension of the Hausman and Taylor (1981) estimator to quantile regression settings and can be performed in this case by simply computing the $\tau$ quantile of teacher input quality within each state-by-year cell (a "group") and treating this computed quantile as the dependent variable in an OLS regression, with one observation per group, to estimate model (1). Standard quantile regression (Koenker and Bassett 1978) would be inconsistent in this setting given the existence of group-level unobservables, $\varepsilon_{st}$.[17] This approach also has computational advantages over standard quantile regression, which can be quite slow and often fail to converge in standard statistical packages when many right-hand-side variables are included, such as the state and year effects in (1). The grouped quantile approach, on the other hand, handles any group-level right-hand-side variables through a simple OLS regression, greatly reducing computation times.[18]

One robustness test of (1) can be performed by including state-specific linear time trends, yielding the following model:

$$q_{st}^{\tau} = \gamma_{s0}(\tau) + \gamma_{s1}(\tau)t + \lambda_t(\tau) + Law_{st}'\delta(\tau) + \varepsilon_{st}(\tau) \tag{2}$$

---

[16]Angrist and Guryan (2008) regress average teacher quality within a school-district-by-year cell on state-by-year-level requirements, arguing that this state-by-year variation in licensing laws serves as a valid instrument for district-by-year-level requirements, thus treating this state-by-year variation as exogenous, as in this paper.

[17]The standard quantile regression model of Koenker and Bassett (1978) solves the following problem for a given $\tau$:

$$\min_{(\gamma_s(\tau),\lambda_t(\tau),\delta(\tau))} \sum_i \rho_\tau \left( q_{ist} - \gamma_s(\tau) - \lambda_t(\tau) - Law_{st}'\delta(\tau) \right),$$

where $q_{ist}$ represents the individual-level data on teacher $i$ in state $s$ in year $t$ and where $\rho_\tau(\cdot)$ is a "check" function of Koenker and Bassett (1978), i.e. $\rho_\tau(x) = (u - I\{x < 0\})x$. Unlike the grouped quantile approach, standard quantile regression does not allow for additively-separable group-level unobservables ($\varepsilon_{st}(\tau)$) and such unobservables would lead to inconsistency of the standard quantile regression estimator (see Hausman 2001).

[18]I also estimated the standard quantile regression analog of (1) (i.e., assuming $\varepsilon_{st}(\tau) = 0$) and found it to be more than 100 times slower than the grouped approach.

Equation (2) controls for unobserved factors that may affect the quality distribution within a state and that change continuously over time. If equation (1) is well identified, one would expect similar estimates of the effects of licensing on quality from (1) and from (2).

To incorporate data from other experience cohorts, I modify (1) by writing the $\tau^{th}$ quantile of input quality of teachers in state $s$ who had $c$ years of experience when they were surveyed in survey year $t$ as

$$q^\tau_{stc} = \gamma_s(\tau) + \lambda_t(\tau) + \alpha_c(\tau) + Law'_{st}\delta(\tau) + \varepsilon_{stc}(\tau) \tag{3}$$

where $\gamma_s$ and $\lambda_t$ are as above, and $\alpha_c$ represents a cohort fixed effect. $Law_{stc}$ is a three-element vector containing dummies equal to 1 if a subject test, basic skills test, or professional knowledge test was required in state $s$ for teachers surveyed in year $t$ belonging to cohort $c$ when these teachers were initially licensed. Therefore, the test law dummies correspond to the year in which the teacher cohort would have received initial certification. The term $\varepsilon_{stc}(\tau)$ represents unobserved factors at the state-year-experience cohort level that affect the distribution of teacher input quality. Grouping cohorts in this fashion and pooling the data together in estimation forms a "synthetic panel" or "pseudo panel" (see Verbeek and Nijman 1992 or Deaton 1985).

One concern with pooling all experience cohorts is that the experienced teachers, for example, those with five years of experience, who are present in survey year $t$ consist of a group which has self-selected to remain in the profession five years. This is not a problem if the experience/cohort effect does not differ by state, because this effect is controlled for by $\alpha_c$. However, if the effect of experience on quality differs by state, and if this effect is related to policy changes occurring simultaneously with changes in teacher certification requirements, ordinary least squares estimates of $\delta$ would be biased. One specification check I employ in Section 5 below is to replace $\alpha_c$ above with state-specific returns to experience, $\alpha_c + \alpha_s c$, similar to the state-specific linear time trends in (2).

The model in (1) above can be modified to estimate the effect of teacher certification laws on the distribution of output quality (student test scores from the NAEP) by letting $q^\tau_{st}$ be the $\tau^{th}$ quantile of the student test score distribution among students taught by teachers who began teaching in state $s$ and survey year $t$. Equation (3) is similarly modified, with the $q^\tau_{stc}$ being the $\tau^{th}$ quantile of the student test score distribution among students taught by teachers with $c$ years of experience in state $s$ and survey year $t$. Therefore, for students appearing in the NAEP sample in survey year $t$ taught by cohort $c$ in state $s$, the certification test law indicators correspond to the year in which the teaching cohort that taught these students would have received their initial certification.

In the results discussed below in Section 5, some specifications also include an additional vector of state-year observable covariates, $X_{st}$, including student enrollment; the percent of students eligible for free lunch; the percent minority enrollment; the proportion of the state categorized as suburban, city,

or rural; and a quadratic term in the state unemployment rate. Output quality regression models also include the percentage of students who report having an encyclopedia in the home as a simple control for students' educational environment at home.

## 4.2 Estimating heterogeneous effects by income

To estimate the effect of occupational licensing on average quality for school districts of differing income levels, I estimate regressions of quality on an interaction term of test law dummies with the percent of students qualifying for free lunch. These regressions are of the following form:

$$q_{idst} = \gamma_s + \lambda_t + Law'_{st}\delta + Lunch_{dst} * Law'_{st}\psi + Lunch_{dst}\xi + \nu_{idst} \tag{4}$$

where $q_{idst}$ is the input quality of teacher $i$ in school district $d$ in state $s$ and survey year $t$. $Lunch_{dst}$ represents the percent of students in district $d$, in state $s$, in survey year $t$ who qualify for free school lunch. Thus, a higher value of this variable represents a lower income school district. Unlike (1), the regression in (4) is a model of the individual quality of teacher $i$, not of a group-level (state-by-year) measure. Also, unlike (1), the regression in (4) is not a model of conditional quantiles, and hence all unobserved heterogeneity is captured in the term $\nu_{idst}$. Thus, ordinary least squares estimation of this model measures the effect of stricter licensing requirements on average quality for areas of differing income. In the case where all experience cohorts are pooled together, I modify (4) to include cohort effects as in (3). The effect of teacher certification laws on the average input quality of teachers found in a school district in which $\%I$ of students qualify for free lunch is given by $\delta + I\psi$. The parameter $\xi$ captures the main effect of the percent of students qualifying for free lunch.

To examine heterogeneous effects on output quality, I replace $q_{idst}$ with the NAEP test score of student $i$ in school $d$ in state $s$ in survey year $t$.[19] In the NAEP datasets, the percentage of students qualifying for free lunch is not recorded as a continuous variable. Instead, discrete data are recorded, classifying the school as having either 0–5%, 6–10%, 11–25%, 26–50%, 51–75%, or 76–100% of its students qualifying for free lunch. Therefore, rather than containing an interaction with a single continuous variable, the output quality regression is fully saturated with interactions between law dummies and percentage free lunch dummies:

$$q_{idst} = \gamma_s + \lambda_t + \sum_{j=1}^{6} p_{dst}^j * Law'_{st}\psi_j + \sum_{j=1}^{6} p_{dst}^j * \xi_j + \nu_{idst} \tag{5}$$

---

[19]In the NAEP, the percent free lunch variable is recorded at the school level, whereas in the SASS data it is recorded at the district level.

where $p_{dst}^j$ are dummies specifying which of the six free lunch categories contains school $d$ in state $s$ in survey year $t$. The parameters $\xi_j$ capture the main effect of the percent free lunch variables. A similar modification yields the pooled cohorts specification.

As highlighted above, it is important to note that if unobserved factors at the group (state-by-year or state-by-year-by-cohort) level are correlated with changes in teacher certification testing requirements, each of the approaches described above will yield biased estimates of the effects of testing laws on teacher quality. For example, other education policies may have changed simultaneously with teacher certification requirements. If such policy changes did occur, and these policies are complements with teacher certification testing, the measured effects could simply be interpreted as the total effect of the policy changes rather than the specific portion of the effect attributable to certification tests. If, however, unobserved policy changes and certification testing are substitutes, the estimated effects will be difficult to interpret. While I am unable to rule out all possible confounds, below I check the robustness of the results to the inclusion of time and experience trends and present an event study to explore whether the timing of effects is consistent with the changes in licensing laws.

# 5 Results: The effects of licensing on the distribution of quality, and differential effects of licensing by income level

In this section I present the results of the estimation procedure described in Section 4. I first present the grouped quantile regression estimates of the effects of certification test laws on the distributions of input quality (i.e. the selectivity of a teacher's undergraduate institution as measured by the institution's average SAT score for entering freshmen) and output quality (student NAEP test scores). I then present the estimates of the effects of certification testing on input and output quality in regions differing by income level. Finally, I estimate the effects of licensing on outcomes other than quality to examine whether certification tests lead low-income areas to substitute away from hiring licensed teachers.

## 5.1 The effects of licensing on input quality

### 5.1.1 Effects on average input quality

Before addressing the question of whether stricter occupational licensing laws lead to an increase in the lower tail of the distribution of teacher input quality, I first estimate a regression of average teacher input quality on teacher testing laws. The first three columns of Table 3 display the results of estimating (1), i.e. using first-year teachers only, where the dependent variable is replaced with $\bar{q}_{st}$, the mean of input quality

in state $s$ in survey year $t$. Column 1 includes only state and year effects, column 2 adds state-year level demographic controls as explained above, and column 3 adds a state-specific linear time trend. Columns 4–6 repeat the exercise for the pooled sample of teachers. Column 7 adds a term capturing state-specific returns to teacher experience, and column 8 includes both the time trend and the experience trend.

Columns 1–3 suggest that none of the test laws have a statistically significant impact on the average input quality of first-year teachers, which is consistent with the results of Angrist and Guryan (2008). However, because surveys occurred at 4–6 year intervals, estimation with only first-year teachers does not take advantage of all historical variation in teacher certification test laws. The pooled sample of teachers, on the other hand, uses all teacher cohorts, thus using a larger sample and using variation in certification test laws over the entire period (1983–2008).

Column 4 indicates that subject test laws are associated with a statistically significant increase of 0.08 standard deviations in average teacher input quality among the pooled teachers sample. This suggests that, controlling for years of teaching experience, teachers who remain in the occupation multiple years and who were required to pass a subject test prior to initial licensure are more likely to come from highly selective undergraduate institutions than those who did not face this requirement. The basic test law and professional knowledge test laws do not have a significant effect. Columns 5–8 indicate that these findings are robust to the inclusion of demographics, time trends, experience trends, or both.

### 5.1.2  Quantile effects for input quality

To examine how teacher certification laws affect the distribution of input quality, I estimate (1) separately for each decile ($\tau = 0.1, 0.2, ..., 0.9$) of teacher input quality. The results are shown in Figure 4. Panels on the left of Figure 4 (i.e. (a), (c), and (e)) display the effects of subject, basic skills, and professional knowledge test laws at each quantile using the sample of first-year teachers, and panels on the right (i.e. (b),(d), and (f)) use the pooled sample of experienced teachers. All regressions include demographic variables and state and year fixed effects, as in columns 2 and 5 of Table 3. Point-wise 90% confidence intervals are represented in gray.

Estimation using the first-year teacher sample demonstrates a generally downward-sloping quality effect for subject and basic skills tests, implying that, among first-year teachers, certification test laws may have a more positive impact on the lower tail than on the upper tail of quality, although at most quantiles the effect is not significant. For subject test laws, the decrease in the upper tail (the 0.8 quantile) is statistically significant. This decrease in the upper tail of teacher input quality presents some evidence in favor of the hypothesis of Ballou and Podgursky (1998) that occupational licensing can drive away highly qualified candidates. Note that this effect is not possible to detect when looking only at effects

on average quality as in Table 3. Figure 4 does not show strong evidence of certification tests weeding out less-qualified candidates, as the estimates for each sample and test type are not significantly positive at lower quantiles, although the point estimate is positive for the basic skills test at the 0.1 quantile for first-year teachers.

The pooled sample of teachers yields a significant effect of subject test laws on the distribution of teacher qualifications, as shown in panel (b) and at some quantiles of panel (d) of Figure 4. The effect of licensing is positive, implying that the sample of teachers who remain in the occupation for multiple years is more likely to come from highly selective undergraduate institutions when subject test laws are in place than when they are not. Interestingly, this result is relatively flat across the distribution, indicating that subject test laws do not appear to differentially affect the lower tail of the distribution and upper tail of the distribution. To examine the robustness of these results, Appendix Figure A.1 displays the same results as in panel (b) of Figure 4, that is, the effect of subject test laws using the pooled sample of teachers, but with time and experience trends included. The results do not change drastically with the inclusion of these trends, consistent with treatment and control states not differing drastically in their input quality trends prior to changes in licensing laws.

