NATIONAL
CENTER *for* ANALYSIS *of* LONGITUDINAL DATA *in* EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

A program of research by the American Institutes for Research with Duke University, Northwestern University, Stanford University, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

The Gateway to the Profession: Assessing Teacher Preparation Programs Based on Student Achievement

DAN GOLDHABER AND
STEPHANIE LIDDLE

WORKING PAPER 65 • JANUARY 2012

The Gateway to the Profession: Assessing Teacher Preparation Programs Based on Student Achievement

Dan Goldhaber
Center for Education Data & Research
University of Washington Bothell

Stephanie Liddle
Center for Education Data & Research
University of Washington Bothell

Contents

Acknowledgements

The research presented here utilizes confidential data from Washington State supplied by the Office of Superintendent for Public Instruction (OSPI). We gratefully acknowledge the receipt of these data. We wish to thank The Carnegie Corporation for support for this research. This paper has benefited from helpful comments from Joe Koski, John Krieg, and Jim Wyckoff. Finally, we wish to thank Roddy Theobald for research assistance and Jordan Chamberlain for editorial assistance. The views expressed in this paper do not necessarily reflect those of the University of Washington Bothell, Washington State, or the study's sponsor. Responsibility for any and all errors rests solely with the authors.

CALDER • American Institutes for Research
1000 Thomas Jefferson Street N.W., Washington, D.C. 20007
202-403-5796 • www.caldercenter.org

The Gateway to the Profession: Assessing Teacher Preparation Programs Based on Student Achievement

Dan Goldhaber and Stephanie Liddle

CALDER Working Paper No. 65
January 2012

Abstract

With teacher quality repeatedly cited as the most important *schooling* factor influencing student achievement, there has been increased interest in examining the efficacy of teacher training programs. This paper presents research examining the variation between and impact that individual teacher training institutions in Washington state have on the effectiveness of teachers they train. Using administrative data linking teachers' initial endorsements to student achievement on state reading and math tests, we find the majority of teacher training programs produce teachers who are no more or less effective than teachers who trained out-of-state. However, we do find a number of cases where there are statistically significant differences between estimates of training program effects for teachers who were credentialed at various in-state programs. These findings are robust to a variety of different model specifications.

Introduction

Focusing on Teacher Training

Policymakers at all levels of government are looking for ways to improve the teacher workforce. Unfortunately, evidence of effective policy that would improve teacher quality remains elusive, as it is clear teacher quality is only weakly related to readily quantifiable teacher attributes like licensure status, degree, and experience levels (Goldhaber, 2002; Hanushek, 1986; 1997).

State policymakers arguably have a great deal of leverage over teacher quality: by setting standards for teacher preparation programs, teacher licensure, and recertification, states determine who is eligible to join and remain in the teaching profession.[1] While there has been a marked increase over the last decade in the number of teachers entering the profession through alternative routes, most teachers train at traditional state-approved colleges and universities.[2] Thus, changing the way teachers in traditional training programs are selected or prepared could significantly influence the teacher workforce.

Ostensibly, states regulate specific teacher preparation programs through evaluation and accreditation, but these quality control mechanisms are deemed by many to be ineffective. For instance, as Arthur Levine, the president of Teachers College, Columbia University, notes, "Under the existing system of quality control, too many weak programs have achieved state approval and been granted accreditation" (2006, p. 61). The perceived lack of quality control within the teacher preparation system paints a discouraging picture of the system's prospects for improving the teacher workforce.[3] Not surprisingly then, there have been calls for reform that include closer monitoring of programs and

[1] Local school systems may have the final word on who teaches in the classroom, but the state ultimately sets the parameters of their choice set.
[2] According to Feistritzer & Haar (2010), the number of teachers certified through alternative routes has grown from 275 in 1985-86 to 59,000 in 2008-09. The National Center for Education Statistics estimates that 327,000 teachers were hired in the US in 2007, which suggests that less than 20 percent of new teachers are entering through alternative routes.
[3] See, for instance, Cochran-Smith and & Zeichner (2005), Crowe (2010), Duncan (2010), NCATE (2010).

holding them accountable for student achievement results (Teaching Commission, 2006).[4] Evaluating teacher training programs based, at least in part, on the performance of their trainees has emerged as an education reform strategy in several states and was a central tenet of the Race to the Top grant competition.[5]

But teacher training often gets painted with a broad brush, despite the fact that there are over 2000 traditional teacher training programs in the United States.[6] And rhetoric about teacher training aside, there exists relatively little quantitative information linking programs with the quality of their graduates, or how specific approaches to teacher preparation are related to the effectiveness of teachers in the field (National Research Council, 2010). Much of the existing academic literature on teacher preparation has focused on differences in the effectiveness of teachers who enter the profession through alternative versus traditional pathways (Glazerman et al., 2006; Goldhaber and Brewer, 2000; Xu et al., 2007). Researchers have only recently used administrative databases to draw the link from teacher preparation programs program to in-service teachers and then to student achievement in order to draw conclusions about the efficacy of different teacher training programs (Harris & Sass, 2007; Boyd et al., 2009; Noell et al., 2008 Henry et al., 2011).[7]

Like every state, Washington has longstanding requirements for initial entry into the teacher workforce. Washington's standards are relatively stringent in the sense that, to date, it has not included alternative routes into the profession as a source of new teachers (National Council on Teacher Quality,

---

[4] Some states have considered ways to hold teacher preparation programs accountable for their graduates (Johnston, 1997), but they are in the minority and there is little evidence, even in these states, that weak education programs are systematically closed. It is not surprising, then, that one of Levine's five major recommendations for improving pre-service training is to "close failing teacher education programs" (2006, p. 111).

[5] In both Colorado and Tennessee, for instance, the renewal or non-renewal of teacher and principal preparation programs depend at least in part on teacher evaluations that are themselves dependent on student achievement (Colorado Senate Bill 10-191; Tennessee Senate Bill 7005).

[6] In line with Constantine, et al. (2009), we define "traditional training programs" as those from which new teachers must complete all their certification requirements before beginning to teach.

[7] In practice it is difficult, if not impossible to definitively assess the causal impact of training institutions on teacher candidates since the effectiveness of in-service teachers is likely to depend on both their individual attributes as well as what they learned while being trained.

2007).[8] The great majority of its teachers are trained at one of the 21 state-approved teacher credentialing programs (most of which are private, four-year universities).

In this paper we present research on the extent to which teacher training programs explain the variation in teacher effectiveness, as well as how individual training programs in the state of Washington compare to one another, as judged by the effectiveness of the teachers they produce. This assessment is based on analyses of administrative data that links each teacher's initial teaching endorsement ("qualification to teach") to student achievement on state tests in reading and math. We find that the majority of programs produce teachers who are no more or less effective than teachers who are credentialed outside of the state. However, we see meaningful variation in the effectiveness of teachers who come from the various in-state programs. As an example, in math, the average difference between teachers who received a credential from a program with the lowest performing teachers and those who received a credential from the program with the highest performing teachers is about 7 percent of a standard deviation, which is roughly as large as the average difference in performance between students eligible for free or reduced-price lunches and those who are not. In reading, this same difference is about 10 percent of a standard deviation, which is roughly equivalent to the average performance difference between students with learning disabilities and those without.

Teacher Training and Student Achievement

The policy direction and value of teacher training are hotly debated topics. Much of this debate is fueled by comparisons of teachers who hold either a traditional or alternative license.[9] Findings that suggest there is often little difference between teachers who enter the profession through different routes (and usually have a different licensure status) have led some to conclude that there is little value in traditional teacher training (Gatlin, 2009; Stotko, Ingram & Beaty-O'Ferrall, 2007; Wilson, et al., 2002;

[8] In 2009-10, for example, 248 out of 57,881 employed teachers were credentialed through alternative routes.
[9] See, for instance, Darling-Hammond (1999), Goldhaber and Brewer (2000), Glazerman et al. (2006).

Rochester, 2002; Gansle et al., 2010). There is some evidence to support this proposition. For example, both experimental and non-experimental research on Teach for America (TFA) — probably the best-known alternative route into the classroom — suggests TFA teachers compare favorably, in terms of student achievement, with other teachers in schools which employ TFA members (Decker et al. 2004; Xu et al. 2007). TFA, however, represents a very selective group of teachers in terms of academic preparation; one might guess that a different comparison group would yield different results, but a more recent study (Constantine et al., 2009) on less selective alternative routes to the classroom also shows little difference in the test achievement of students whose teachers received traditional training and those who entered teaching through alternative routes.[10]

Findings like those described above are not surprising given that there is likely significant variation in effectiveness amongst teachers who received training in either traditional or alternative programs. The broad classification of teachers by route of entry into the profession may mask the fact that when it comes to teacher preparation, programs themselves may be more different than they are alike. Corcoran, for example, describes a chaotic pre-service training system in which "visions of good teaching, standards for admission, rigor and amount of subject matter preparation, clinical experiences…and quality of assessments" differ widely among and within traditional training programs (2007, p. 314). Levine (2006) presents a similar picture, highlighting the vast "disparities in institutional quality" across teacher preparation programs.

The notion that focusing on within-route differences in training programs might be the best way to determine what kinds of selection processes or training experiences predict teacher effectiveness is buttressed by two recent research studies. Specifically, both Boyd et al. (2009) and Kane et al. (2007) investigate teachers employed in New York City who entered the profession through different routes. Each finds there is far more quality variation *within* a certification category than there is *between*

[10] The study also found little evidence that either the content or extent of teacher training coursework completed by alternative route teachers was associated with the achievement of their students.

certification categories of teachers.[11] But there are relatively few studies that look at the connection between individual teacher training programs and student achievement. Harris and Sass (2007), for example, investigate the relationship between teacher training characteristics and teacher productivity in Florida. Their data link student test scores, teachers' professional development programs, pre-service teacher training programs, college coursework, and pre-college entrance exam scores (which allows them to address selection effects in training programs).[12] In addition, Noell et al.'s (2008) work in Louisiana also capitalizes on the ability to link student demographic, attendance, and test score data to teacher demographic, attendance, and certification data through a curriculum database (as well as classroom and school characteristics). These data allow them to run value-added analyses to assess the effectiveness of new versus more experienced teachers from various teacher preparation programs based on a five-point performance band rating system. Based on these comparisons, they find considerable variability in effectiveness across teacher preparation programs. These findings appear to be fairly consistent over time (Noell et al., 2009; Gansle et al., 2010).