This finding for the pooled sample of teachers is in contrast to the Ballou and Podgursky (1998) hypothesis and the results in panel (a) of Figure 4, which found a decrease in the upper tail of input quality due to licensing among first-year teachers. Together, these results suggest that stricter licensing requirements may drive away some highly qualified *first-year* teachers in the short run but have a positive longer-run effect on the quality. That is, if licensing drives out high-quality first-year teachers, panel (b) suggests that these teachers would likely not have remained in the occupation anyway.

## 5.2   The effects of licensing on output quality

### 5.2.1   Effects on average output quality

Before estimating the effects of occupational licensing on the distribution of output quality, I first estimate (1) with the dependent variable being the average eighth grade math NAEP score within a given group (state-by-year or state-by-year-by-cohort). The results are displayed in Table 4. For the pooled sample columns, the dependent variable is the mean among students taught by teachers in cohort $c$ in state $s$ in survey year $t$. Subject test laws appear to have no effect, while basic skills test laws appear to have an effect of about 0.05 standard deviations, but this effect is not robust to the inclusion of time trends. Professional knowledge test laws appear to have a small but imprecisely-measured negative effect. As a useful benchmark, other interventions in education, such as the class size experiments of Krueger (1999), have led to increases in student test scores of approximately 0.2 standard deviations.

17

### 5.2.2　Quantile effects for output quality

The effects on the distribution of output quality (NAEP student test scores) are displayed in Figure 5. The grouped quantile approach makes it possible to study effects on the entire distribution, as advocated by Brown and Saks (1975) and the ensuing literature, which recommended examining the variance of student outcomes rather than simply the mean. As above, the results in Figure 5 come from estimating (1) at each decile of output quality, controlling for demographics and state and year effects and, in the case of the pooled sample, experience effects. Panel (a) demonstrates that, in the first-year teacher sample, subject test laws are associated with a 0.15 standard deviation drop in the lower tail of the distribution. The effect also appears to be increasing with the quantile. In the pooled sample, panel (b), the effect on the lower tail is smaller and insignificant, but the increasing shape is preserved, and a positive impact of 0.05 standard deviations is detected at the 0.9 quantile. Panels (c) and (d) demonstrate that basic skills test laws increase test scores by 0.05–0.10 standard deviations across most of the distribution, while panel (f) shows a marginally significant negative impact of 0.05 standard deviations for professional knowledge test laws on the middle and far right quantiles. Overall, Figure 5 suggests that most of the positive effects of these teacher certification tests accrue to the upper half of the distribution, implying a widening of the student test score distribution. While is not obvious why this widening would occur, one possibility is these certification tests screen teaching candidates on abilities best suited for providing value to students who perform at or above the median—a question meriting further research. Neal and Schanzenbach (2010) documented related results which suggest that educational interventions such as changes in proficiency requirements may lead teachers to focus on students at certain points of the test score distribution.

Appendix Figure A.2 checks the robustness of the effects of subject test laws and basic skills test laws in the pooled sample. The subject test results remain similar even with the inclusion of experience trends. The basic skills test results change shape somewhat when state specific time trends are included, in particular in the upper tail, suggesting that in the pooled teachers sample the argument for a causal relationship of basic skills test laws on student test scores is less strong than for subject test laws as treated vs. untreated states appear to not be experiencing parallel output quality trends prior to changes in basic skills test laws.

## 5.3 Heterogeneous effects of licensing on input quality for high vs. low-income areas

The next question of interest is whether occupational licensing laws differentially affect areas of differing income levels. Shapiro (1986) and Friedman and Friedman (1962) predict that licensing may result in higher quality for high income consumers and lower quality for low income consumers. I test this prediction by estimating (4). The results of this estimation for input quality are displayed in Table 5 and Figure 6. First, Table 5 displays the results when the teacher testing law effect is evaluated at the mean of the percent-free-lunch variable (i.e. $\hat{\delta} + \bar{I}\hat{\psi}$, where $\bar{I}$ is the mean of percent-free-lunch). Standard errors, clustered at the state-year level, are reported in parentheses. In the pooled teacher sample, subject test laws appear to have a positive impact of about 0.06 standard deviations, similar to the average effect measured in Table 3 above. In the first-year teacher sample, in column 3, basic skills test laws appear to have a large negative effect on the average teacher input quality, but this effect is only significant when time trends are included. Other estimates appear relatively similar with or without time and experience trends. As explained above, each of these regressions also includes the main effect of the percent free lunch variable, not shown; the estimated coefficient on this main effect is significant and negative, implying that teachers in lower-income areas tend to come from less-selective universities.

Figure 6 displays the estimated average effect of certification test laws on input quality evaluated at each quantile of the percent-free-lunch variable. Note that, unlike the figures of distributional effects in Sections 5.1 and 5.2, all of the points plotted in panels (a),(c), and (e) of Figure 6 come from the same regression using first-year teachers, with the different estimates representing the effect of teacher testing laws evaluated at different quantiles of the percent-free-lunch distribution. Similarly, all of the points plotted in panels (b), (d), and (f) come from the a single regression using the pooled sample.

Figure 6 indicates that stricter occupational licensing laws are associated with an increase in teacher qualifications, but only at the wealthiest school districts—those with a lower percentage of students who qualify for free lunch. Panel (b) indicates that, among experienced teachers, at the 0.1 quantile of the percent free lunch variable, subject tests are associated with an increase of nearly 0.15 standard deviations in average teacher qualifications. This effect decreases as income decreases, as shown by movement to the right on the horizontal axis. For the poorest districts shown, those at the 0.9 quantile, the point estimate of the effect of subject test laws is negative (although insignificant). Panel (f) of Figure 6 demonstrates that professional knowledge test laws have the same relationship as subject test laws, leading to increases in teacher input quality in wealthier areas of approximately 0.07 standard deviations and a decrease of similar magnitude for the lowest income districts. The remaining panels display a similar downward slope but the relationship is imprecisely measured.

19

These results are consistent with the predictions of Shapiro (1986) and Friedman and Friedman (1962), that any improvements in quality due to stricter licensing will accrue primarily to high-income areas at the expense of low-income areas. One potential cause of this differential treatment effect would be that stricter licensing is associated with higher wages and that schools in high-income areas are more capable of increasing wages in response to changes in licensing and thus can continue to attract higher quality teachers, as explored in Section 5.5 below. Another possibility is that higher quality teachers simply prefer higher-income areas for a variety of reasons, as suggested by Boyd, Lankford, Loeb, and Wyckoff (2005a,b). Regardless of the cause, these findings suggest a positive relationship between the impact of licensing stringency and the wealth of the school district.

## 5.4   Heterogeneous effects of licensing on output quality for high vs. low-income areas

The effect of licensing on student test scores for areas of differing income levels is displayed in Figure 7. A table analogous to Table 5 is omitted because the mean value of the percent free lunch variable is not reported in the data; instead, the percent free lunch variable is reported in six bins. 90% confidence intervals are displayed around each estimate, where standard errors are calculated by clustering at the state-year level.

In Figure 7, the first-year teacher sample displays noisily measured effects, while the pooled sample shows stronger patterns. For example, panel (a) does not display a significant effect of subject test laws for the wealthiest schools, and displays a significant negative effect for schools in the second quartile of income (26–50%). In the pooled sample, subject and basic skills test laws are associated with a larger increase in test scores (0.05 standard deviations for subject test laws and 0.15 standard deviations for basic skills test laws) for higher income districts. Appendix Figure A.3 focuses on the basic skills test results and demonstrates that they are robust to the inclusion of trends.

In Figure 7, professional knowledge test laws display the opposite effect of basic skills test, with a large and significant decrease of 0.15 standard deviations for the wealthiest districts. While the cause of this opposite effect is unclear, it may be that the pedagogy skills required for these exams are more valuable in low-income areas.

## 5.5   Substitution away from licensed professionals

Given the results reported thus far, it is natural to ask how, if at all, licensing induces consumers (in this case, schools and school districts) to substitute away from licensed professionals. Previous studies (e.g.

Angrist and Guryan 2008 and others cited in Section 2) have demonstrated that more stringent licensing requirements, including teacher certification tests, lead to higher wages. This is as theory would predict: licensing requirements restrict supply and hence increase wages. If it is the case that higher income schools are better equipped than lower income schools to raise wages in response to this restriction of supply, lower income schools might be expected to respond to stricter occupational licensing by substituting away from licensed professionals. Specifically, lower income schools could increase class sizes or hire more emergency-certified teachers, i.e. those who did not have to pass certification tests. For example, Jepsen and Rivkin (2009) found some evidence that lower income schools were less able to retain teachers during supply shortages. The authors studied California's 1996 law to reduce class sizes, finding that the law led to a shortage of teachers. Many teachers left low-income schools to fill coveted vacancies at higher income schools. Below I examine effects of certification testing on emergency-certified hirings, class sizes, and wages in high vs. low-income areas.

To study whether teacher certification tests lead to larger increases in the use of emergency-certified teachers in low-income areas than in high-income areas, I turn again to the SASS survey, which contains an indicator for whether or not a given teacher was emergency certified. Summary statistics for this variable as well as the other variables introduced in this section appear in Appendix Table A.1. Emergency certified teachers do not meet the same testing requirements as other teachers and are hired when a school is unable to meet staffing requirements using fully certified teachers. I estimate (4) with the dependent variable replaced with the emergency-certified indicator. The results are displayed in the first column of Figure 8. None of the results are significant, although for basic skills and professional knowledge test laws the point estimates are increasing in the percent free lunch variable, consistent with the idea of low-income areas being more likely than high income areas to increase the hiring of emergency certified teachers in response to stricter occupational licensing.

I explore changes in class size in the second column of Figure 8, which displays the results of estimating (4) with the dependent variable replaced by the pupil-to-teacher ratio as recorded in the Common Core of Data (CCD) from the NCES, using all school districts in the database from the years 1988 to 2010.[20] Panel (b) demonstrates that there is not a significant effect of subject test laws on class size. Panel (d) indicates that there is a significant increase of approximately 0.25 in the student to teacher ratio at the lowest-income school districts, those in which 80% or more qualify for free school lunch. At higher-income schools, a marginally significant decrease in the pupil to teacher ratio of about 0.25 occurs. This is consistent with the hypothesis that occupational licensing may lead to larger class sizes at these schools.

---

[20]I include the years 2009–2010 as well because, unlike the SASS and NAEP datasets, the CCD contains data for these years, and my compiled certification test law data also contains these years. Years prior to 1988 are not available in the CCD for the variables I study.

Professional knowledge tests, on the other hand, are associated with larger class sizes (an increase of 0.3-0.4 in the student to teacher ratio) at both high and low income schools.

Figure 9 demonstrates some weak evidence that basic skills test laws tend to increase wages primarily in high income areas. The estimates in this figure come from (4) with the dependent variable replaced by the log of teacher salary (in 2008 dollars) for teachers with a bachelor's degree, as reported in the SASS survey years. Panel (b) shows that, in higher-income schools, teacher salary increases with basic skills test laws by approximately 2%. In lower-income schools, those with a higher percentage of students qualifying for free lunch, the estimated increase is not statistically significant. The significant results for basic skills test laws, however, were not robust to the inclusion of state time trends (not shown). The results are difficult to interpret given that Figure 6 panel (d) showed that basic skills test laws did not have a significant effect on teacher qualifications in areas varying by income. In panels (a) and (c), estimates of the effects of subject and professional knowledge test laws are not significant. Note, however, that these salary regressions use much less variation in licensing regimes, as salary data is only recorded in survey years.

## 5.6 Event study

I now introduce an event study framework to test whether the observed changes in input and output quality occurred *after* the changes in occupational licensing laws. This event study focuses solely on subject test law changes because, as demonstrated in Table 1, few states eliminated the requirement of a subject test for certification once this requirement was in place, making it simple to determine the year in which the law change took place. Moreover, certification tests of subject matter have been the focus of legislation (Kuenzi 2009).

For state $s$ in year $t$, let $d_{st}^{(-2)}$ be a dummy variable that is equal to 1 if year $t$ was two years prior to the year in which state $s$ changed its subject test law. For one year prior to a law change, $d_{st}^{(-1)}$ is defined similarly. Let $d_{st}^{(0)}$ be an indicator for whether state $s$ changed its subject test law in year $t$. The variables $d_{st}^{(1)}$, $d_{st}^{(2)}$, $d_{st}^{(3)}$, and $d_{st}^{(4+)}$ are dummies for year $t$ being 1, 2, 3, or 4 or more years after state $s$ changed its law. If a state never instituted a subject test law, all of the dummies are set to zero. Let the vector of these dummies be given by

$$d_{st} = [d_{st}^{(-2)} \ d_{st}^{(-1)} \ d_{st}^{(0)} \ d_{st}^{(1)} \ d_{st}^{(2)} \ d_{st}^{(3)} \ d_{st}^{(4+)}]'.$$

To estimate the effects of subject test laws on the distribution of teacher input quality in the framework of this event study, I estimate (1) replacing the subject test dummy in the $Law_{st}$ vector with the vector of event dummies, $d_{st}$. I modify the pooled regression model similarly.

The results are displayed in Appendix Figure A.4. I estimate nine regression, one regression for each decile ($\tau = 0.1, 0.2, ...0.9$), with each decile displayed in its own panel of Appendix Figure A.4. Only the pooled teacher sample results are reported. Thus, the results in Appendix Figure A.4 are an event study version of those in panel (b) of Figure 4. The results demonstrate that in the two years prior to the law change, the effect is zero or insignificant at each quantile of input quality. In the year of the law change and the two years immediately following the change, there is also no significant effect. However, at three years and at four or more years after the law change, there is a significant, positive effect on the distribution at many of the quantiles. This lends support to the evidence from Section 5.1, which suggested that subject test laws raise the distribution of teacher input quality among experienced teachers (demonstrating that the *cause*—the change in subject test law—appears to occur prior to the *effect*—the change in the distribution of teacher qualifications). In particular, the principal effect of the law change appears to occur several years after the law is first implemented. This phenomenon may arise from announcements of law changes affecting decisions of undergraduates to enter the teaching occupation or from a phase-in period or lag in effective enforcement of new certification laws.

Appendix Figure A.5 displays the same estimation but for output quality. That is, these results can be compared with those in panel (b) of Figure 5. Recall that Figure 5 showed a positive impact of 0.05 standard deviations at the 0.9 quantile for the pooled teachers sample, with insignificant results at other quantiles. A similar pattern is seen in Appendix Figure A.5, with the 0.9 quantile, panel (i), demonstrating a marginally significant effect, with the effect arriving at approximately two years after the law change.

To apply this event study to the measurement of heterogeneous effects by income, I estimate (4) with the subject test law dummy in the $Law_{st}$ variable replaced by the event dummies, $d_{st}$, and similarly for the pooled sample model. I interact each of the event dummies with the percent free lunch variable. The results of this estimation for the pooled sample are displayed in Appendix Figure A.6. Once again, at most quantiles, no significant effects are measured prior to or coincident to the law change. Rather, the effects appear to occur four or more years after the change takes place.

To examine the timing of the effects of subject test laws on output quality, I estimate (5) including interactions of the percent free lunch dummies with the event study dummies in $d_{st}$. The results are displayed in Appendix Figure A.7. The effects of the subject law change are larger for the wealthier areas (panels (a) and (b)), and these effects arrive mainly one and two years after the law change took place.

23

# 6  Conclusion

This study used state-level variation in teacher certification testing requirements over the period from 1983 to 2008 to examine how more stringent occupational licensing standards—in the form of teacher certification tests prior to initial licensure—affect the composition of candidates who meet these standards and enter the occupation (input quality, measured here by the selectivity of a teacher's undergraduate institution), and the distribution of the output in the profession (output quality, measured here by student test scores). I found that subject area certification tests appear to raise the distribution of input quality among teachers who remain in the occupation multiple years, but decrease the upper tail of input quality among first-year teachers. The distribution of output quality appears to widen when certification test laws are in place. I also found that increases in input or output quality due to stricter licensing accrue primarily to high-income areas; low-income areas experience little or no benefit from stricter teacher licensing requirements.

The results of this study suggest several avenues for future research. First, the empirical findings can provide guidance for theoretical models of the how teacher certification exams affect outcomes. For example, the results suggest that models of the process through which certification tests screen candidates of different quality levels should take into account candidates' heterogeneity in the costs of passing exams, heterogeneity in outside career options, or heterogeneity in preferences for a teaching career. Given that the effects of licensing differed for teachers who only remained in the profession for a short time period, models should also consider heterogeneity in teachers' propensity to exit the profession. Finally, the results suggest a role for modeling theoretically the link between the effects of licensing on the quality distribution and the effects of licensing on high vs. low income areas and how schools may compete to attract high-quality teachers.

Although the focus of this study was on the teaching profession, the questions and approaches employed here can play a valuable role in the study of other occupations as well. In particular, the licensing literature to date has focused on the effect of licensing on changes in average levels of quality, while the stated purpose of licensing by many proponents is instead to provide a minimum level of quality that candidates and services must meet. Future research could look for distributional effects in other occupations, especially those in which consumer safety or asymmetric information of professionals vs. consumers is a larger concern, such as medicine or law, to see if licensure laws do in fact prevent the worst candidates from entering the occupation or prevent the worst outcomes from occurring.

Future research could also aid in determining the effects of occupational licensing on heterogeneous consumers, examining whether these laws tend to harm low-income consumers relative to high-income consumers. This study sheds some light on this question, as well as on how low-income areas may substi-

tute away from licensed professionals, indicating that policymakers may do well to consider potentially unintended consequences of licensing on low-income areas. However, the question merits further research, again particularly in professions providing services with a greater safety or informational concern, or in occupations in which safety concerns are small, as licensing regulations in these occupations may be more likely to result from professionals' rent-seeking behavior.

# References

AACTE (1990, 1991, 1993-1996): *Teacher Education Policy in the States: A 50-State Survey of Legislative and Administrative Actions.* American Association of Colleges for Teacher Education, Washington, DC, Six certification manuals.

AKERLOF, G. (1970): "The market for 'lemons': Quality uncertainty and the market mechanism," *Quarterly Journal of Economics*, 84(3), 488–500.

ANGRIST, J., AND J. GURYAN (2004): "Teacher testing, teacher education, and teacher characteristics," *American Economic Review Papers and Proceedings*, pp. 241–246.

——— (2008): "Does teacher testing raise teacher quality? Evidence from state certification requirements," *Economics of Education Review*, 27(5), 483–503.

ARROW, K. (1963): "Uncertainty and the welfare economics of medical care," *American Economic Review*, 53(5), 941–973.

ASTIN, A., K. C. GREEN, W. S. KORN, AND M. J. MAIER (1983): *The American Freshman: National Norms for Fall 1983.* Cooperative Institutional Research Program of the American Council on Eduaction, Higher Education Research Institute, Graduate School of Education, UCLA.

BACOLOD, M. (2007): "Do alternative opportunities matter? The role of female labor markets in the decline of teacher quality," *Review of Economics and Statistics*, 89(4), 737–751.

BALLOU, D., AND M. PODGURSKY (1998): "The case against teacher certification," *Public Interest*, pp. 17–29.

BERGER, M., AND E. TOMA (1994): "Variation in state education policies and effects on student performance," *Journal of Policy Analysis and Management*, 13(3), 477–491.

BOYD, D., H. LANKFORD, S. LOEB, AND J. WYCKOFF (2005a): "The draw of home: How teachers' preferences for proximity disadvantage urban schools," *Journal of Policy Analysis and Management*, 24(1), 113–132.

——— (2005b): "Explaining the short careers of high-achieving teachers in schools with low-performing students," *American Economic Review*, pp. 166–171.

BOYDSTON, J. E. (1995-2010): *Teacher Certification Requirements in All 50 States: How and Where to Get a Teaching Certificate.* Teacher Certification Publications, Sebring, FL, Fifteen cerfication manuals from 1995 to 1999 and from 2001 to 2010.

BROWN, B. W., AND D. H. SAKS (1975): "The production and distribution of cognitive skills within schools," *Journal of Political Economy*, pp. 571–593.

CAMPO-FLORES, A. (2011): "In Florida, interior designers have designs on deregulation," *Wall Street Journal*, Posted April 15, 2011. Retrieved from http://online.wsj.com.

CARROLL, S., AND R. GASTON (1981): "Occupational restrictions and the quality of service received: Some evidence," *Southern Economic Journal*, pp. 959–976.

CHETVERIKOV, D., B. LARSEN, AND C. PALMER (2015): "IV quantile regression for group-level treatments, with an application to the distributional effects of trade," NBER Working Paper 21033.

COLEY, R. J., AND M. GOERTZ (1990): *State Educational Standards in the 50 States: 1990*. Educational Testing Service, Princeton, NJ.

CURRIE, J., AND V. J. HOTZ (2004): "Accidents will happen? Unintended childhood injuries and the effects of child care regulations," *Journal of Health Economics*, 23(1), 25–59.

DARLING-HAMMOND, L. (1997): *Doing What Matters Most: Investing in Quality Teaching*. National Commission on Teaching and America's Future, Kutztown, PA.

DEATON, A. (1985): "Panel data from time series of cross-sections," *Journal of Econometrics*, 30, 109–126.

DUNCAN, A., AND E. OCHOA (2002): *The Secretary's Seventh Annual Report On Teacher Quality: A Highly Qualified TeacherIn Every Classroom*. US Department of Education, Office of Postsecondary Education, Washington, DC.

FEDERMAN, M., D. HARRINGTON, AND K. KRYNSKI (2006): "The impact of state licensing regulations on low-skilled immigrants: The case of Vietnamese manicurists," *American Economic Review*, 96(2), 237–241.

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

FRIEDMAN, M., AND R. FRIEDMAN (1962): *Capitalism and Freedom*. University of Chicago press.

FRYER, R. G. (2011): "Racial inequality in the 21st century: The declining significance of discrimination," in *Handbooks in Economics, Handbook of Labor Economics, Vol 4B*, ed. by O. Ashenfelter, and D. Card, pp. 855 – 971. Elsevier.

GARDNER, D. (1983): *A Nation At Risk*. The National Commission on Excellence in Education, Washington, DC.

GODDARD, R. E. (1983-1993): *Teacher Certification Requirements in All 50 States: How and Where to Get a Teaching Certificate*. Teacher Certification Publications, Grover Hill, OH, Eleven cerfication manuals from 1983 to 1993.

GOLDHABER, D., AND D. BREWER (2000): "Does teacher certification matter? High school teacher certification status and student achievement," *Educational Evaluation and Policy Analysis*, 22(2), 129.

GROSS, S. (1986): "Professional licensure and quality: The evidence," Cato Policy Analysis no. 79. Cato Institute, Washington, D.C.

HANUSHEK, E., AND R. PACE (1995): "Who chooses to teach (and why)?," *Economics of Education Review*, 14(2), 101–117.

HANUSHEK, E. A. (1979): "Conceptual and empirical issues in the estimation of educational production functions," *Journal of Human Resources*, 14(3).

HAUSMAN, J. (2001): "Mismeasured variables in econometric analysis: Problems from the right and problems from the left," *Journal of Economic Perspectives*, 15(4), 57–67.

HAUSMAN, J., AND W. E. TAYLOR (1981): "Panel data and unobservable individual effects," *Econometrica*, pp. 1377–1398.

HOXBY, C. M., AND A. LEIGH (2004): "Pulled away or pushed out? Explaining the decline of teacher aptitude in the United States," *American Economic Review*, pp. 236–240.

JACOB, B. A. (2007): "Test-based accountability and student achievement: An investigation of differential performance on NAEP and state assessments," *NBER Working Paper 12817*.

JEPSEN, C., AND S. RIVKIN (2009): "Class size reduction and student achievement," *Journal of Human Resources*, 44(1), 223–250.

KANE, T., J. ROCKOFF, AND D. STAIGER (2008): "What does certification tell us about teacher effectiveness? Evidence from New York City," *Economics of Education Review*, 27(6), 615–31.

KANE, T., AND D. STAIGER (2008): "Estimating teacher impacts on student achievement: An experimental evaluation," *NBER Working Paper 14607*.

KLEINER, M. M. (2000): "Occupational licensing," *Journal of Economic Perspectives*, 14(4), 189–202.

——— (2006): *Licensing Occupations: Ensuring Quality or Restricting Competition?* WE Upjohn Institute for Employment Research, Kalamazoo, MI.

KLEINER, M. M., AND A. B. KRUEGER (2010): "The prevalence and effects of occupational licensing," *British Journal of Industrial Relations*, 48(4), 676–687.

KLEINER, M. M., AND R. T. KUDRLE (2000): "Does regulation affect economic outcomes? The case of dentistry," *Journal of Law and Economics*, 43(2), pp. 547–582.

KLEINER, M. M., AND D. PETREE (1988): "Unionism and licensing of public school teachers: Impact on wages and educational output," in *When Public Sector Workers Unionize*, ed. by R. B. Freeman, and C. Ichniowski. Chigago: University Of Chicago Press.

KOENKER, R., AND G. BASSETT (1978): "Regression quantiles," *Econometrica*, pp. 33–50.

KOLSTAD, A. (2006): *Psychometric procedures used in NAEP*. Presentation at the NAEP Database Training Seminar, Washington, DC.

KRUEGER, A. B. (1999): "Experimental estimates of the education production function," *Quarterly Journal of Economics*, 114(2), 497–532.

KUENZI, J. J. (2009): "A highly qualified teacher in every classroom: Implementation of the No Child Left Behind Act and reauthorization issues for the 111th Congress," Congressional Research Service, Report RL33333.

KUGLER, A., AND R. M. SAUER (2005): "Doctors without borders? Relicensing requirements and negative selection in the market for physicians," *Journal of Labor Economics*, 23(3), 437–465.

LAW, M. T., AND M. S. MARKS (2009): "Effects of occupational licensing laws on minorities: Evidence from the progressive era," *Journal of Law and Economics*, 52(2), pp. 351–366.

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

LITAN, R. (2015): "In Obama's budget, an effort to rein in occupational licensing," *Wall Street Journal*, Url: http://blogs.wsj.com/washwire/2015/02/09/in-obamas-budget-an-effort-to-rein-in-occupational-licensing/ Posted Feb. 9, 2015. Retrieved Apr. 20, 2015.

MAURIZI, A. (1980a): "The impact of regulation on quality: The case of California contractors," in *Occupational Licensure and Regulation*, ed. by S. Rottenberg, pp. 299–333. American Enterprise Institute for Public Policy Research, Washington, DC.

———— (1980b): "Mandatory licensure of registered nurses: Introduction and impact," in *Occupational Licensure and Regulation*, ed. by S. Rottenberg, pp. 47–72. American Enterprise Institute for Public Policy Research, Washington, DC.

MEHTA, J. (2013): "Teachers: Will we ever learn?," *New York Times*, Url: http://www.nytimes.com/2013/04/13/opinion/teachers-will-we-ever-learn.html. Posted Apr. 12, 2013. Retrieved Apr. 15, 2014.