More recent work by Henry et al. (2011), using data from North Carolina, look at differences between teachers trained in-state versus out-of-state. They find teachers trained in-state tended to be slightly more effective than those trained in out-of-state institutions, but they also find considerable overlap in the estimated effectiveness of North Carolina institutions: students taught by teachers trained at in-state institutions outperformed students taught by out-of-state programs in 14 comparisons, under-performed in 9, and were not significantly different for 74 comparisons.

Boyd et al.'s (2009) examination of the distribution of teacher performance from different training programs in New York City more clearly suggests that there is significant variation in the effectiveness of

[11] Kane et al. (2007) find, for example, that the gap in teacher effectiveness (measured by value-added) within each certification category is about ten times larger than the average gap between certification categories.

[12] Harris and Sass find that teacher training generally has little effect on teacher productivity. Content-focused teacher professional development programs positively influences teacher effectiveness, but only in middle and high school math. They find no evidence of a relationship between teacher pre-service (undergraduate) training and teacher productivity no matter what type of undergraduate degree the teacher holds.

teachers graduating from different programs and, moreover, that some program characteristics (e.g., timing of student teaching) predict program effectiveness. The difference between teachers from the average institution and highest performing institution is about as large as the average difference between students who are eligible for free or reduced lunch and students who are not. This degree of variation is similar for both math and language arts. Furthermore, institutions that produce effective math teachers also tend to produce effective language arts teachers. And yet, to the degree that the quality of training institutions is contingent on where a teacher works (i.e., that some institutions may serve districts with certain student populations particularly well), a study of a singular district like New York begs important questions about teacher preparation in other types of schools and districts (e.g., in rural and suburban areas).

While each of the above studies focuses on teacher training, it is important to note it is not possible to entirely disentangle the extent to which differences are a result of the type of teacher candidates selected by programs or the training that individuals receive while in a program. In our analysis, we try to tease out the effects of teacher selection into certain training programs and the efficacy of those teacher training programs. But, given data limitations, this is to a large extent impossible. Therefore, in step with previous research, the program estimates discussed below reflect combined selection and training effects.

Analytic Approach

A key to our project is the estimation of models that identify the effectiveness of teachers who obtain their initial teaching credential from different training programs in Washington—as well as those who received their credential from any out-of-state program and were subsequently approved by the

Washington State Office of Superintendent of Public Instruction (OSPI) to teach in Washington.[13] To do this, we estimate standard education production function models at the elementary level, where we can link teachers to their students through either the proctor name listed on the state assessment or a unique within-school course id. There is a growing body of literature that uses value-added models (VAMs) to identify causal impacts of schooling inputs, and in some cases the individual contributions of teachers, toward student learning gains on standardized tests.[14] There is, however, no universally accepted estimation specification for this purpose (NRC, 2010), and empirically derived program estimates involve making a number of strong assumptions about the nature of student learning.[15]

In this study we estimate several variants of the following model:

$$A_{ijklst} = \beta_1 A_{is(t-1)} + \beta_2 X_{it} + \beta_3 C_{jt} + \beta_4 T_{jt} + \beta_5 P_j + \beta_6 S_{kt} + \beta_7 D_{lt} + \gamma_{it} + \varepsilon_{ijklst} \qquad (1)$$

In equation (1), i represents students, j represents teachers, k represents schools, l represents districts, s represents subject area (math or reading), and t represents the school year. The achievement of students in a subject at time t, A_{ijklst}, which is standardized with subject, grade, and year, is regressed against: prior student achievement, $A_{is(t-1)}$; a vector of student background characteristics, X_{it}; a vector of classroom characteristics, C_{jt}; a vector of teacher characteristics and credentials, T_{jt}; a vector of credentialing program indicators, P_j; a vector of school characteristics, S_{kt}; a vector of district characteristics, D_{lt}; and grade dummies, γ_{it}. The error term associated with a particular student in a particular year, ε_{ijklst}, is assumed to be N(0, σ^2_{jt}).[16]

[13] In Washington, the legislature passes the laws about educator credentialing, the Professional Educators Standards Board (PESB) then writes the administrative code to interpret the law, and finally OSPI executes the administrative code.

[14] See Aaronson et al., (2007), Boyd et al., (2009), Clotfelter et al. (2007), Goldhaber (2007), Rockoff (2004) as examples of studies that attempt to isolate the impact of schooling inputs from other factors (such as family background or class size) that influence student growth on standardized tests.

[15] For a discussion of this in relation to the derivation of individual teacher effects, see Todd and Wolpin (2003).

[16] Note that we focus on self-contained classrooms so that subject area does not vary by teacher, class, or school. However, since we observe students multiple times we cluster standard errors at the student level to account for unique (and potentially unobservable) student-level factors that could influence a student's performance over time. Also note that results below are from models that do not include year fixed effects. However, the results do not change when we run models that do include year fixed effects.

The key area of interest is in the estimated coefficients for the program credentials, β_5.[17] Interpretation of these program credential indicators is complicated by the fact that individuals are selected into preparation programs and credentialed teacher candidates select into particular school districts, schools, and classrooms. As we noted above, in practice it is not possible to definitively assess the causal impact of training programs on teacher candidates since the effectiveness of in-service teachers likely depends on both their individual attributes as well as what they learned while being trained.

We attempt to account for selection in a variety of ways. First, in some specifications of (1), we include additional controls various measures of institutional selectivity or for the tests that prospective teachers take *prior* to entering a training program. To the extent that these measures control for individual pre-training teaching ability, they help account for selection into program. To account for the potential that teacher candidates from different preparation programs are funneled into school districts or schools that are systematically different from each other in ways that are not accounted for by variables included in our models, we estimate several variants of model (1) that include either school district or school fixed effects. These different models identify teacher training program estimates in different ways. Specifically, the student-, teacher-, classroom-, school-, and district-level observables included in the base specification are an attempt to isolate the effect of the training program net of other factors that could influence student achievement. However, it is well known that teachers and students are not randomly matched together, holding constant observable characteristics (Boyd et al, 2002; Clotfelter et al., 2006; Rothstein, 2010). Thus, we also estimate district fixed effects specifications in which program credentials are identified based on within-district differences in teachers, and school fixed effects specifications in which the differences are identified based on within-school differences between teachers.

[17] The referent category for these institution dummies is teachers who were trained outside of Washington State and received their initial teaching certificate from the state's OSPI.

Data

Information on teachers and students for this paper are derived from secondary data from six administrative databases prepared by OSPI: the *Washington State S-275* personnel report, the *Washington State Credentials* database, the *Core Student Record System (*CSRS*),* the *Comprehensive Education Data and Research System* (CEDARS), the *Washington Assessment of Student Learning (WASL)* database, and the *Washington State Report Card*.

The *S-275* contains information from Washington State's personnel-reporting process; it includes a record of all certified employees in school districts and educational service districts (ESDs), their place(s) of employment, and annual compensation levels. It also includes gender, race/ethnicity, highest degree earned, and experience which are useful for this study's purposes. (See **Table A1** in the appendix for means of selected teacher characteristics by training program).

The *Washington State Credentials* database contains information on the licensure/certification status of all teachers in Washington, including when and where teachers obtained their initial teaching certificates.[18,19] This database also includes teachers' testing outcomes on the Washington Educator Skills Test – Basic, or WEST-B, a standardized test that all teachers must take prior to entering a teaching training program.[20]

---

[18] From this database, we identify the institution from which a teacher received his or her first teaching certificate, which may or may not be where a teacher did his or her undergraduate work. OSPI's coding schema for, first-issue teaching certificates (i.e., what we call "initial" certificates) has changed over time. Under 1961 guidelines, individuals were issued *provisional* certificates. In 1971, additional guidelines were created to issue *initial* certificates. In 2000, guidelines changed once again to the current categorization of *residency* certificates. Note however, that after a major guideline change there is still a period during which certificates may be issued under their former names. So, even in 2000, some individuals may have received certificates under previous guidelines. We code all initial certificates to account for these historical changes.

[19] The "recommending agency" variable in these data identifies the college/university that did all of the legal paperwork to get an individual issued a teaching certificate. Thus, while likely that the recommending institution was also the institution where teachers were trained, the variable itself does not necessarily mean that the person graduated from the recommending agency.

[20] Since August 2002, candidates of teacher preparation programs in Washington State have been required to meet the minimum passing scores on all three subtests (Reading, Mathematics, and Writing) of the WEST-B as a prerequisite for admission to a teacher preparation program approved by the PESB. The same is also required of out-of state teachers seeking a Washington State residency certificate. This test is designed to reflect knowledge and

Information on teachers in the S-275 and the Washington State Credentials database can be linked to students via the state's *CSRS*, *CEDARS*, and *WASL* databases. The CSRS includes information on individual student background including gender, gender race/ethnicity, free or reduced-price lunch, migrant, and homeless statuses, as well as participation in the following programs: home-based learning, learning disabled, gifted/highly capable, limited English proficiency (LEP), and special education for the 2005-06 to 2008-09 school years. In 2009-10, CEDARS replaced the CSRS database. It contains all individual student background characteristics, but in addition, includes a direct link (a unique course ID within schools) between teachers and students. The WASL database includes achievement outcomes on the WASL, an annual state assessment of math and reading given to students in grades 3 through 8 and grade 10. (See **Table A2** in appendix for means of selected student characteristics by training program).

In addition to various teacher- and student-level data, we have compiled institution-level data from *The College Board* containing annual (since 1990) measures of selectivity based on the high school grades and standardized test scores of incoming freshman or admissions rates. We also use school- and district-level data from the *Washington State Report Card,* which includes aggregated information on student and teacher demographics and student test scores (assessed by the WASL).

We combine data from these sources to create a unique dataset that links teachers to their schools and, in some cases, their students in grades 3 through 6 for the 2005-06 to 2009-10 school years in both math and reading. Due to data limitations, not all students in grades 3 through 6 across these five school years can be linked to their teachers. This is largely due to the fact that, until recently, the state has only kept records of the names of individuals who proctored the state assessment to students,

skills described in textbooks, the Washington Essential Academic Learning Requirements (EALRs), curriculum guides, and certification standards.

not necessarily the students' classroom teacher.[21] Across all these grades and years we were able to match about 70 percent of students to their teachers and estimate value-added models (VAMs) of teacher effectiveness.[22]

Our analytic sample includes 8,732 teachers for whom we can estimate VAMs and for whom we know their initial teacher training program as being either from one of 20 state accredited teacher preparation programs or from outside of the state.[23] These teachers are linked to 293,994 students (391,922 student-years) who have valid WASL scores in both reading and math for at least two consecutive years.