NADLER, D., AND P. E. PETERSON (2009): "What happens when states have genuine alternative certification? We get more minority teachers and test scores rise," *Education Next*, Url: http://educationnext.org/what-happens-when-states-have-genuine-alternative-certification/. Posted Winter 2009. Retrieved Apr. 15, 2014.

NASDTEC (1991, 1994, 1996, 1998, 2000-2002): *National Association of State Directors of Teacher Education Certification (NASDTEC) Manual on the Preparation and Certification of Education Personnel*. Kendall/Hunt Publishing Company, Dubuque, IA, Seven certification manuals.

NASDTEC KNOWLEDGEBASE (2003–2010): *National Association of State Directors of Teacher Education Certification (NASDTEC) Manual on the Preparation and Certification of Education Personnel and NASDTEC Knowledgebase*. Kendall/Hunt Publishing Company, Dubuque, IA, As published on National Center for Education Statistics (NCES) website: http://nces.ed.gov/programs/digest/. Eight tables from years 2003-2010.

NCTAF (1996): *Teaching for Americas Future*. National Commission on Teaching and America's Future, New York.

NEAL, D., AND D. W. SCHANZENBACH (2010): "Left behind by design: Proficiency counts and test-based accountability," *Review of Economics and Statistics*, 92(2), 263–283.

NEUWEILER, H. B. (1988): *Teacher Education Policy in the States: A 50-State Survey of Legislative and Administrative Actions*. The American Association of Colleges for Teacher Education, Washington, DC.

PAIGE, R., S. SROUP, AND J. ANDRADE (2002): *Meeting the Highly Qualified Teachers Challenge: The Secretary's Annual Report on Teacher Quality*. US Department of Education, Office of Postsecondary Education, Washington, D.C.

PELTZMAN, S. (1976): "Toward a more general theory of economic regulation," *Journal of Law and Economics*, 19, 211–240.

RAMSEYER, J. M., AND E. B. RASMUSEN (2015): "Lowering the bar to raise the bar: Licensing difficulty and attorney quality in Japan," *Journal of Japanese Studies*, 41(1), 113–142.

ROCKOFF, J., B. JACOB, T. KANE, AND D. STAIGER (2008): "Can you recognize an effective teacher when you recruit one?," *Education Finance and Policy*, 6(1), 43–74.

ROGERS, A., AND J. J. STOECKEL (2008): *NAEP 2007 Mathematics, Reading, and Writing Assessments Restricted-Use Data Files Data Companion.* U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics., Washington, DC.

RUDNER, L. (1987): *What's Happening in Teacher Testing: An Analysis of State Teacher Testing Practices.* US Dept. of Education, Office of Educational Research and Improvement, Educational Resources Information Center.

SASS, T. R. (2011): "Certification requirements and teacher quality: A comparison of alternative routes to teaching," Working Paper, Georgia State University.

SHAPIRO, C. (1986): "Investment, moral hazard, and occupational licensing," *Review of Economic Studies*, 53(5), 843.

SMITH, A. (1776): *An Inquiry into the Nature and Causes of the Wealth of Nations.*

STIGLER, G. (1971): "The theory of economic regulation," *Bell Journal of Economics and Management Science*, II, 3–21.

TENN, S. (2001): "Three Essays on the Relationship Between Migration and Occupational Licensing," Ph.D. thesis, University of Chicago, Dept. of Economics.

VERBEEK, M., AND T. NIJMAN (1992): "Can cohort data be treated as genuine panel data?," *Empirical Economics*, 17(1), 9–23.

WEDDLE, E. (2014): "Teachers object to weakened licensing rules," *Indianapolis Star*, Url: http://www.indystar.com/story/news/2014/01/14/teachers-object-to-weakened-licensing-rules/4480923/. Posted Jan. 14, 2014. Retrieved Apr. 15, 2014.

WISWALL, M. (2007): "Licensing and occupational sorting in the market for teachers," *Working Paper, NYU Department of Economics.*

Figure 1: Illustration of possible changes to quality distribution
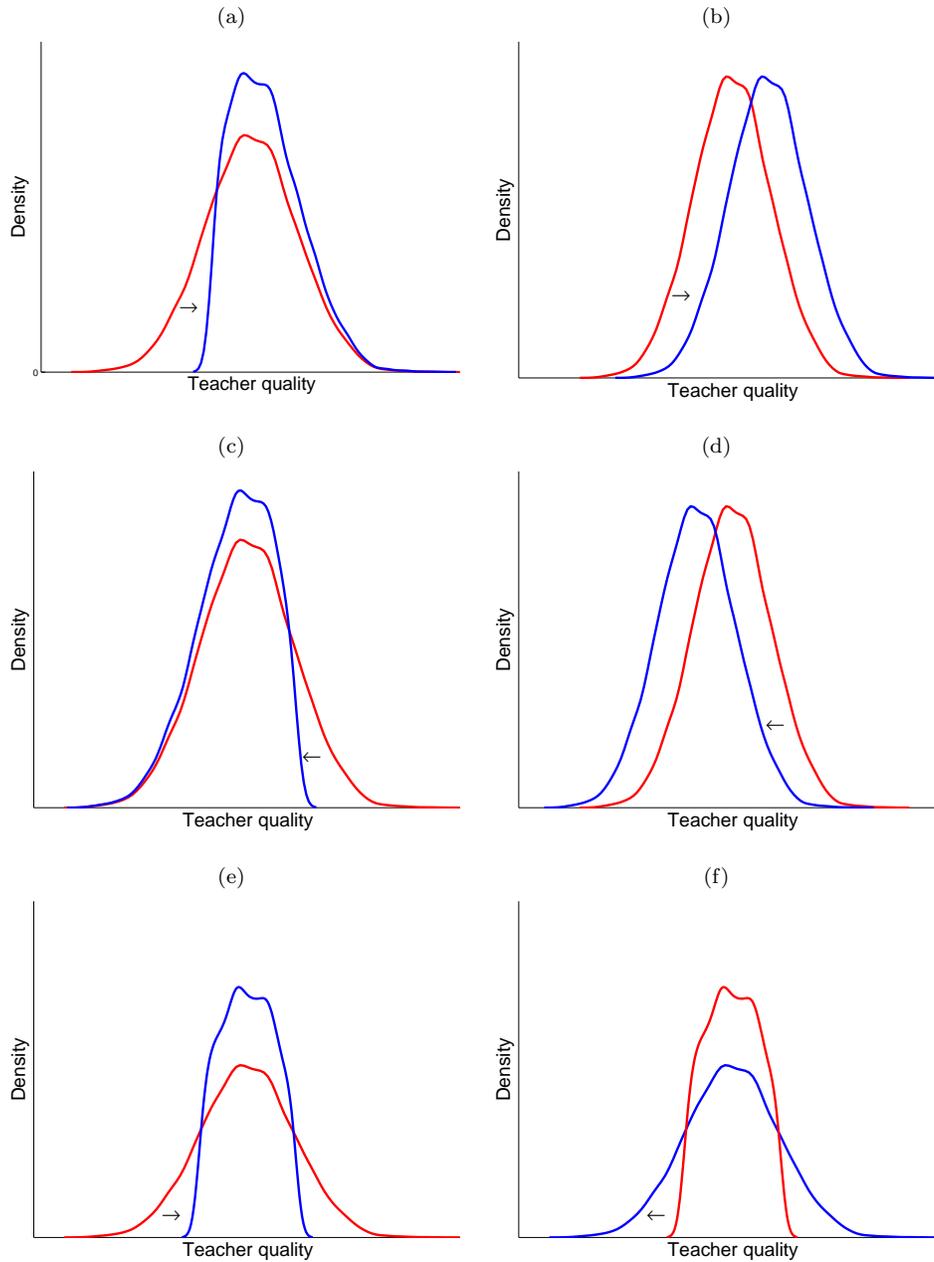


Notes: Figure illustrates potential effects of an increase in licensing stringency on the distribution of quality. In each panel licensing causes a shift from the red quality distribution to the blue quality distribution. Panels (a) and (b) both display an *increase* in average quality; in (a) this occurs through an improvement in the lower tail and in (b) through an improvement in the whole distribution. Panels (c) and (d) both display a *decrease* in average quality; in (c), this occurs through a decrease in the upper tail and in (d) through a decrease in the whole distribution. Panels (e) and (f) display cases where the mean is unchanged but licensing tightens the distribution in (e) and widens the distribution in (f).

Figure 2: Illustration of possible heterogeneous effects of licensing by income level
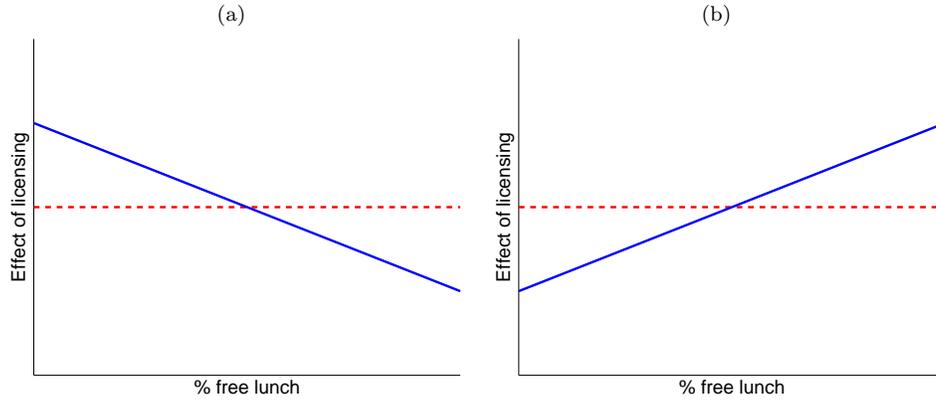


Notes: Figure displays potential heterogeneous effects (the solid blue line) of licensing by income level relative to a constant average effect (the dashed red line). In (a), at higher income levels (lower percent free lunch), the effects of licensing are larger and at lower income levels (higher percent free lunch) the effects of licensing are smaller. In (b), at higher income levels (lower percent free lunch), the effects of licensing are smaller and at lower income levels (higher percent free lunch) the effects of licensing are larger.

Figure 3: Proportion of states requiring certification tests, 1983-2010.
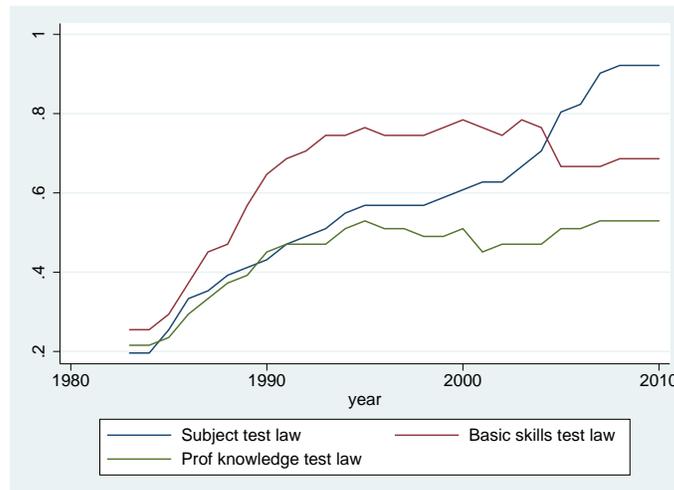


Notes: Figure displays (in blue) the proportion of states in each year which required teachers to pass a subject test prior to initial licensure. A similar measure is shown in red for basic skills test requirements and in green for professional knowledge test requirements.