Table 1 reports selected student characteristics for the 2008-09 school year for an unrestricted sample of students, i.e., those in grades 4–6 who have a valid WASL math or reading score but could not be matched to a teacher, and our restricted analytic sample described above. T-tests show that while nearly all of these differences are statistically significant, none of them are very large.[24]

As we noted above, a difficulty in separating selection from teacher training effects is that individuals and training programs jointly choose each other. Some argue that most of the nation's teachers are prepared in programs with relatively lenient admission and graduation standards (Levine,

---

[21] The proctor of the state assessment was used as the teacher-student link for the data used for analysis for the 2005-06 to 2008-09 school years. The assessment proctor is not intended to and does not necessarily identify the subject-matter teacher of a student. The "proctor name" might be another classroom teacher, teacher specialist, or administrator. Because a listed proctor may not necessarily be a student's subject-matter teacher, we take additional measures to reduce the possibility of inaccurate matches. For example, we limit our analyses to elementary school data where most students have only one primary teacher and only include matches where the listed proctor is reported (in the S-275) as being a certified teacher in the student's school and, further, where he or she is listed as 1.0 FTE in that school, as opposed to having appointments across various schools. With all that said, in the beginning of 2009-10, OSPI revised their data system to maintain a direct link between students and their classroom teachers via a unique "course id". And for the 2009-10 school year, we are able to check the accuracy of these proctor matches using the state's new Comprehensive Education Data and Research System (CEDARS) that matches students to teachers through a unique course ID. Our proctor match agrees with the student's teacher in the CEDARS system for about 95% of students in both math and reading.

[22] For the 2005-06 to 2008-09 school years where the proctor name was used as the student-teacher link, student-to-teacher match rates vary by year and grade with higher match rates in earlier years and lower grades. For a breakdown of match rates by subject, grade, and year see **Table A3** in the appendix.

[23] Lesley University produced its first teachers in 2009-10. So, although it is an accredited institution in Washington State, our observation window precludes it from being included in our analysis.

[24] Comparisons for other years reveal similar results.

2006); however, in Washington there is clearly heterogeneity in the selectivity of programs preparing teachers. For instance, the University of Washington Seattle (UW Seattle) is considered the flagship university in the state and in 2009, the 75th percentile composite SAT score of incoming UW freshman was about 1330. Nearly every other program in the state had lower 75th percentile SAT scores ranging between 1070 and 1290.[25] And, a few accredited programs do not require applicants to submit admissions test results in order to be considered for admission.[26]

Yet despite the differences in university selectivity, there is significant heterogeneity in at least one measure of pre-service academic skills for teachers who attend different training programs (and these, of course, are just those who ultimately make their way into the teaching profession). Specifically, since 2002, prospective teachers in the state have been required to take the Washington Educator Skills Test-Basic (WEST-B).[27] As is apparent from the boxplots in **Figure 1**, which show the distribution of teachers' first score on the WEST-B math and reading tests, by program, for teachers in our analytic subsample, there is significant overlap in teacher scores across programs in Washington.[28]

Results

Prior to describing our findings on teacher credentialing programs, a few peripheral findings warrant brief notice. As is typically found in the literature (Jackson & Bruegmann, 2009; Jacob, Lefgren, and Sims, 2008; Rivkin et al., 2005), in both reading and math and across model all specifications reported below, there are significant differences between student subgroups in achievement. Specifically, we find in both math and reading that American Indian/Alaska Native, Hispanic, and black

[25] The one exception is the University of Puget Sound whose composite 75th percentile SAT score was 1340.
[26] Heritage University has an open enrollment policy. City University and Antioch University both focus on adult learning and bachelor's degree completion suggesting less stringent entrance requirements.
[27] An individual's performance on the WEST-B is evaluated against a standard established by the PESB and was based on the professional judgments and recommendations of Washington educators. Test results are calculated as scaled scores in a range from 100 to 300; a score of 240 is the minimum passing score. (http://www.west.nesinc.com/WA9_passingrequirements.asp)
[28] These test scores distributions for our subsample look very similar to those for all teachers who graduated after 2002 and were teaching in the state any time between 2006-07 and 2009-10.

students score lower than their white counterparts; Asian/Pacific Islander students score higher than that same referent group; female students score higher than males in reading, but lower than males in math; students who are eligible for free or reduced-price lunch programs, those with reported learning disabilities, and those enrolled in LEP or special education classes also score lower in both math and reading than similar students who are not eligible for or enrolled in those programs; and students involved in gifted/highly capable programs score higher in both math and reading than those students who are not.[29]

At the classroom level, class size plays an influential role in explaining student achievement. In our models for both subjects, the coefficient on class size is negative (-0.002 in math and reading) and statistically significant, meaning that students in larger classes typically score lower than similar students in smaller classes, though the effects are pretty small. For instance, a ten student decrease in class size is estimated to increase student achievement by only 2 percent of a standard deviation.

Additionally, the model specifications described below include teacher covariates (which are not reported, but available upon request).[30] Again, consistent with the literature (Boyd et al., 2005; Clotfelter et al., 2007; Rivkin et al., 2005; Rockoff, 2004), we see the largest productivity gains for teachers early in their careers. For instance, students with a teacher who has one to two years of experience outperform students with novice teachers by about 4 percent of a standard deviation (depending on model specification), and students with teachers who have three to five years of experience tend to outperform those with one to two years of experience by about an *additional* 2 percent of a standard deviation (though the difference is not statistically significant across all model specifications). We find little evidence, however, of statistically significant productivity gains associated with increases in experience beyond five years.

---

[29] We do not report these coefficients in the tables below, as they are well-known findings, but the results are available from the authors upon request.
[30] Estimates of institutional effects change very little whether or not teacher covariates are included in the model.

Research on the attainment of graduate degrees offers mixed findings (Hanushek, 1997; Goldhaber & Brewer, 2000; Harris and Sass, 2007). We find small, but statistically significant positive effects for students whose teachers hold at least a master's degree.[31] To assess whether the attainment of an advanced degree affects teacher productivity, we estimate models that include teacher fixed effects, along with an indicator for the years that a teacher held a master's or higher degree. In these models, which identify the effect of an advanced degree based on within-teacher variation of effectiveness, the coefficient for holding at least a Master's degree remains positive, but is not statistically significant at conventional levels.[32] Thus, consistent with recent evidence from Florida (Harris and Sass, 2007), our findings show little evidence that the process of obtaining graduate degrees is beneficial to students, even if those who do obtain the degree are generally more productive.[33]

Unfortunately, our data do not allow us to assess the difference in effectiveness between teachers who got their bachelor's degree and those who got their master's degree from a given program. That is, although we know which degree a teacher has earned and we know which degrees teacher preparation programs currently offer (e.g., teaching certificate only, Bachelor's, Master's) we do not know which degree a teacher earned in conjunction with his or her initial teaching certificate.

How Much of Student Achievement is Explained by Credentialing Programs?

In this section, we investigate how much of the variation in student achievement/teacher effectiveness can be explained by teachers' training programs. We begin by noting that F-tests confirm

[31] This finding is consistent across all model specifications ranging from 0.005 to 0.007 in math and remaining around 0.010 in reading. Put in context, these effects are, at most, about one quarter the size of the difference between a teacher with 1 to 2 years of experience and a novice teacher.

[32] For math, the coefficient for obtaining an advanced degree is 0.013 with a standard error of 0.010. For reading, it is 0.018 with a standard error of 0.011.

[33] One shortcoming of this research on degree level is that it does not account for the timing of the degree receipt. It is possible that a masters program raises teacher effectiveness early in the program, prior to teachers receiving the credential, so that the within teacher estimator that compares teacher productivity pre- and post-degree finds little change.

the joint significance of training program indicators.[34] ANOVA models are used to explain the variation in student achievement explained by training programs and other teacher characteristics. Given the match between teachers and training programs is not random, we estimate both Type I and Type III sum of squares for the program indicators from estimating equation (1) above; this provides an upper and lower bound of the proportion of the variation of student achievement (in math and reading) explained by the training programs.[35] Finally, we replace all time-invariant teacher variables in the model with teacher fixed effects to determine the total variation of student achievement explained by the stable component of teacher effectiveness.[36]

Table 2 shows that, in practice, it matters little whether we use the Type I or Type III estimates, as the difference between them is small. In either case, only a small proportion of the total variation in student achievement is explained by program indicators; and less than one percent of the total explained sum of squares is accounted for by the program indicators. However, these program indicators still explain more of the variation in student achievement in both math and reading than do exogenous teacher characteristics (i.e., gender and race). And, in math, these programs indicators explain more of the variation in student performance than do teacher credentials (i.e., degree level and experience).[37] These results suggest that while there is information about teacher effectiveness that can be derived based on knowing where teachers received their credentials, the vast majority of the heterogeneity of teachers is not associated with the program from which they received their

[34] The joint F-tests for all of the institution dummies in math and reading models are significant at the 1 percent confidence level.

[35] In estimating the Type I sum of squares, we include the training institutions before teacher credentials (but after exogenous teacher characteristics, like race and gender) so this estimate is the upper bound on the proportion of variation explained by programs.

[36] Results are similar whether we use teacher observables or teacher fixed effects. But, note that teacher effectiveness may change over time with experience or additional training. Sass (2008), for instance, estimates that roughly 50 (at the elementary level) to 70 (at the middle school level) percent of the total variation in teacher effectiveness is between teachers, with the remainder being within teacher variation in effectiveness over time.

[37] In reading, teacher credentials explain slightly more than program indicators. Ultimately we note that as proportion of the total variance all of these differences are rather small.

endorsements. In the next section, we explore how *individual* teacher training program estimates compare to one another.

Assessment of Training Programs

We present the individual teacher training program coefficient estimates in **Table 3**. Columns 1 and 5 report the base model (the model with school and district covariates rather than fixed effects) estimates for the full analytic sample in math and reading respectively.[38] The Empirical Bayes (EB) adjusted estimates of the program indicators (a one standard deviation change in the estimates) is 0.01 in math and 0.02 in reading.[39] To put this in context, these differences are roughly equivalent to our estimates of changes in class size by between five to ten students.

Using the same model to estimate individual *teacher* effects (except excluding those district-level variables where there is little variation within teachers), the standard deviation is 0.18 in math and 0.20 in reading. So, the standard deviation of the program estimates is about 7 to 9 percent of the standard deviation of the teacher effects. This finding differs from Boyd et al. (2009) who find that "a one standard deviation change in the effectiveness of the preparation program corresponds to more than a third of a standard deviation in the teacher effect for new teachers" (p.429). We can only speculate why we find less heterogeneity in our program estimates than Boyd et al. It is possible these differences result from the fact that New York has more training programs (i.e., 30) than does Washington, that New York training programs draw potential teachers from a different distribution, or that training programs in Washington are more tightly regulated and so more similar to one another.

[38] As noted above, there are currently 21 regionally accredited programs in Washington State. Lesley University, however, did not begin graduating students until 2009-10 and is therefore not included in our analysis.

[39] To do this we calculate estimates for the program estimates that use as many years of data that are available to inform each individual estimate and then adjusts for variance inflation by shrinking the estimates back to the grand mean of the population in proportion to the error of the individual estimate. For more detail on this Empirical Bayes methodology, see Aaronson et al. (2007).

Many of these programs produce teachers who are statistically indistinguishable from teachers who were trained out-of-state and later approved by OSPI to teach in Washington, but the point estimates for the individual program indicators show meaningful differences in teacher effectiveness (this is illustrated in **Figure 2**, which plots the point estimates and 95 percent confidence intervals from our base model for each program in both math and reading).[40] For instance, in the base model (which omits district or school fixed effects), the average difference between teachers who receive a credential from the least and most effective programs is about 7 percent of a standard deviation in math and 10 percent of a standard deviation in reading.[41]

To put the above findings in context, the average expected difference in student performance between having a math teacher from the most effective program and the least effective program is equivalent to the regression-adjusted difference between students who are eligible for free or reduced-price lunches and those who are not (ß=-0.075). For reading, this same difference is roughly comparable to the difference between students with learning disabilities and those without (ß=-0.091). These differences are roughly twice as large as the estimated return to one or two additional years of teaching experience beyond the first year of teaching, which is about 4 percent of a standard deviation.

Focusing on the individual estimates in the base math and reading models (columns 1 and 5, respectively), there are a few surprises. For instance, there are relatively selective institutions (based on SAT scores) that do not appear to be graduating particularly effective teachers and less selective institutions that appear to be producing very effective teachers. UW Tacoma, for example, whose composite 75th percentile SAT score of 1120 was the fourth lowest of all other institutions requiring SATs

[40] In **Figure 2**, a value of zero represents our referent category, i.e., teachers who were credentialed outside of Washington State. When comparing teachers from any given institution to those from out-of-state, we look to see where the point estimate falls (above or below zero) and whether the confidence interval crosses zero. When the confidence interval does not cross zero, the effect is statistically significant. For example, we see that UW Tacoma produces both math and reading teachers who are significantly more effective than out-of-state teachers. On the other hand, teachers from Eastern Washington University are significantly less effective at teaching both math and reading compared to out-of-state teachers.

[41] These findings are roughly consistent with Boyd et al. (2009) who report a difference of 0.07 standard deviations in both math and reading between the highest performing institution and the average institution (p.428).

in 2009, has graduated some of the most effective math and reading teachers in the state. Whereas more selective institutions, such as Seattle University and Gonzaga University, with composite 75th percentile SAT scores in 2009 that were above 1270, graduated teachers who are significantly less effective in math (Seattle University) or reading (Gonzaga) compared to out-of-state teachers or graduates from a handful of other in-state programs. This perhaps points to the importance of the training prospective teachers receive, however, this is largely speculation because the SAT selectivity measure applies to all students at an institution, not just teacher trainees.

We attempt to account for selection into training programs in our models with the inclusion of various measures of institutional selectivity: the composite (math and verbal) 75th percentile score on the SAT for incoming freshman, the percent of incoming freshman whose high school grade point average was above 3.0, and admissions rates for incoming freshman (i.e., total admitted/total applied).[42] Joint f-tests show that together these three selectivity measures significantly improve model fit in both subjects.

Turning to the coefficients of these measures we see small, positive effects for the composite 75th percentile SAT score of the institution and the percent of incoming freshman whose high school GPA was above 3.0 for both subjects; only the latter is statistically significant in reading (ß = 0.001). The effect of admissions rates is statistically significant for both math (ß = 0.001) and reading (ß = 0.002). However, its *positive* direction is somewhat surprising given that we might expect *more* selective institutions—those with *lower* admissions rates—would graduate *more* effective teachers.[43]

[42] Since most training programs are part of four-year institutions, we subtract four years from a teacher's certification date to approximate their entry year into their certifying institution. Teachers who were certified out of state (22%), entered school before 1990 (31%), or graduated from institutions that don't report selectivity data (9%) are missing institutional selectivity data, but are included in the regression with a dummy indicator for missingness.

[43] We further investigate this finding by quantifying admissions rates as quartiles and re-running the model with the highest quartile as the referent category. Each of the coefficients for lower three quartiles are increasingly negative (Q1= -0.015, Q2= -0.019. Q3= -0.023) and statistically significant reiterating our rather unusual finding that teacher effectiveness is positively correlated with admissions rates.

Columns 2 and 6 in **Table 3** show the program estimates of these selectivity models for math and reading, respectively.[44] Importantly, although institutional selectivity plays in influential role in student achievement outcomes, as noted above, accounting for such selectivity does not noticeably change the individual program estimates. Indeed, the spearman rank correlation coefficients between the program estimates of the base model and the program estimates of the selectivity model are 0.90 for math and 0.85 for reading (see **Table 4**).

The above controls are for the selectivity of the institutions, and may be a poor proxy for individuals' capacity at the point that they enter teacher training. We attempt to account for individual teaching capacity with a sub-sample of teachers who were required to pass basic skills tests (the WEST-B tests, described more fully above in the data section) prior to entering a state-approved training program. Since individuals are allowed to take the exam more than once and may be contingent on the number of times that the test is taken, we use teachers' first scores on the WEST-B tests as controls for teacher selection. Also note since we only observe test scores for trainees who made it into the teaching profession, this sample is necessarily selective.[45] The WEST-B coefficients are positive and statistically significant for both subjects indicating that students with higher-scoring teachers score higher on their own tests.[46] Comparing program estimates from this WEST-B model to those of the base model suggests important differences as the Pearson correlations between the two models are only 0.24 in math and 0.59 in reading. However, it is premature to jump to any strong conclusions mainly because the WEST-B sample is substantially different from the full analytic sample.[47] To determine whether the differential we observe is due to the controls for selection or the change in sample, we re-estimate the base model for the WEST-B sample without including the actual test scores and find that

[44] Spearman correlations across model specifications (i.e., base model, district fixed effects, and school fixed effects) are all above 0.51 for math and 0.86 for reading.
[45] Of our full sample of 8,732 teachers, only 1,471 (16.8%) have WEST-B scores largely because the majority of teachers (roughly 85% percent) were enrolled in a program before the WEST-B requirement was put into effect.
[46] The WEST-B coefficient for the math model is 0.02 with a standard error of 0.003. For reading, the coefficient is 0.01 with a standard error of 0.003.
[47] See **Table A4** in the appendix for the number of teachers who graduate from each program.

the correlation between these models is above 0.98 for each subject. This implies that the differentials in programs estimates are driven by the sample, not the inclusion of the WEST-B scores. Moreover, it suggests training program estimates change significantly over time since the since the WEST-B sample only includes recent program graduates. (We return to this issue below in sub section D.)

The second type of selection we are concerned about is related to the possibility that teacher candidates from different preparation programs are funneled into school systems or schools that are systematically different from each other in ways that are not accounted for by variables included in our models.[48] It is fairly standard to account for the type of districts and schools in which teachers are employed by estimating model specifications that include district or school fixed effects, as we do in columns 3 and 4 of **Table 3,** for math, and columns 7 and 8, for reading. However, it is important to consider what these models actually tell us. To the degree that they capture any time-invariant district or school effects that may confound our program estimates in the base model, they provide a more accurate picture of the effect of training programs. However, it is not totally clear that district or school fixed effects models will yield unbiased program estimates. The reason is that the estimates are based solely on within district or school differences in teacher effectiveness and some of the differences between programs may lead to systematic sorting across different types of districts or schools. Imagine, for instance, that there are large differences between programs, but schools tend to employ teachers of a similar effectiveness level. In this case, a school that employs teachers that are average in effectiveness, from multiple programs, would tend to have some of the least effective teachers from the best training programs and most effective teachers from the worst training programs and thus the within school comparison would tend to show little difference between the programs. The true

[48] This wouldn't be a concern if program trainees were randomly sorted across districts and schools, but to the extent that this sorting process is nonrandom (e.g., programs act as "feeder institutions" to districts), it may be that districts and/or school effects confound the estimate of programs effects.

differences between programs are one of the reasons for the sorting of teachers across schools so the within school comparisons lead to a washing out of the program estimates.[49]

There are several other issues that arise in estimating fixed effects specifications of Equation (1). Mihaly et al. (2011), for example, discuss the assumptions and consequences of including school fixed effects in a model measuring the effectiveness of teacher preparation program graduates. In particular, they argue that for these models to produce consistent and unbiased estimates of programs estimates, data must meet two assumptions: identifiability and homogeneity. Identifiability refers to the connectedness of training programs and schools and the representation of teachers from different preparation programs in the same schools. If, for example, a school employs teachers from only one preparation program then that school is not connected to other teacher training programs and, in a model with school fixed effects, their school effect is collinear with their program estimate and those teachers and will drop from the model without contributing to the overall estimation of the program estimate leading to variance inflation. Homogeneity, on the other hand, refers to the assumption that programs estimates for "highly centralized" schools (those with teachers from four or more preparation programs) are not significantly different from those for less connected schools. In other words, if highly centralized schools serve a substantially different population of students than other schools in the state, and if the student covariates in the model do not fully account for these differences, then the assumption of homogeneity does not hold and the program estimates could be biased.[50]

We follow Mihaly et al.'s (2011) procedures for testing these assumptions and conclude there is minimal concern about either in our data. We are not overly concerned with identifiability since only about 15 percent of the schools in our sample (which employed about 5 percent of all teachers) were

---

[49] The analogous issue arises when estimating individual teacher effectiveness and making decisions of whether or not to include school level fixed effects. We thank Jim Wyckoff for his insights on this matter. (Personal Communication, August 2011).