Figure 4: Effects of certification test laws on input quality distribution
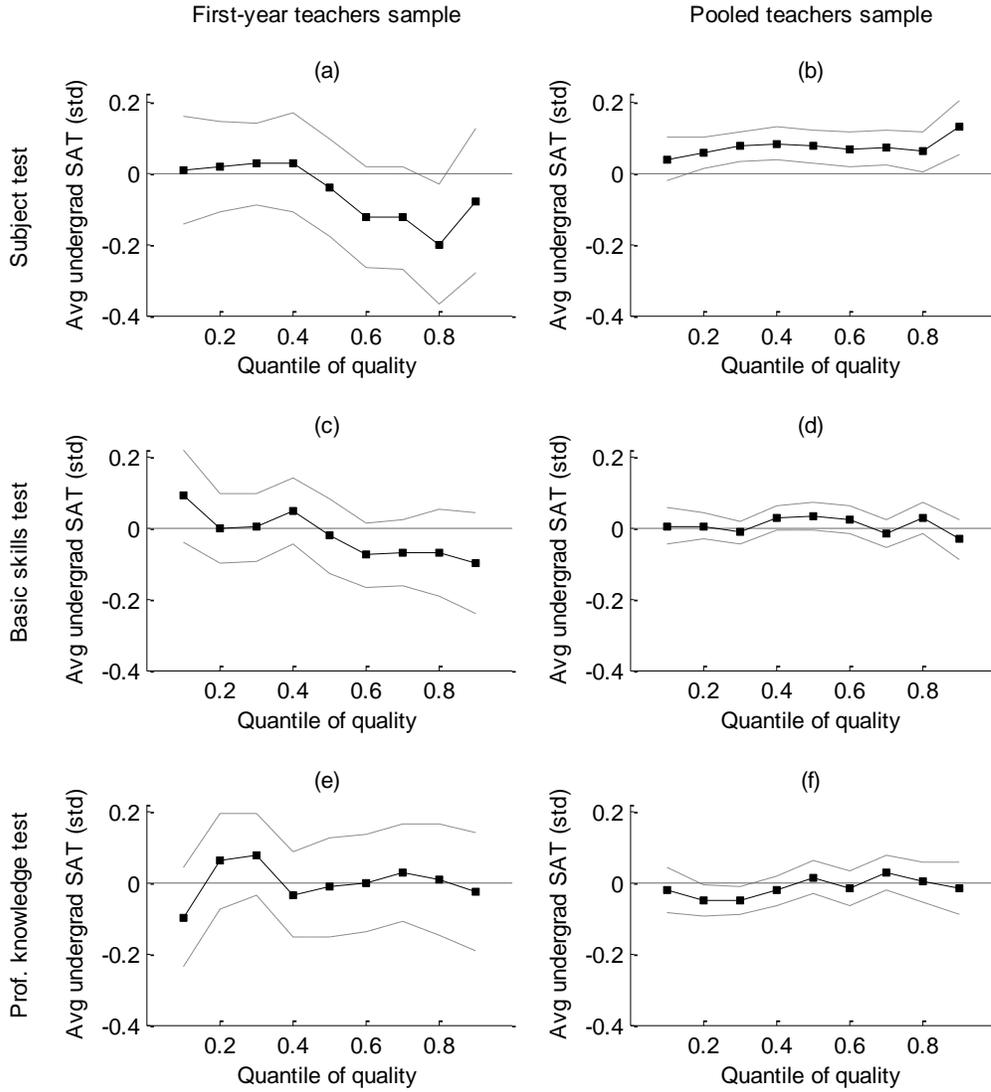
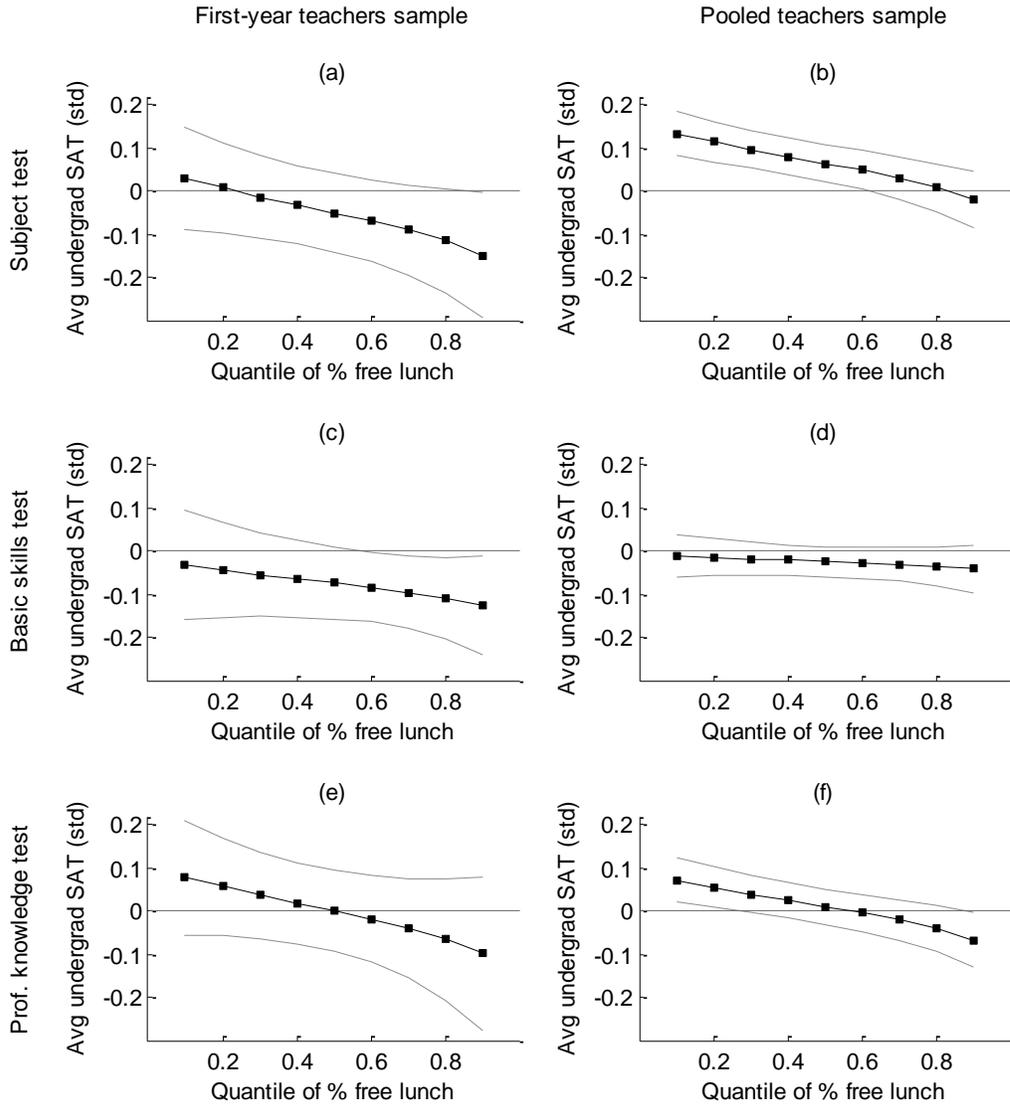First-year teachers sample          Pooled teachers sample



Notes: Effects of subject test law, basic skills test law, and professional knowledge test law on quantiles of teacher input quality distribution. Panels on the left display first-year teacher sample and on the right display pooled teacher sample. Robust, pointwise 90% confidence bands are displayed by dashed lines.

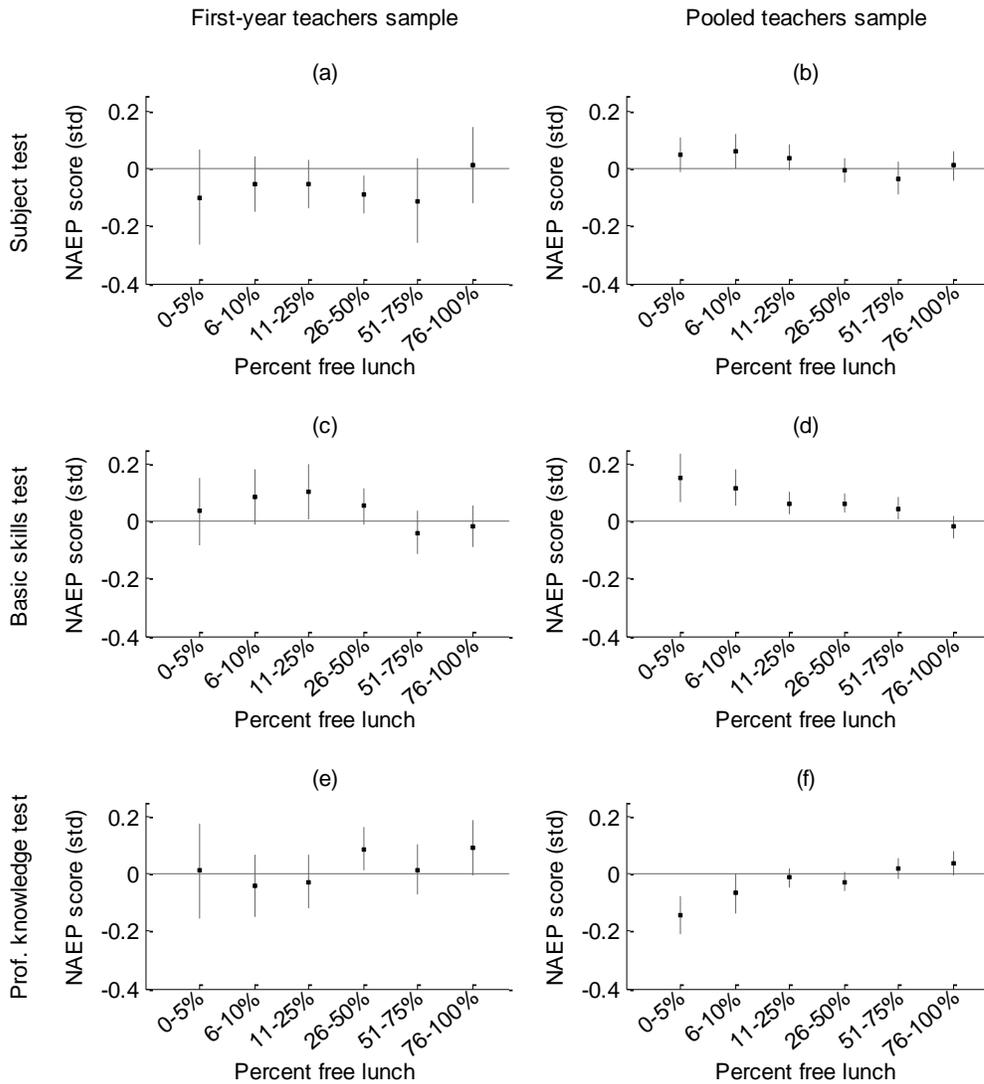Figure 5: Effects of certification test laws on output quality distribution
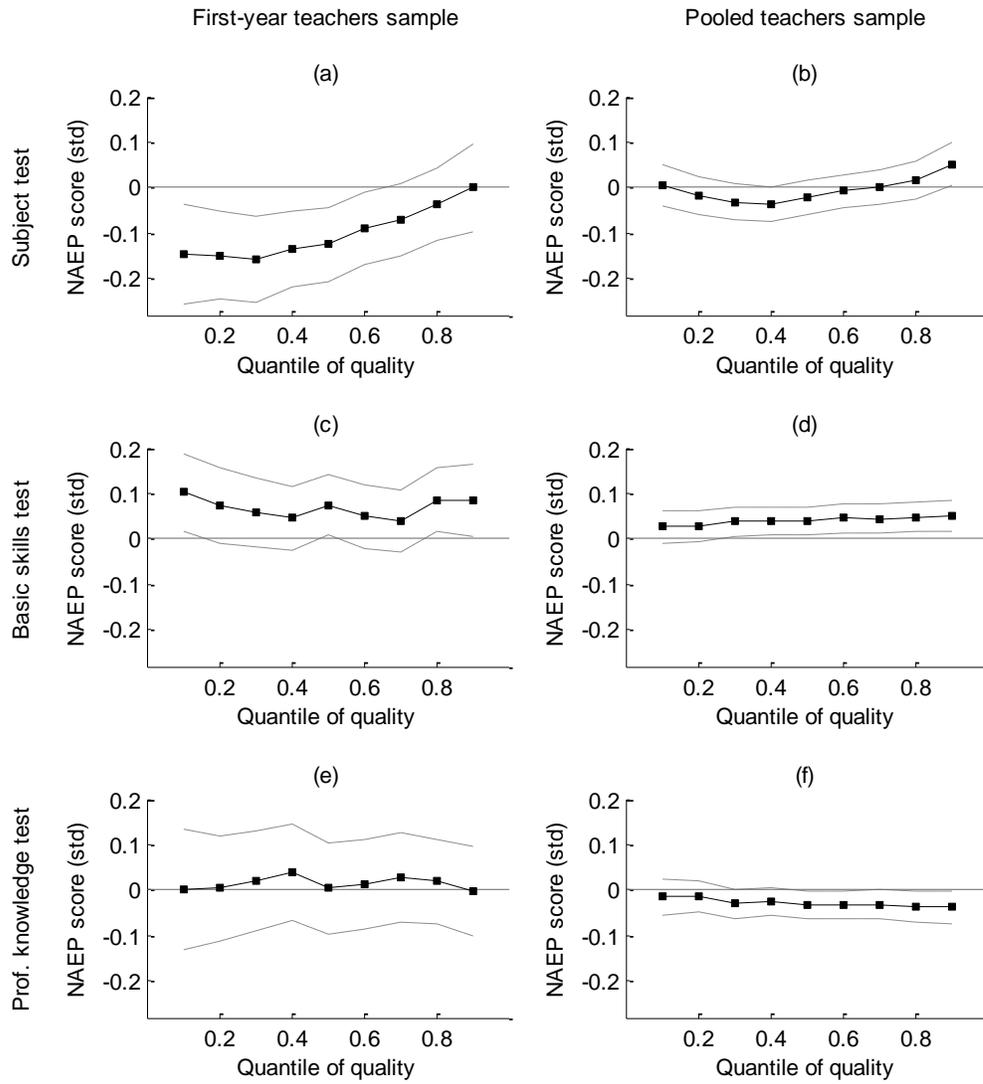


Notes: Effects of subject test law, basic skills test law, and professional knowledge test law on quantiles of output quality distribution. Panels on the left display first-year teacher sample and on the right display pooled teacher sample. Robust, pointwise 90% confidence bands are displayed by dashed lines.

Figure 6: Heterogeneous effects of certification test laws on input quality by income



Notes: Effects of subject test law, basic skills test law, and professional knowledge test law on teacher input quality in areas differing by income, where income is measured by the percent of students eligible for free lunch. Panels on the left display first-year teacher sample and on the right display pooled teacher sample. Robust, pointwise 90% confidence bands are displayed by dashed lines.

Figure 7: Heterogeneous effects of certification test laws on output quality by income
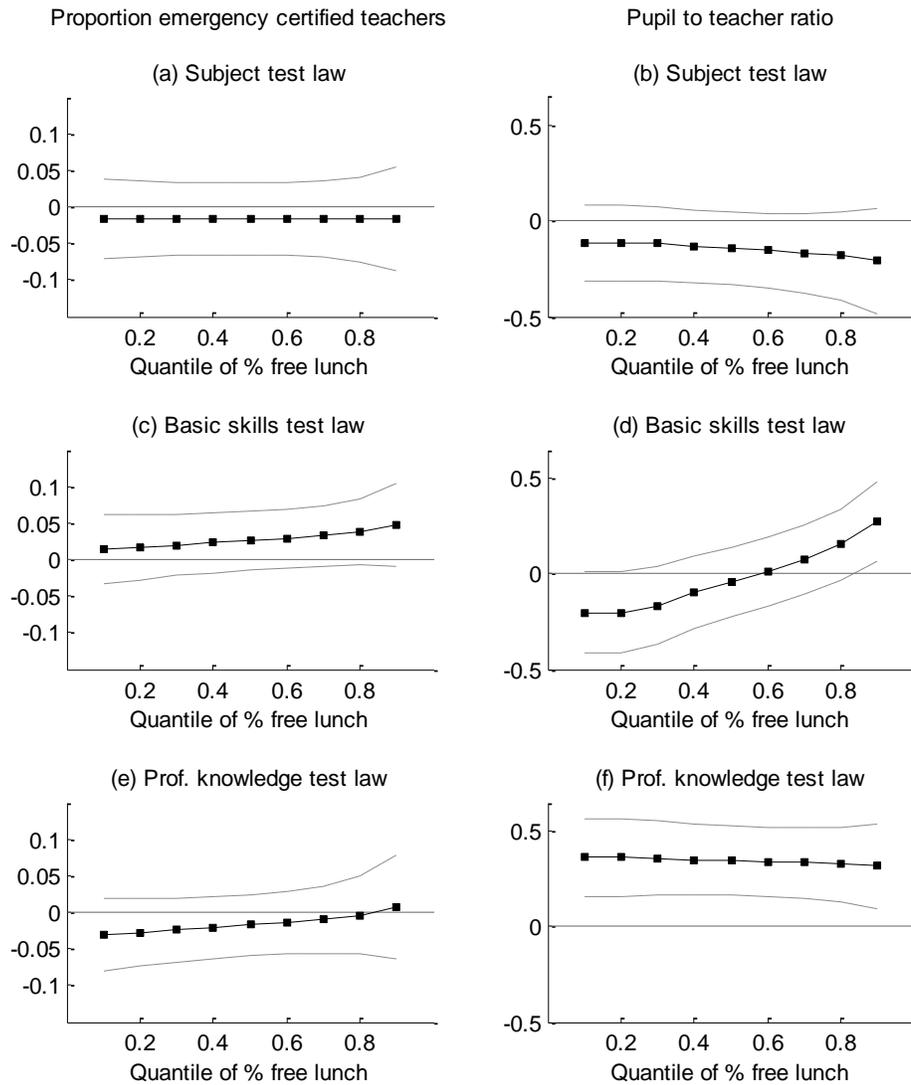


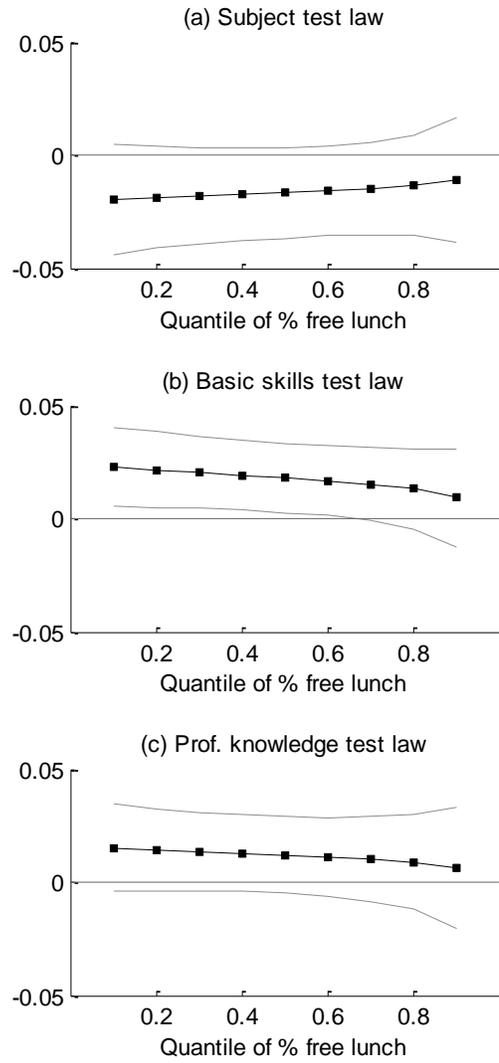Notes: Effects of subject test law, basic skills test law, and professional knowledge test law on teacher output quality in areas differing by income, where income is measured by the percent of students eligible for free lunch. Panels on the left display first-year teacher sample and on the right display pooled teacher sample. Robust, pointwise 90% confidence bands are displayed by vertical lines.

Figure 8: Heterogeneous effects of certification test laws on proportion of emergency-certified teachers and pupil-to-teacher ratio

Proportion emergency certified teachers

Pupil to teacher ratio

(a) Subject test law

(b) Subject test law

(c) Basic skills test law

(d) Basic skills test law

(e) Prof. knowledge test law

(f) Prof. knowledge test law

Notes: Effects of subject test law, basic skills test law, and professional knowledge test law on proportion of emergency certified teachers (observed in the SASS dataset) and pupil to teacher ratio (observed in the CCD) in areas differing by income, where income is measured by the percent of students eligible for free lunch. Robust, pointwise 90% confidence bands are displayed by dashed lines.

Figure 9: Heterogeneous effects of certification test laws on log of teacher salary



(a) Subject test law

(b) Basic skills test law

(c) Prof. knowledge test law

Notes: Effects of subject test law, basic skills test law, and professional knowledge test law on teacher salary in areas differing by income, where income is measured by the percent of students eligible for free lunch. Results correspond to first-year teacher sample only. Robust, pointwise 90% confidence bands are displayed by dashed lines.

Table 1: Number of states changing certification test laws over the period 1983-2008.

| Number of states | Subject | Basic skills | Prof knowledge |
|---|---|---|---|
| With test law over whole sample period | 8 | 8 | 9 |
| With no test law over whole sample period | 4 | 5 | 18 |
| Changing from having no test law to having a test law | 39 | 34 | 23 |
| Changing from having a test law to not having a test law | 2 | 13 | 9 |

Notes: Table displays, in the first column, the number of state which required a subject test for initial licensure in every year of the sample period, the number of states which never required a subject test during the sample period, the number of states which changed from not requiring to requiring a subject test, and the number of states which changed from requiring to not requiring a subject test. The second and third columns show similar measures for basic skills and professional knowledge tests.

Table 2: Summary statistics of SASS and NAEP variables of interest

| | Mean | Standard deviation | Min | Max |
|---|---|---|---|---|
| SASS Variables (60,820 observations) | | | | |
| Teacher undergraduate institution average SAT | 915.98 | 106.65 | 570 | 1,410 |
| Percent eligible for free lunch | 39.11 | 24.17 | 0 | 100 |
| Teacher years of experience | 5.29 | 3.07 | 1 | 11 |
| Cohort size | 32.76 | 17.23 | 10 | 113 |
| NAEP Variables (366,100 observations) | | | | |
| Eighth grade math scores | 261.85 | 39.03 | 79.86 | 388.65 |
| Teacher years of experience | 4.54 | 2.58 | 1 | 11 |
| Cohort size | 203.00 | 106.76 | 10 | 1,439 |

Notes: Summary statistics for SASS and NAEP datasets. In SASS dataset, cohort size represents the number of teachers in a state/year/experience level cell. In NAEP dataset, cohort size represents the number of students in a cell taught by teachers in the same state/year/experience level.

Table 3: Average effect of certification test laws on input quality

| | Dependent Variable: Average input quality within group | | | | | | | |
| | First-year teachers only | | | All experience levels pooled | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Subject test law | -0.02 | -0.04 | -0.01 | 0.08*** | 0.08*** | 0.07*** | 0.08*** | 0.07*** |
| | (0.06) | (0.06) | (0.09) | (0.02) | (0.02) | (0.03) | (0.02) | (0.03) |
| Basic skills test law | -0.02 | -0.02 | -0.07 | 0.01 | 0.01 | 0.00 | 0.01 | 0.00 |
| | (0.04) | (0.05) | (0.07) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| Prof. knowledge test law | 0.00 | -0.01 | -0.02 | 0.02 | 0.01 | 0.01 | 0.01 | 0.01 |
| | (0.05) | (0.06) | (0.08) | -(0.02) | -(0.02) | -(0.02) | -(0.02) | -(0.02) |
| Demographic controls | -- | Yes | Yes | -- | Yes | Yes | Yes | Yes |
| State-specific time trend | -- | -- | Yes | -- | -- | Yes | -- | Yes |
| State-specific exp. effect | -- | -- | -- | -- | -- | -- | Yes | Yes |
| Sample Size | 245 (state-year cells) | | | 2,318 (state-year-cohort cells) | | | | |

Notes: Table reports results from group-level quantile regressions of input quality on test law dummies. An observation is a group, which is a state-year cell in columns (1)–(3) and a state-year-cohort cell in columns (4)–(8). Dependent variable is the average input quality within a group, weighting by teacher sample weights when aggregating to the group level. Input quality at the individual teacher level is the average SAT for the teacher's undergraduate institutional, standardized to have mean 0 and variance 1 in the sample. All regressions include state and year fixed effects. All pooled regressions include experience cohort fixed effects. Robust standard errors are reported in parentheses. $* = p < 0.10, ** = p < 0.05, *** = p < 0.01$.

Table 4: Average effect of certification test laws on output quality

| | Dependent Variable: Average output quality within group | | | | | | | |
| | First-year teachers only | | | All experience levels pooled | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Subject test law | -0.05 | -0.09* | -0.06 | 0.02 | 0.00 | 0.00 | 0.01 | 0.01 |
| | (0.05) | (0.05) | (0.06) | (0.03) | (0.02) | (0.02) | (0.02) | (0.03) |
| Basic skills test law | 0.05 | 0.07* | 0.03 | 0.06*** | 0.04** | 0.03* | 0.05** | 0.03 |
| | (0.04) | (0.04) | (0.06) | (0.02) | (0.02) | (0.02) | (0.02) | (0.03) |
| Prof. knowledge test law | 0.02 | 0.00 | 0.00 | -0.02 | -0.03 | 0.00 | -0.04** | 0.02 |
| | (0.07) | (0.06) | (0.05) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| Demographic controls | -- | Yes | Yes | -- | Yes | Yes | Yes | Yes |
| State-specific time trend | -- | -- | Yes | -- | -- | Yes | -- | Yes |
| State-specific exp. effect | -- | -- | -- | -- | -- | -- | Yes | Yes |
| Sample Size | 259 (state-year cells) | | | 2277 (state-year-cohort cells) | | | | |

Notes: Table reports results from group-level quantile regressions of output quality on test law dummies. An observation is a group, which is a state-year cell in columns (1)–(3) and a state-year-cohort cell in columns (4)–(8). Dependent variable is the average output quality within a group, weighting by student sample weights when aggregating to the group level. Output quality at the individual student level is the eighth grade NAEP math score, standardized to have mean 0 and variance 1 in the sample. All regressions include state and year fixed effects. All pooled regressions include experience cohort fixed effects. Robust standard errors are reported in parentheses. $* = p < 0.10, ** = p < 0.05, *** = p < 0.01$.

Table 5: Average effect of certification test laws on input quality controlling for heterogeneous income effects

| | Dependent Variable: Input quality | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | First-year teachers only | | | All experience levels pooled | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Subject test law | -0.06 | -0.06 | 0.01 | 0.06** | 0.06** | 0.05* | 0.06** | 0.05 |
| | (0.06) | (0.06) | (0.06) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) |
| Basic skills test law | -0.07 | -0.08 | -0.17** | -0.03 | -0.02 | -0.03 | -0.03 | -0.04 |
| | (0.05) | (0.05) | (0.07) | (0.02) | (0.02) | (0.02) | (0.02) | (0.03) |
| Prof. knowledge test law | 0.00 | -0.01 | 0.01 | 0.01 | 0.01 | 0.03 | 0.01 | 0.04 |
| | (0.06) | (0.06) | (0.07) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) |
| Demographic controls | -- | Yes | Yes | -- | Yes | Yes | Yes | Yes |
| State-specific time trend | -- | -- | Yes | -- | -- | Yes | -- | Yes |
| State-specific exp. effect | -- | -- | -- | -- | -- | -- | Yes | Yes |
| Sample Size | | 7,900 | | | | 62,810 | | |

Notes: Table reports results from regressing input quality on certification test dummies and interaction of certification test dummies with percent of students qualifying for free lunch, evaluated at the mean of the percent free lunch variable. An observation is a single teacher. Columns (1)–(3) use only first-year teachers; columns (4)–(8) use all experience levels. Input quality is the average SAT for the teacher's undergraduate institutional, standardized to have mean 0 and variance 1 in the sample. All regressions include state and year fixed effects and a main effect for the perfect free lunch, and are weighted by teacher sampling weights. Standard errors, clustered at the state-year level, are reported in parentheses. $* = p < 0.10, ** = p < 0.05, *** = p < 0.01$.