[50] In theory these same concerns could arise in the case of district effects, but, in practice, there is considerable overlap of teachers from different training institutions in the same districts.

staffed with teachers from a single training program.[51] As to the concern about homogeneity, our two independent sample t-tests comparing the characteristics of students and teachers in schools with teachers from one training program to those in schools with teachers from four or more training programs reveal that there only a few significant differences (i.e., students in schools with teachers from only one training program are significantly more likely to be American Indian or white and have teachers with at least a master's degree, whereas students in schools with teachers from four or more training programs are significantly more likely to meet state math standards, be Black, Asian, and/or bilingual). And while these differences are not large, our training program estimates could be biased if the student-level covariates do not fully control for these differences.

Given the above discussion, we are not terribly concerned that our findings are likely to be sensitive to model specification and this is verified by scanning across the different specifications (i.e., base models, district fixed effects, and school fixed effects) reported in **Table 3**. These program estimates appear to be similar both within subjects, across model specifications, and within model specifications across subjects. This is verified by the Spearman correlations reported in **Table 4**, which are all above 0.57 for math and 0.80 for reading.[52]

Investigating the Potential of Program Specialization

The above findings provide program estimates in general but do not allow for the possibility that training programs may specialize in preparing teachers to serve particular types of students.[53] Indeed, as Boyd et al. (2009) point out, the degree to which the quality of training programs is contingent on

[51] This contrasts sharply with Mihaly et al's sample of Florida teachers wherein 54.1 percent of schools had teachers from a single training institution suggesting that programs in Florida serve much more localized markets than those in Washington State.

[52] As reported in the Table, most of the Pearson correlations are also high. Additionally, there is also a fair degree of consistency of program estimates across subject. As we report in **Table 4**, for instance, the within-model, across-subject Spearman correlations are all are above 0.43. Pearson correlations are generally lower, but all are above 0.26.

[53] The above results also do not consider the possibility that training institutions and school systems may collaborate in ways that lead to school district specialization (e.g., student teaching, curricula use).

where a teacher works (i.e., that certain programs may serve districts with particular student populations particularly well), begs important questions about teacher preparation. For example, Heritage University's website notes that the institution graduates "more English Language Learner endorsed educators than any other institution in the state."[54] And UW Seattle's teacher preparation program promotes itself as offering "fieldwork opportunities in [a] network of partner schools, all located in culturally diverse urban communities around the Puget Sound area."[55]

To investigate whether there are differences in program effectiveness for different student subgroups, we estimate the base model for both math and reading with interactions between program indicators and selected student characteristics (i.e., whether a student is eligible for free lunch, is receiving LEP services, is Black, is Asian, and/or is Hispanic). The estimates and standard errors for the interaction terms from these models are reported in **Table 5**.[56] Interactions terms for *all* five subgroups (100 total interactions for each subject) are jointly significant for both math and reading. We caution, however, not to overstate these findings. Joint F-tests for the interaction effects of *each* of the five subgroups show that only two subgroups (free lunch and Asian) significantly improve model fit for both math and reading.[57]

For math and reading, a few (13 percent) of the 100 interaction effects are statistically significant.[58] This is a greater proportion than one would expect from chance alone and the findings across subjects suggest some consistency in the subgroup effects in the sense that there are a few

[54] Source: http://www.heritage.edu/LinkClick.aspx?fileticket=sBIF4rVDnV0%3d&tabid=269

[55] Source: http://education.washington.edu/areas/tep/

[56] We also estimated these same models with district and school fixed effects. Spearman rank correlations of the interaction effects across models within subject are all above 0.88 in math and 0.95 in reading.

[57] More specifically, F-tests show that interaction effects for the Free Lunch, Asian, and Hispanic subgroups are each jointly significant for the math model while Free Lunch, Asian, and LEP subgroups are each jointly significant for the reading model. These patterns of significance are the same for model specifications with district or school fixed effects with the one exception that the interaction effects for the LEP subgroup are not jointly significant in the school fixed effect model for reading, at least at conventional levels. They are, however, marginally significant.

[58] These findings are robust to model specification (i.e., across the base, district fixed effects, and school fixed effects models) as Pearson correlations of interaction effects across models are all above 0.93.

programs that have the same direction of effects across disadvantaged student subgroups and a few programs are significant in both math and reading for the same subgroups.[59]

In a few instances, we see that teachers from a program that is generally less effective are actually more effective at teaching specific subgroups of students and vice versa. For example, columns 1 and 5 in **Table 3** show that, on average, teachers from Eastern Washington University are significantly less effective than teachers trained out of state by about 1 to 2 percent of a standard deviation in math and reading, respectively. But, interaction effects in **Table 5** show that these same teachers appear to be significantly better at teaching LEP students relative to non-LEP students, by about 5 percent of a standard deviation in math and 7 percent of a standard deviation in reading. The opposite is true for math teachers from UW Seattle who are, on average, more effective than out-of-state-trained teachers by roughly 2 percent of a standard deviation (see column 1 in **Table 3**) but appear less effective at teaching students who are eligible for free or reduced-price lunch compared to students who are not, by about 5 percent of a standard deviation in math and 2 percent of a standard deviation in reading (see **Table 5**).

In addition to having differential effectiveness by student subgroups, it may be that teachers from particular programs are better prepared to teach students in or near the school districts where they were trained because pre-service teachers work extensively with local in-service teachers in the field as part of their practicum training. This would make sense since there is evidence that teacher labor markets are quite localized. Boyd et al. (2005), for example, report that "an individual is twice as likely to teach in a region that is within five miles of his or her hometown as one 20 miles away and about four times as likely to teach in a region within five miles of his or her hometown as one 40 miles away" (pg. 123). To test how proximity to training institutions teacher effectiveness in our sample, we

[59] Pearson correlations of interaction effects across subjects within subgroups are: 0.66 for Free Lunch, 0.45 for LEP, 0.05 for Black, 0.51 for Asian, and 0.71 for Hispanic. These correlations are similar for models with district or school fixed effects.

calculate three mutually exclusive proximity indicators to capture whether a teacher, in any given year, taught in school district within 10, 25, or 50 miles of their training institution and then re-estimate our base model including these proximity indicators.[60][61] Results from the student achievement models that include indicators for teacher employment proximity to training institution show inconsistent results across math and reading achievement. In math the interaction is positive and significant for teaching in a school district within 10, 25, or 50 miles from one's training institution.[62] One might think, therefore, that training programs and local school districts somehow collaborate on curricula use or pedagogy thus making graduates from nearby programs especially effective teachers for local student populations. However, we caution that this seemingly strong evidence is tempered by the results from the reading model, which shows negative and at least marginally significant effects for teaching within 10 and 50 miles of one's training institution (the indicator is positive but insignificant for teaching within 25 miles).[63] Thus, it is difficult to know what to make of these findings.

Testing for Programmatic Change

It is quite possible that teacher training programs change over time due to state mandates or institutional initiatives designed to improve teacher training and subsequent teacher effectiveness across the board. In 2008, for example, the PESB implemented the WEST-E, a content knowledge exam required for all candidates applying for endorsements on their initial teaching certificate designed to

[60] These distances were calculated using ESRI's ArcGIS software. More specifically, the Generate Near Table tool in ArcMap calculates the distance between an XY point (i.e., a training institution) and the closest line of a polygon (i.e., a school district) within a specified search radius. If any portion of a school district falls within the given radius, it is included in that proximity measure.

[61] Although we know the name of the institution that issued a teacher's credential, we do no know the specific physical site a teacher attended to obtain that credential (e.g., the number of total sites by program with satellite sites are: Central University: 6, City University: 7, Eastern Washington University: 2, Heritage University: 5, Saint Martin's University: 3, Washington State University: 4, and Western Washington University: 4.) We assume that most of the teachers in our sample attended the pro Have chosen to focus on the primary site for each institution. But, to the extent that large percentages of teachers in our sample attended other sites our results will be bias.

[62] In math, the coefficient for teaching in a school district within 10 miles is 0.014 while the effects for teaching in schools district within either 25 or 50 miles are 0.023 and 0.008, respectively.

[63] In reading, the coefficient for teaching in a school district within 10 miles is -0.008 while the effects for teaching in schools district within either 25 or 50 miles are 0.003 and -0.007, respectively.

more fully align with Washington standards than was the case with previous exams. The expectation is that teachers who were required to pass this exam should be more effective in the classroom (especially in math) than teachers from earlier cohorts who were not held to such a standard. We cannot currently test whether this requirement has affected the quality of the teacher workforce since the test was only required of teachers who could have begun teaching in 2009-10 and too few of these teachers are in our analytic sample, but we will be able to assess this as new cohorts of teacher data become available.

It is also the case that the relative magnitude of individual program estimates could change over time due to institutional changes of individual programs.[64] To test this, we estimate variants of our base model that include interaction terms between program dummies and grouped cohorts of recent graduates.[65] More specifically, we focus on two grouped cohorts: (1) graduates between 2005 and 2009 and (2) graduates between 2000 and 2004. There is nothing special about those particular years. However, we wished to pick a time span over which the program estimates would be informed by a reasonably large number of teachers (at least 30).[66]

Prior to reporting whether individual program estimates have changed over time, it is worth noting that we compare the effectiveness of teachers trained in Washington State, as a whole, to those who received their training out-of-state and credentials from OSPI, and, in particular, whether there was evidence of changes in relative effectiveness over time. To do this, we estimate a model that includes an in-state training indicator variable and interactions between this variable and having been trained recently, i.e. whether a teacher was certified within the last five years (2005 to 2009) or the five years

[64] In 2003, for example, the University of Washington received a matching grant for 5 million dollars over a five-year period (2003 to 2008) from The Carnegie Corporation of New York's Teachers for a New Era Project to implement and examine changes in its teacher training program. However, future teachers from UW did not immediately experience such changes. For example, the four cohorts from 2004-05 to 2007-08 each received only parts of the renewed program. The first cohort in the fully renewed program got their certification in spring 2009. Because of the induction year, they didn't finish the full program until spring 2010.

[65] Interaction effects from models with district or school fixed effects look similar to those from the base models and all have spearman rank correlations above 0.87 in math and 0.72 in reading.

[66] Because some programs graduated few (if any) graduates during these years, we focus the twelve programs that graduated at least 30 new teachers during the years specified by each cohort grouping. All other programs are collapsed and their combined effects are reported together.

prior to that (2000 to 2004).[67] In math, the interaction terms for both certification cohorts are positive and significant. In reading, both interaction terms are positive, but only the one for the most recent cohort (2005-09) is statistically significant. [68] These results suggest teachers who were trained in Washington State within last five to ten years are relatively more effective than those who had been credentialed in-state prior to 2000, at least as compared to teachers who were credentialed by OSPI.