# A   Additional tables and figures

Table A.1: Summary statistics for SASS, NAEP, and CCD variables

| | Mean | Standard deviation | Min | Max |
|---|---|---|---|---|
| SASS Variables (60,820 observations) | | | | |
| District enrollment | 67,771 | 204,023 | 0 | 1,197,117 |
| Percent minority enrollment | 35.59 | 29.37 | 0 | 100 |
| Suburb dummy | 0.47 | -- | -- | -- |
| Rurual dummy | 0.25 | -- | -- | -- |
| Emergency certified dummy | 0.13 | 0.34 | 0 | 1 |
| Salary (2008 dollars) | 27,674 | 7,395 | 11,800 | 51,400 |
| Log salary (2008 dollars) | 10.45 | 0.14 | 9.59 | 10.98 |
| CCD Variables (328,810 observations) | | | | |
| Pupil to teacher ratio | 14.94 | 3.90 | 4.00 | 26.67 |
| NAEP Variables (366,100 observations) | | | | |
| School percent black | 16.03 | 23.57 | 0 | 100 |
| School percent Hispanic | 14.68 | 23.13 | 0 | 100 |
| NAEP Dummy Variables (means) | | | | |
| Encyclopedia in home dummy | 0.61 | Enrollement: 1-99 | | 0.01 |
| Suburb dummy | 0.46 | Enrollment: 100-299 | | 0.07 |
| Rural dummy | 0.25 | Enrollment: 300-499 | | 0.18 |
| Black dummy | 0.17 | Enrollment: 500-749 | | 0.29 |
| Hispanic dummy | 0.15 | Enrollment: 750-999 | | 0.23 |
| Asian dummy | 0.04 | Enrollment: 1000-1499 | | 0.17 |
| Male dummy | 0.50 | Enrollment: 1500+ students | | 0.05 |

Notes: Summary statistics for variables not displayed in Table 2. Data on salaries and emergency certification comes from the SASS sample of first-year teachers, and consists of 8,770 observations. Data on pupil to teacher ratios comes from the CCD from 1988-2010.

Figure A.1: Trend robustness of input quality distributional effects



(a) State time trends

(b) State experience trends
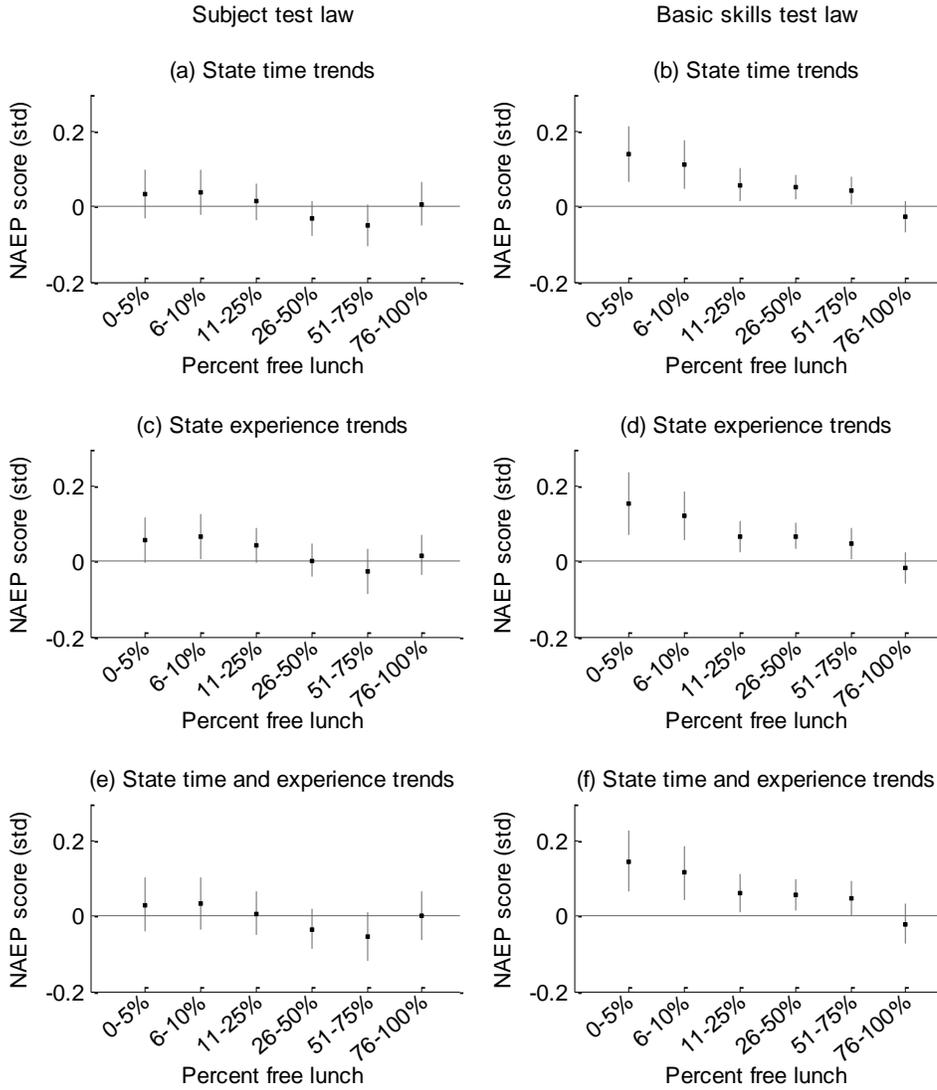
(c) State time and experience trends
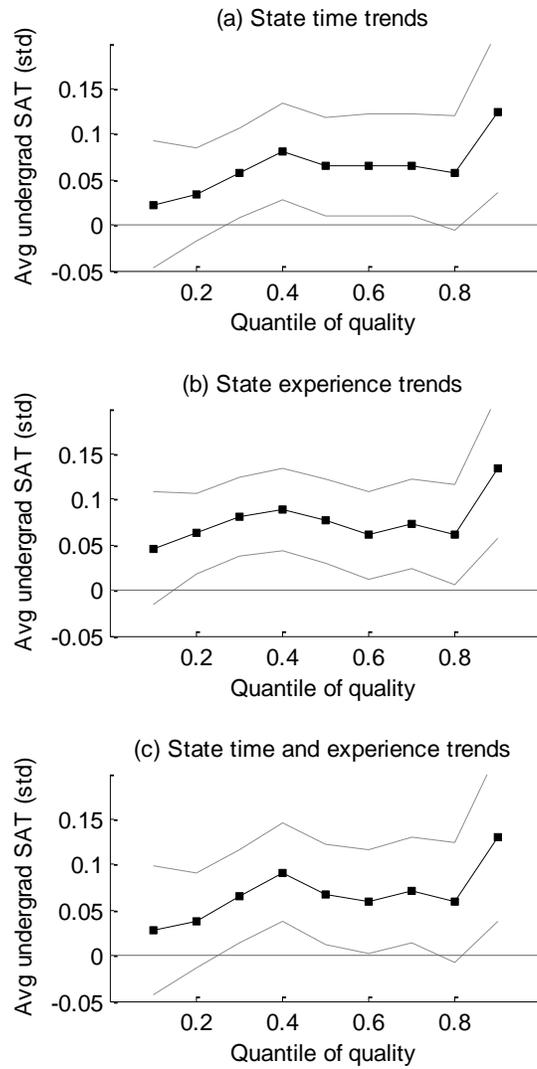
Notes: Effects of subject test law on distribution of input quality with pooled teacher sample, with state-specific linear time trends and/or state-specific returns to teaching experience included.

Figure A.2: Trend robustness of output quality distributional effects
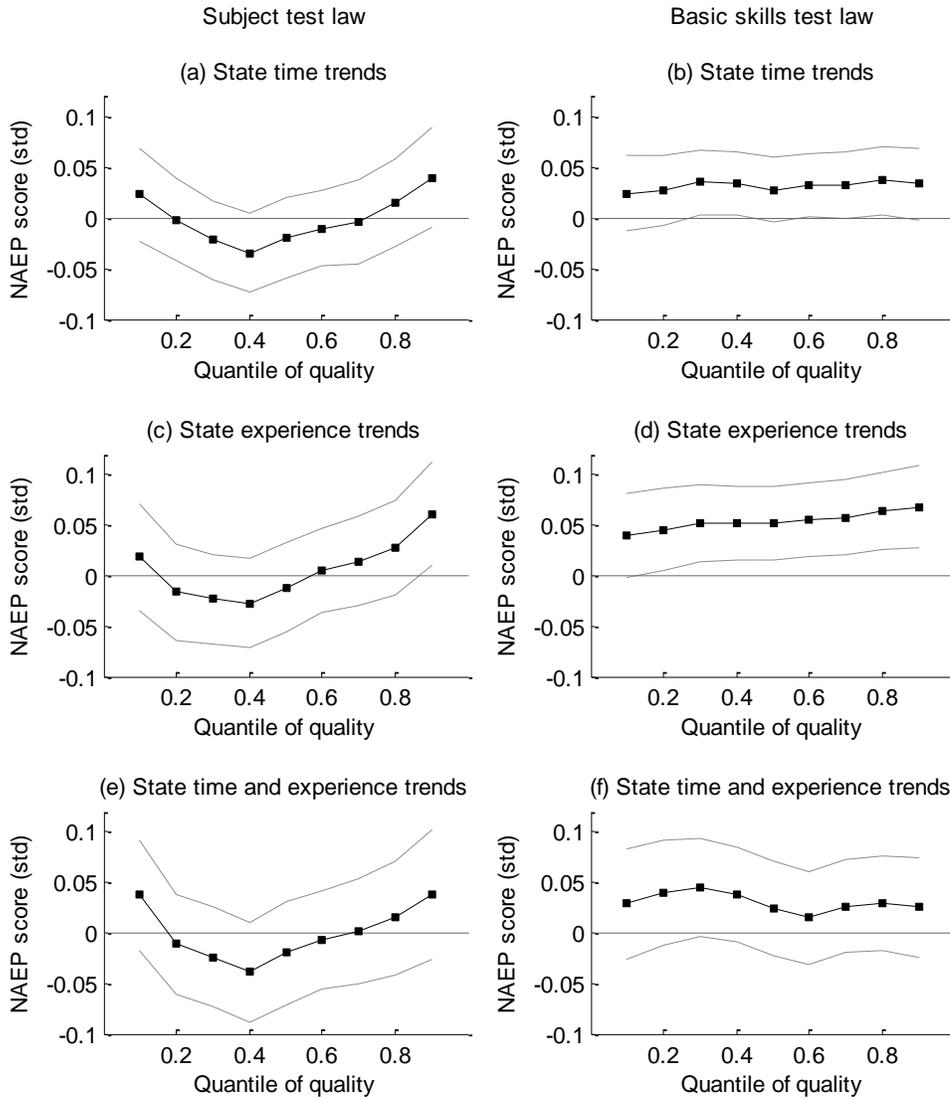
Notes: Effects of subject test law on distribution of output quality with pooled teacher sample, with state-specific linear time trends and/or state-specific returns to teaching experience included.

Figure A.3: Trend robustness of output quality effects by income level



Notes: Effects of basic skills test law on output quality by bins of percent free lunch distribution using pooled teacher sample, with state-specific linear time trends and/or state-specific returns to teaching experience included.

Figure A.4: Event study: effect of subject test laws on input quality distribution
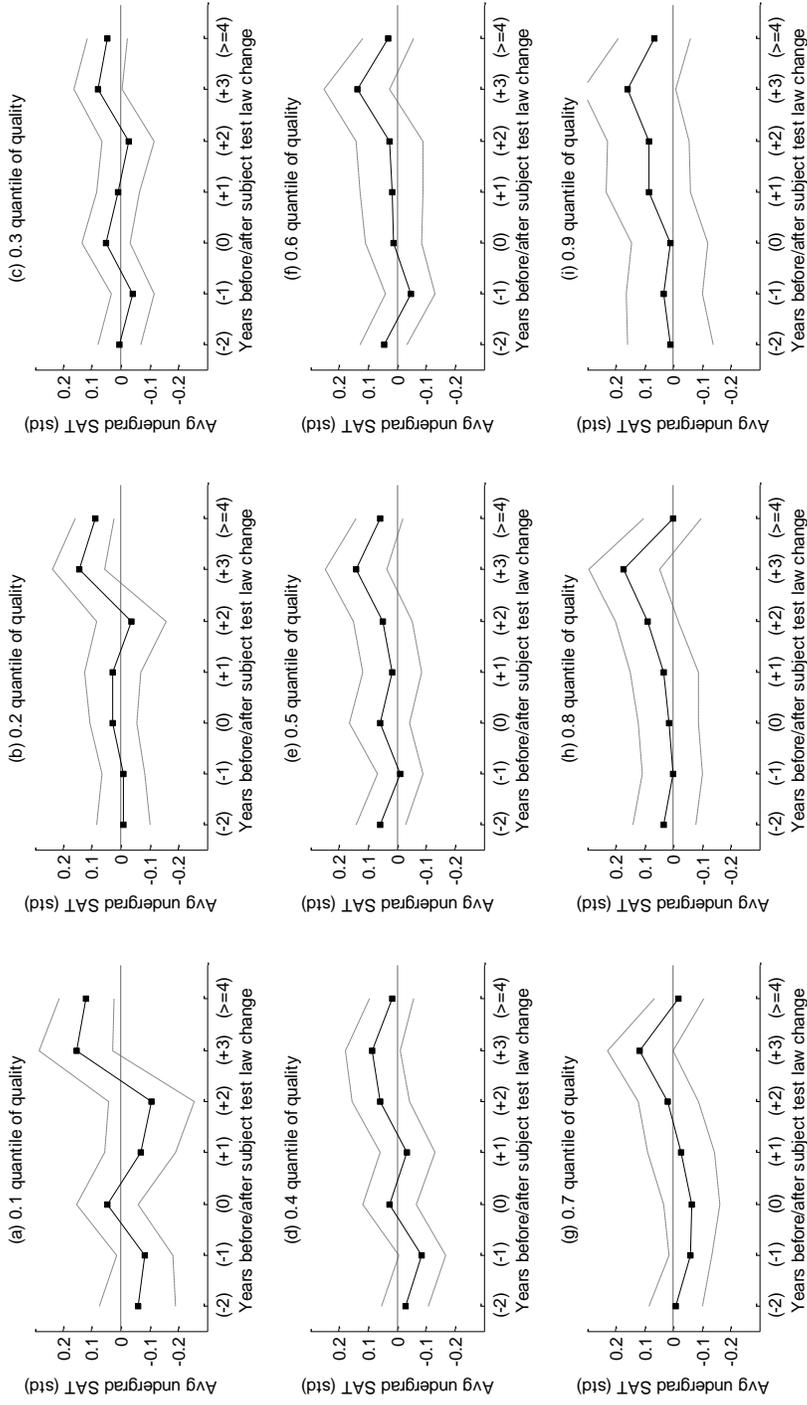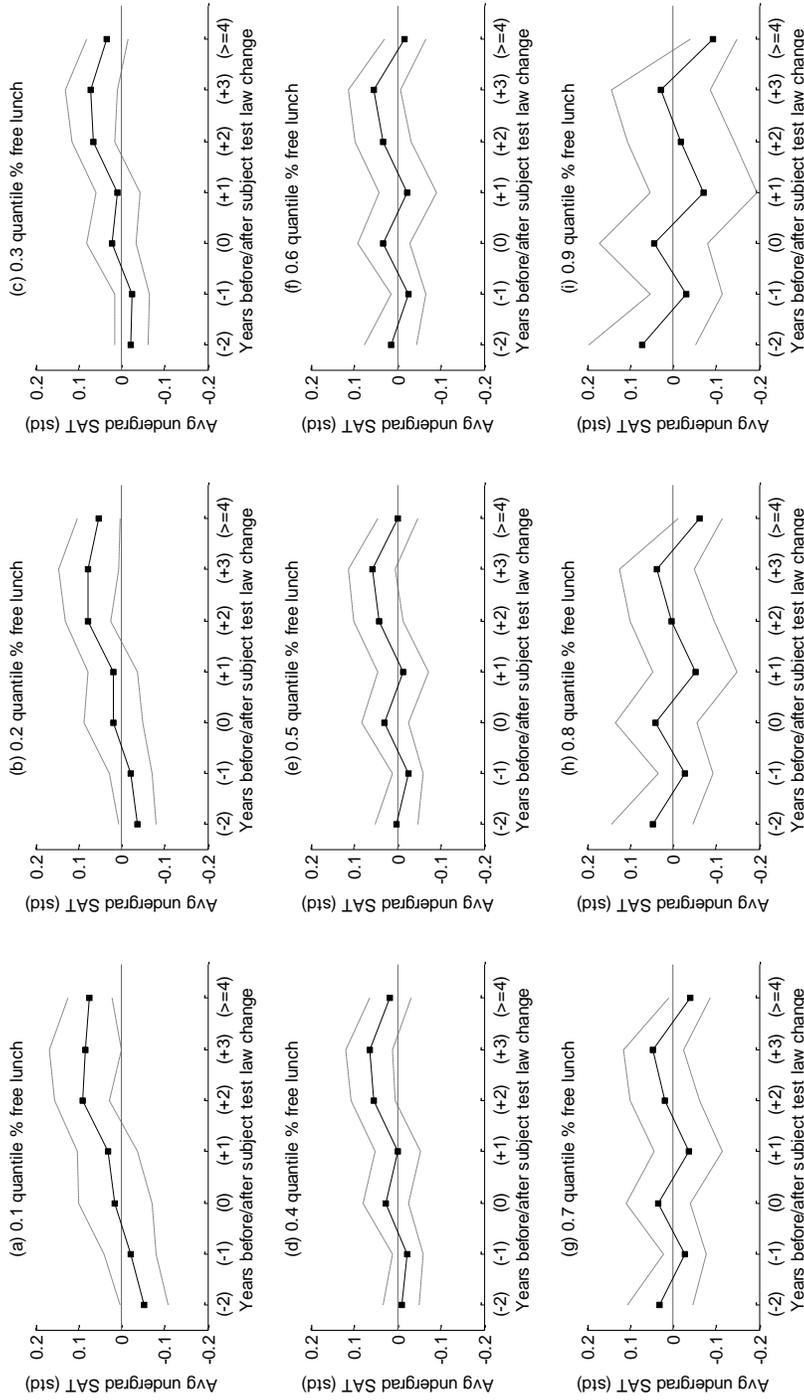


Notes: Event study analysis of subject test law on quantiles of teacher input quality distribution using only states which changed from having no subject test law to having a subject test law. Each panel displays the effect for a particular decile of the distribution. Within a given panel, the plot corresponds to the effect two years before the subject test law changed, one year before, the year of the change, one year after, etc. The final point corresponds to the effect four or more years after. Robust, pointwise 90% confidence bands are displayed by dashed lines.

Figure A.5: Event study: effect of subject test laws on output quality distribution
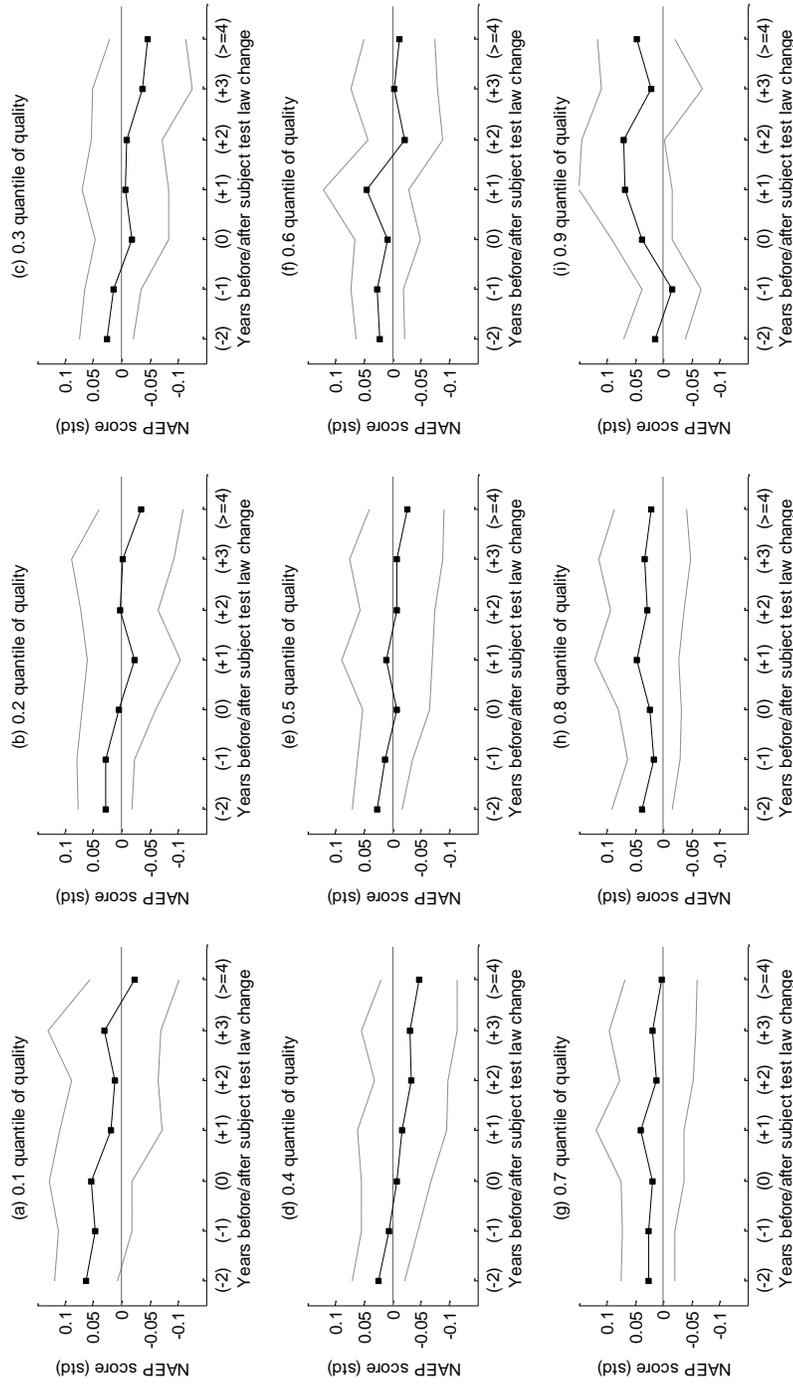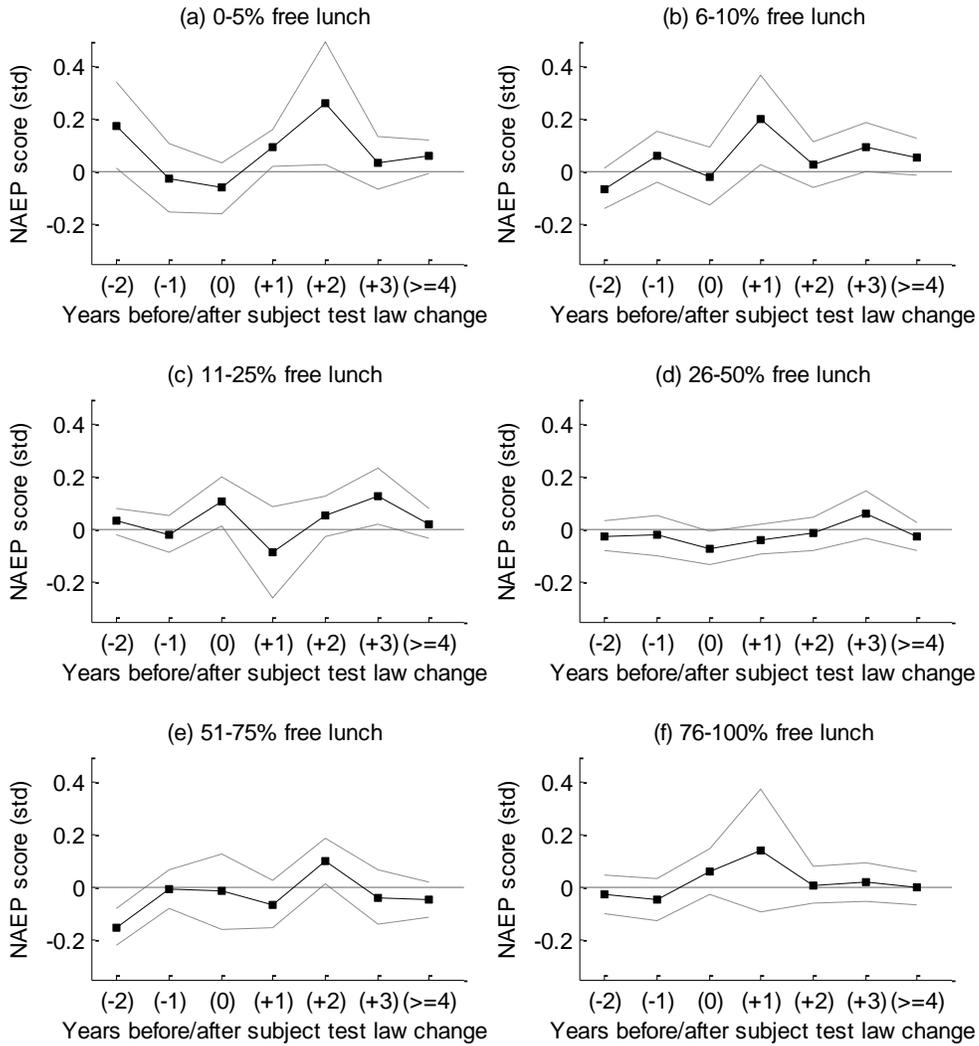


Notes: Event study analysis of subject test law on quantiles of teacher output quality distribution using only states which changed from having no subject test law to having a subject test law. Each panel displays the effect for a particular decile of the distribution. Within a given panel, the plot corresponds to the effect two years before the subject test law changed, one year before, the year of the change, one year after, etc. The final point corresponds to the effect four or more years after. Robust, pointwise 90% confidence bands are displayed by dashed lines.

Figure A.6: Event study: heterogeneous effects of subject test laws on input quality by income



Notes: Event study analysis of subject test law on teacher output quality in areas differing by income and using only states which changed from having no subject test law to having a subject test law. Each panel displays the effect for a particular decile of the percent free lunch variable. Within a given panel, the plot corresponds to the effect two years before the subject test law changed, one year before, the year of the change, one year after, etc. The final point corresponds to the effect four or more years after. Robust, pointwise 90% confidence bands are displayed by dashed lines.

Figure A.7: Event study: heterogeneous effects of subject test laws on output quality by income



Notes: Event study analysis of subject test law on teacher output quality in areas differing by income and using only states which changed from having no subject test law to having a subject test law. Each panel displays the effect for a particular bin of the percent free lunch variable. Within a given panel, the plot corresponds to the effect two years before the subject test law changed, one year before, the year of the change, one year after, etc. The final point corresponds to the effect four or more years after. Robust, pointwise 90% confidence bands are displayed by dashed lines.