We now turn to changes in individual program effectiveness. **Table 6** shows results for two models (math and reading) that include interaction effects for each of the programs by cohort grouping.[69] Columns 1 and 4 give the main program estimates for graduates before 2000 in math and reading respectively. Columns 2 and 5 show the interaction effects for teachers who graduated from each program between 2000 and 2004. Columns 3 and 6 show the interaction effects for teachers who graduated from each program between 2005 and 2009.[70] As we might expect given the above finding regarding the effectiveness of teachers who are trained in-state versus out-of-state, several of the interaction effects are positive and significant for each subject and cohort grouping. This suggests that, for some programs, more recent cohorts of teachers who graduated from a particular program are more effective than teachers who graduated from that program before 2000 relative to changes in effectiveness for teachers trained outside of Washington. For example, teachers trained at UW Bothell before 2000 were essentially indistinguishable from teachers trained out-of-state during that same time period. But by 2000-2004, UW Bothell graduates were more effective at teaching reading than prior

[67] F-tests show that the interaction terms are jointly significant for both math and reading models.
[68] For math, the coefficient for the interaction between being trained in Washington and being certificated between 2005 and 2009 is 0.020 with a standard error of 0.007. The coefficient for this interaction with the 2000 to 2004 cohort is 0.015 with a standard error of 0.006. For reading, the interaction effect for the in-state trained 2005-09 cohort is 0.023 with a standard error of 0.007. This same interaction term for the 2000-04 cohort is 0.008 with a standard error of 0.007.
[69] F-tests show that interaction effects are jointly significant for both math and reading models.
[70] The total estimated effectiveness (relative to a teacher who is trained out-of-state and, hence, whose credential is obtained from OSPI) for a teacher in a recent cohort is the *sum* of the main effective and the cohort-time period interaction term. For instance, relative to a teacher trained out-of-state, the estimated effectiveness for math teachers obtained a credential from Seattle University in the 2000-04 time period is .026, the sum of the main effect (-.058) and the 2000-04 interaction term (.084).

graduates and by 2005-2009 they were also more effective at teaching math (relative to any changes in effectiveness for out-of-state teachers).

Thus, we find some evidence that the relative effectiveness of graduates from various programs has changed over time and, as a whole, Washington teachers recently trained in-state appear relatively more effective than those who received training in the past. However, all programs are measured relative to those teachers who received training out-of-state, so we cannot say whether this is a reflection of the effectiveness of teachers who received an in-state credential or the possibility that there is a change in the quality of teachers who are coming in from out-of- state.[71]

Conclusions

It is important to note several caveats about the analyses we presented here. First, the samples used to detect differential program estimates for student subgroups and programmatic change over time were relatively small so it is conceivable that effects do exist but their magnitude is too small to detect with the sample at hand. Second, our analyses are focused entirely on elementary schools and teachers. It is conceivable that comparable analyses would yield results that look quite different at the secondary level. Third, students' outcomes on standardized tests are only one measure of student learning, so it is possible that value-added approaches miss key aspects of what different training institutions contribute to teacher candidates.

Our findings suggest that where teachers are credentialed explains only a small portion of the overall variation in the effectiveness of in-service teachers. This is now a common finding in the educational productivity literature; it appears that the best assessments of teachers are those based on actual classroom performance rather than pre- or in-service credentials. That said, the differential in the average effectiveness of the teachers credentialed by various programs is meaningful, in fact it is at least

[71] And note that even for recent cohorts of teachers, out-of-state trained teachers are still relatively effective compared to those who received an in-state credential.

as important as years of experience and degree level. This means that improving teacher training has the potential to greatly enhance the productivity of the teacher workforce.

We find that programs credentialing teachers who are more effective in math are generally also credentialing more teachers who are more effective in reading. Moreover, the fact that teachers from different programs in Washington State tend to mix within districts and schools, and that the findings are robust to model specification suggests that our estimates reflect something about the selection or preparation of teachers from different institutions rather than the potential that the findings are driven by the districts and schools they sort into. Of course it is not possible with state administrative data to definitively determine whether these, or any of the program estimates, are related to the selection of individuals into teacher training programs or the training individuals in the programs receive. For some purposes, e.g., accreditation and accountability, the distinction between selection and training effects may not be relevant. However, to understand what aspects of teacher training appear relevant we clearly care about the distinction.

Perhaps the most important findings are that there is speculative evidence of student subgroup and regional specialization, and evidence that the program estimates do change over time. The former finding suggests that something about a specific program may lead teachers to be more or less prepared to teach specific students while the latter may be a reflection of selection or training. In either case, the very fact that there is evidence of specialization and changes over time suggests that programs may be differently focused and that improvement is possible.

There is no doubt that evaluating teacher training programs based on the value-added estimates of the teachers they credential is controversial. It is true that the value-added program estimates do not provide any direct guidance on how to improve teacher preparation programs. However, it is conceivable that it is not possible to move policy toward explaining *why* we see these program estimates until those estimates are first quantified. Moreover, it is certainly the case that some of the policy questions that merit investigation–-e.g., Do we see program change with new initiatives?; Do we see differences between programs within endorsement area?; How much of the difference between programs is driven by selection versus training?—require additional data. Some of these questions could be addressed with larger samples—in this case it is merely a matter of time—but other questions require additional information about individual programs, teacher candidate selection processes, and so forth. The collection of this kind of data, along with systematic analysis, would provide a path towards evidence-based reform of the pre-service portion of the teacher pipeline.

References

Aaronson, D., L. Barrow, and W. Sanders. (2007). "Teachers and Student Achievement in the Chicago Public High Schools." Journal of Labor Economics, 25(1), 95-135.

Ballou, D. (2005). "Value-Added Assessments: Lessons from Tennessee." In *Value Added Models in Education: Theory and Applications*. Edited by Robert Lissitz. Maple Grove, MN: JAM Press, 272-297.

Ballou, D., W. Sanders, and P. Wright. (2004). "Controlling for Student Background in Value-Added Assessment of Teachers." *Journal of Educational and Behavioral Statistics*, 29(1), 37-66.

Boyd, D. Lankford H., Loeb P. & Wyckoff, J. (2005). "The Draw of Home: How Teachers' Preferences for Proximity Disadvantage Urban Youth." Journal of Policy Analysis and Management, 24(1), 113–132.

Boyd, D., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2005). *How changes in entry requirements alter the teacher workforce and affect student achievement*. Cambridge, MA: National Bureau of Economic Research.

Boyd, D., Grossman, P. L., Ing, M., Lankford, H., Loeb, S., & Wyckoff, J. (2008), "Surveying the Landscape of Teacher Education in New York City: Constrained Variation and the Challenge of Innovation. Educational Evaluation and Policy Analysis, 30(4), 319-343

Boyd, D., Grossman, P., Lankford, H., Loeb, S., and Wyckoff, J. (2009). "Teacher Preparation and Student Achievement." Educational Evaluation and Policy Analysis, 31(4), 416-440.

Clotfelter, Charles, Lennie, Ladd, and Vigdor. (2006). "Would Higher Salaries Keep Teachers in Schools? Evidence from a Policy Intervention in North Carolina." National Bureau of Economic Research. Working Paper. June 2006

Clotfelter, C., Ladd, H., Vigdor, J. (2006) "Teacher-Student Matching and the Assessment of Teacher Effectiveness." The Journal of Human Resources, 41(4), 778-820.

Clotfelter, C., Ladd, H., Vigdor, J. (2007) "Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects." CALDER working paper #11, October 2007.

Corcoran, T. B. (2007). "The Changing and Chaotic World of Teacher Policy." In D. Cohen, S. Fuhrman & F. Mosher (Eds.), *The State of Education Policy Research* (pp. 307-335). Mahwah, NJ: Lawrence Erlbaum Associates.

Cochran-Smith, M., & Zeichner, K. M. (Eds.). (2005). *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education*. Washington, D.C.: The American Educational Research Association.

Constantine, J., Player D., Silva, T., Hallgren, K., Grider, M., and Deke, J. (2009). An Evaluation of Teachers Trained Through Different Routes to Certification, Final Report (NCEE 2009-4043). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of

Education Sciences, U.S. Department of Education.

Crowe, E. (2010). *Measuring what matters: A stronger accountability model for teacher education*. Washington, DC: Center for American Progress.

Darling-Hammond, L., Wise, A., & Klein. S. (1999). *A License to Teach: Raising Standards for Teaching*. San Francisco, CA: Jossey-Bass.

Decker,P., Mayer, D., Glazerman,S. (2004) “The Effects of Teach For America on Students: Findings from a National Evaluation.” Mathematica, Policy Research, Inc.

Duncan, A. (2010). “Teacher preparation: Reforming the uncertain profession”, *Education Digest,* 75 (5), 13-22.

Feistritzer, E. & Haar. C. (2010) “Research on Alternative Routes Education Research.” National Center for Alternative Certification. Downloaded from: www.teach-now.org/RESEARCH%20ABOUT%20ALTERNATE%20ROUTES.pdf on 3 August, 2011.

Gansle K., Noell, G., Knox R., Schafer M. (2010) “Value Added Assessment of Teacher Preparation programs in Louisiana: 2005-06 to 2008-09. Accessed from: http://regents.louisiana.gov/assets/docs/TeacherPreparation/2010VATechnical082610.pdf

Gatlin, (2009). “A Pluralistic Approach to the Revitalization of Teacher Education.” *Journal of Teacher Education*, 60(5) 469–477.

Glazerman, S., Mayer, D., Decker,P. (2006). “Alternative Routes to Teaching:The Impacts of Teach for America on Student Achievement and Other Outcomes.” *Journal of Policy Analysis and Management*, 25(1), 75–96.

Goldhaber, D. (2002). “The Mystery of Good Teaching.” *Education Next,* 2(1), 50-55.

Goldhaber, D. (2007). “Everybody's Doing It, But What Does Teacher Testing Tell Us About Teacher Effectiveness.” *Journal of Human Resources,* 42(4), 765-794.

Goldhaber, D., & Hansen, M. (forthcoming). “National Board Certification and Teachers' Career Paths: Does NBPTS Certification Influence How Long Teachers Remain in the Profession and Where they Teach?” *Education Finance and Policy*.

Goldhaber, D., & Hansen, M. (2009). “Race, Gender, and Teacher Testing: How Informative a Tool is Teacher Licensure Testing?” CRPE Working Paper #2008-2. Seattle, WA: Center on Reinventing Public Education, University of Washington.

Goldhaber, D. and Hansen, M. (2010). “Is It Just a Bad Class? Assessing the Stability of Measured Teacher Performance.” Center for Education Data and Research. Working Paper 2010-3.

Goldhaber, D., Brewer, D. (2000) “Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement.” Educational Evaluation and Policy Analysis. 22(2), 129-145.

Hanushek, E. A. (1986). “The Economics of Schooling: Production and Efficiency in Public Schools.” *Journal of Economic Literature,* 24(3), 1141-1178.

Hanushek, E. (1997). "Assessing the Effects of School Resources on Student Performance: An Update." *Educational Evaluation and Policy Analysis*, 19(2), 141-64.

Henry, G. T., Thompson, C. L., Bastian, K. C., Kershaw, D. C., Purtell, K. M., and Zulli, R. A. (2011) “Does teacher preparation affect student achievement?” Chapel Hill, NC: Carolina Institute for Public Policy Working Paper, Version dated February 7, 2011.

Jackson, C. Kirabo, and Elias Bruegmann. (2009). “Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers.” *American Economic Journal: Applied Economics*, 1, 85-108.

Jacob, Brian A., and Lars Lefgren. (2008). “Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education.” *Journal of Labor Economics,* 26, 101-136.

Johnston, Robert C. (1997). “Texas Boosts Accountability for Teacher Preparation.” *Education Week*, October 15. Accessed 3/27/09 at http://www.edweek.org/ew/articles/1997/10/15/07accred.h17.html.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2007). “Photo Finish: Certification Doesn’t Guarantee a Winner.” *Education Next*, 7(2).

Levine, A. (2006). *Educating School Teachers*. Washington, D.C.: The Education Schools Project.

McCaffrey, D., D. Koretz, J. Lockwood, and L. Hamilton. (2004). *Evaluating Value-Added Models for Teacher Accountability*. Santa Monica, CA, RAND Corporation.

Mihaly, K., D. McCaffrey, T. Sass, J.R. Lockwood (2011). “Where You Come From or Where You Go?” Distinguishing Between School Quality and the Effectiveness of Teacher Preparation Program Graduates. RAND Working Paper, March 18, 2011

National Council for Accreditation of Teacher Education (2010). *Transforming teacher education through clinical practice: A national strategy to prepare effective teachers.* Washington, DC.

National Council on Teacher Quality. (2007). “State Teacher Policy Yearbook: Washington State Summary.” Washington, DC. Accessed 3/27/09 at http://www.nctq.org/stpy/reports/stpy_washington.pdf.

National Research Council. 2010. “Better Data on Teacher Preparation Could Aid Efforts to Improve Education.” The *National Academies Office of News and Public Information*. April 29. Accessed from: http://www8.nationalacademies.org/onpinews/newsitem.aspx?RecordID=12882

Noell, G. H., Porter, B. A., Patt, R. M. & Dahir, A. (2008). “Value-added assessment of teacher preparation in Louisiana: 2004-2005 to 2006-2007.” Retrieved from: http://www.regents.state.la.us/Academic/TE/2008/Final%20ValueAdded%20Report%20(12.02.08).pdf

Noell, G.H., Gansle, K.A., Patt, R.M., and Schafer, M.J. (2009) "Value Added Assessment of Teacher Preparation in Louisiana: 2005-2006 to 2007-2008." Retrived from: http://www.laregentsarchive.com/Academic/TE/2009/2008-09VATechnical(8.24.09).pdf

Nye, B., Konstantopoulos, S., and Hedges, L. (2004). "How Large Are Teacher Effects?" Educational Evaluation and Policy Analysis. 26(3), 237-257.

Office of the Superintendent of Public Instruction. (2007). *Professional Certification Handbook: A Guide for Candidates, Facilitators, and Program Directors.* Olympia, WA: Office of Superintendent of Public Instruction. Available online at: http://www.k12.wa.us/certification/Teacher/ProCert-Handbook.aspx.

Rivkin, S., Hanushek, E., & Kane, T. J. (2005). "Teachers, Schools, and Academic Achievement." *Econometrica,* 73(2), 417-458.

Rockoff, J. E. (2004). "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review,* 94(2), 247-252.

Rochester, J. M. (2002). *Class warfare: Besieged schools, bewildered parents, betrayed kids and the attack on excellence*. San Francisco: Encounter Books.

Rothstein, J. (2008). "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." NBER Working Paper No. W14442. Cambridge, MA: National Bureau of Economic Research

Harris, D. & Sass, T. (2007). "Teacher training, teacher quality and student achievement." CALDER Working Paper No. 3. Washington, D.C.: National Center for Analysis of Longitudinal Data in Education Research.

Sass, T. (2008). "The Stability of Value-Added Measures of Teacher Quality and Implications for Teacher Compensation Policy." CALDER Brief 4. November 2008.

Stotko, E. M., Ingram, R., & Beaty-O'Ferrall, M. E. (2007). Promising strategies for attracting and retaining successful urban teachers. *Urban Education,* 42*,* 30-51.

The Teaching Commission. (2006). *Teaching at Risk: Progress & Potholes*. New York, NY.

Todd, P.E., and K.I. Wolpin. (2003). "On the Specification and Estimation of the Production Function for Cognitive Achievement." *Economic Journal* 113, F3-F33.

United States Congressional Senate. "Tennessee First to the Top Act of 2010". Tennessee Senate Bill 7005. Accessed at: http://www.capitol.tn.gov/Bills/106/Bill/SB7005.pdf on 7 September 2011.

United States Congressional Senate. Colorado Senate Bill 10-191. Accessed at: http://www.leg.state.co.us/clics/clics2010a/csl.nsf/fsbillcont3/EF2EBB67D47342CF872576A80027B078?open&file=191_enr.pdf on 7 September 2011.

Wilson, S. M., Floden, R. E., & Ferrini-Mundy, J. (2002). “Teacher preparation research: An insider’s view from the outside.” *Journal of Teacher Education,* 53*,* 190-204.

Xu, Z., J. Hannaway, and C. Taylor. (2007). “Making a Difference? The Effects of Teach For America in High School.” Calder working paper #17.

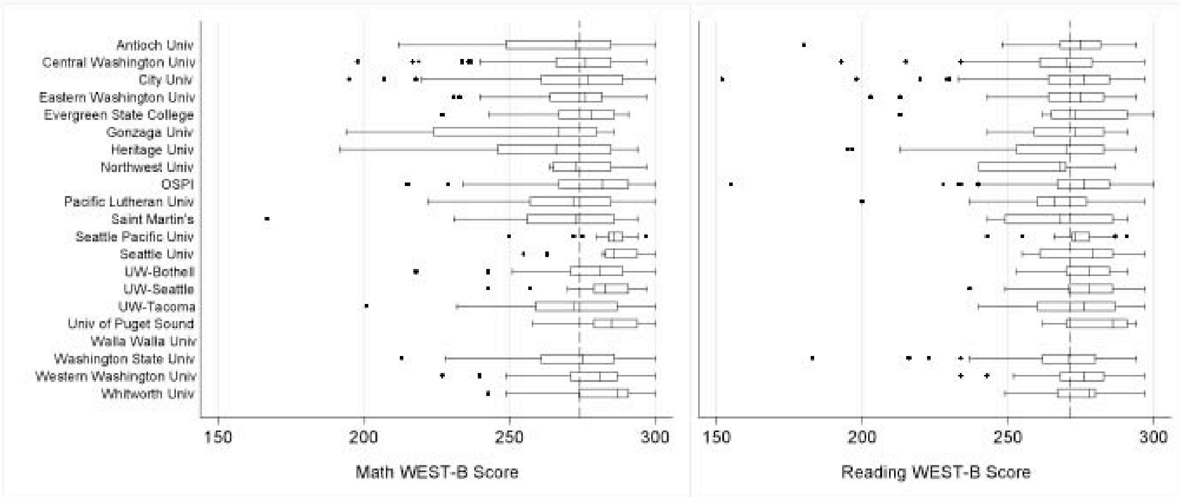
Tables and Figures

Table 1. Means of Selected Student Characteristics in 2008-09 for Unrestricted and Restricted Samples

	Unrestricted	Restricted	Difference
Math WASL Standardized Scale Score	-0.02	0.02	0.04***
Reading WASL Standardized Scale Score	-0.02	0.02	0.04***
Female Students (%)	48.2	49.1	-0.9**
American Indian Students (%)	2.4	2.3	0.1
Asian or Pacific-Islander Students (%)	9.8	8.6	1.2***
Black Students (%)	6.6	6.1	0.5**
Hispanic Students (%)	16.5	15.1	1.5***
White Students (%)	60.1	64.3	-4.2***
Multi-Racial Students (%)	4.1	3.3	0.8***
Migrant Students (%)	1.7	1.2	-0.5***
Homeless Students (%)	1.8	1.4	0.4***
English-speaking Students (%)	81.3	84.0	-2.7***
Students Eligible for Free or Reduced-Price Lunch (%)	47.4	44.7	2.7***
Students with Learning Disabilities (%)	5.8	6.1	-0.4*
Gifted Students (%)	4.9	3.6	1.3***
LEP Students (%)	9.9	6.8	3.1***
Special Education Students (%)	14.0	11.8	2.2***
Total Number of Students	53,177	58,865	-

Note: Students in the unrestricted sample could not be matched to unique proctors.
Statistical significance is denoted with: * if $p<0.05$, ** if $p<0.01$ **, and *** if $p<0.001$.

Figure 1. Boxplots for Math and Reading WEST-B Scores by Program



NOTE: In this figure, each box represents a teacher training program. The line inside each box equals the median score for teachers who obtained their first teaching certificate from that institution. The upper and lower ends of the box are the 75th and 25th percentiles of the score distribution for that institution. The "whiskers" extend to either the maximum and minimum scores unless those scores are 1.5 times the interquartile range (the difference between the 75th and 25th percentiles), in which case they are shown as outlier dots. The dashed vertical lines cutting across all institutions for each subject are the average scores for that exam (274 in math and 271 in reading).

Table 2. ANOVA Models Explaining Teacher Effectiveness

	MATH		READING	
	Type I Sum of Squares	Type III Sum of Squares	Type I Sum of Squares	Type III Sum of Squares
Teacher Characteristics	35.9	35.8	31.2	30.8
Program Indicators	65.5	47.9	55.7	40.0
Teacher Credentials	31.9	28.8	78.2	70.5
Explained Sum of Squares	181,181		149,734	
Total Sum of Squares	277,090		276,006	
Number of Observations	278,867		278,867	
R^2	0.65		0.54	

Table 3. Program Estimates and Standard Errors from Various Model Specifications

	MATH				READING			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(6)
	Base Model	Selectivity Model	District FE	School FE	Base Model	Selectivity Model	District FE	School FE
Antioch Univ	-0.031**	-0.021*	-0.017	-0.018	0.005	0.014	0.011	0.017
	(0.01)	-(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Central Washington Univ	-0.005	-0.016***	-0.005	0.000	0.005	-0.003	0.005	0.009
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
City Univ	-0.002	0.006	0.003	0.001	0.003	0.010*	0.002	0.002
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)
Eastern Washington Univ	-0.014**	-0.025***	-0.034***	-0.031***	-0.024***	-0.033***	-0.028***	-0.034***
	(0.00)	-(0.03)	(0.00)	(0.00)	(0.00)	-(0.03)	(0.00)	(0.01)
Gonzaga Univ	-0.006	-0.019*	-0.028**	-0.030**	-0.030**	-0.041***	-0.038***	-0.044***
	(0.01)	-(0.03)	(0.01)	(0.01)	(0.01)	-(0.04)	(0.01)	(0.01)
Heritage Univ	-0.003	0.004	-0.008	-0.008	-0.006	0.000	-0.013	-0.005
	(0.01)	-(0.01)	(0.01)	(0.01)	(0.01)	-(0.01)	(0.01)	(0.01)
Northwest Univ	-0.046*	-0.039*	-0.046*	0.001	-0.018	-0.008	-0.006	0.012
	(0.02)	-(0.05)	(0.02)	(0.02)	(0.02)	-(0.01)	(0.02)	(0.02)
Pacific Lutheran Univ	0.009	-0.005	0.010	0.019**	-0.014*	-0.028***	-0.018**	-0.013*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	-(0.02)	(0.01)	(0.01)
St Martin's Univ	-0.025**	-0.037***	-0.011	-0.021*	-0.047***	-0.053***	-0.026**	-0.027**
	(0.01)	-(0.01)	(0.01)	(0.01)	(0.01)	-(0.03)	(0.01)	(0.01)
Seattle Pacific Univ	0.008	-0.006	0.005	0.001	0.010	-0.005	0.009	0.013
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Seattle Univ	-0.024**	-0.036***	-0.033***	-0.027**	-0.008	-0.015	-0.018*	-0.007
	(0.01)	-(0.03)	(0.01)	(0.01)	(0.01)	-(0.02)	(0.01)	(0.01)
Evergreen State College	-0.017	-0.039*	-0.023	-0.024	-0.014	-0.020	-0.020	-0.023
	(0.01)	-(0.02)	(0.01)	(0.01)	(0.02)	-(0.02)	(0.02)	(0.02)
U of Puget Sound	0.014*	0.006	0.014*	0.020**	0.010	0.006	0.008	0.016
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
UW Seattle	0.018***	0.009	0.007	0.006	0.006	-0.002	-0.005	-0.005
	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	-(0.01)	(0.01)	(0.01)
UW Bothell	0.020*	0.029**	0.018*	0.017	0.041***	0.050***	0.024*	0.024*
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
UW Tacoma	0.020*	0.029**	0.012	-0.006	0.039***	0.047***	0.034**	0.009
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)
Walla Walla Univ	0.005	0.005	0.009	-0.054	0.046	0.047	0.060*	0.036
	(0.02)	(0.01)	(0.03)	(0.03)	(0.02)	(0.06)	(0.03)	(0.03)
Washington State Univ	0.003	-0.009*	-0.001	-0.004	0.009*	-0.002	0.003	0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Western Washington Univ	0.007*	-0.004	-0.004	-0.007	0.013**	-0.002	0.002	0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Whitworth Univ	-0.004	-0.019**	-0.018*	-0.021**	-0.019*	-0.035***	-0.019*	-0.020*
	(0.01)	-(0.02)	(0.01)	(0.01)	(0.01)	-(0.02)	(0.01)	(0.01)
Number of Observations	384433	384433	385377	388243	384433	384433	385377	388243
R^2	0.65	0.65	0.65	0.66	0.56	0.57	0.57	0.57

All models also include *student covariates* (i.e., student's previous scores on the math and reading WASLs; gender; race/ethnicity; migrant, homeless, home-based, free and reduced-price lunch, and disability statuses; home language; and whether he or she received gifted, LEP, or special education services), *teacher covariates* (i.e., gender, race/ethnicity, degree level, and experience), and *classroom covariates* (i.e., aggregated student demographic characteristics). Base models also include *school* and *district covariates* (i.e., percentages of students who are female, of various racial/ethnic groups, receiving special education services, and/or eligible for free or reduced-price lunches; percent of teachers with at least a Master's degrees, average teacher experience, average class size, and total enrollment).
Models in columns 2 and 4 include school fixed effects. Models in columns 3 and 6 include district fixed effects.
Statistical significance is denoted with: * if p<0.05, ** if p<0.01 **, and *** if p<0.001.

Figure 2. Program Effects and 95% Confidence Intervals for Base Model

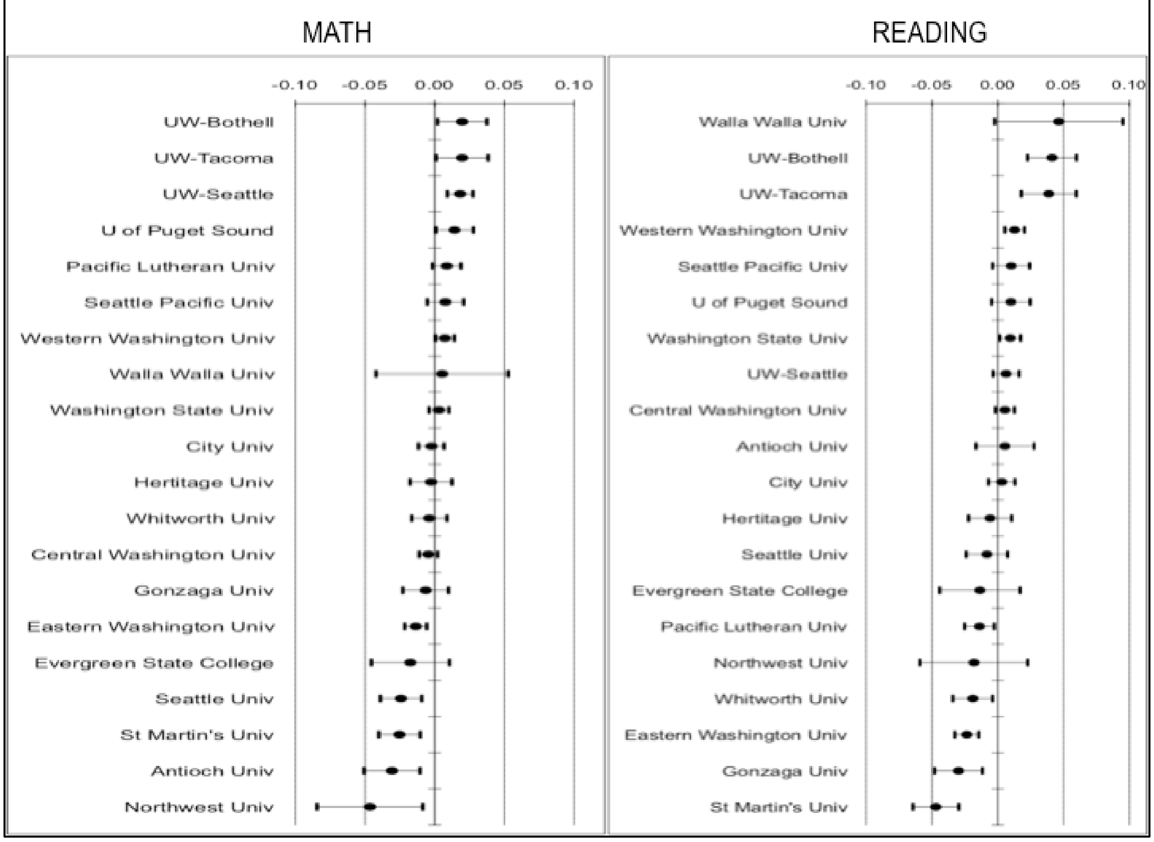


Table 4. Spearman Rank and Pearson Correlations of Program Estimates from Various Model Specifications

		MATH				READING			
	Model Specification	Base Model	Selectivity Model	District FE	School FE	Base Model	Selectivity Model	District FE	School FE
MATH	Base Model	1.00							
	Selectivity Model	0.90/0.87	1.00						
	District FE	0.90/0.87	0.87/0.85	1.00					
	School FE	0.57/0.39	0.53/0.42	0.62/0.48	1.00				
READING	Base Model	0.71/0.62	0.72/0.78	0.75/0.68	0.34/0.16	1.00			
	Selectivity Model	0.52/0.46	0.72/0.75	0.64/0.59	0.35/0.16	0.85/0.95	1.00		
	District FE	0.52/0.43	0.62/0.64	0.68/0.62	0.36/0.02	0.92/0.92	0.88/0.92	1.00	
	School FE	0.36/0.26	0.48/0.51	0.55/0.52	0.43/0.26	0.80/0.83	0.82/0.85	0.93/0.90	1.00

Table 5. Interaction Effects Between Program Indicators and Selected Student Demographics

	MATH					READING				
	Free Lunch	LEP	Black	Asian	Hispanic	Free Lunch	LEP	Black	Asian	Hispanic
Antioch Univ	0.037	-0.023	-0.041	0.003	-0.009	0.008	0.061	-0.024	-0.015	0.032
	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.04)	(0.04)	(0.03)	(0.04)
Central Washington	-0.001	0.017	-0.003	-0.025	-0.002	0.003	0.041**	0.018	-0.018	0.004
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
City Univ	0.000	0.044*	-0.012	-0.019	0.007	0.002	0.034	0.015	-0.028	0.024
	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Eastern Washington	-0.015	0.047**	0.017	-0.006	0.007	-0.016	0.066**	0.031	-0.008	0.027
	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
Gonzaga Univ	0.007	-0.012	-0.034	0.052	-0.052	-0.027	-0.032	0.036	-0.001	0.019
	(0.02)	(0.05)	(0.04)	(0.04)	(0.03)	(0.02)	(0.04)	(0.04)	(0.04)	(0.03)
Heritage Univ	0.066***	0.010	-0.075	-0.029	-0.027	0.007	0.012	-0.002	-0.081	0.015
	(0.02)	(0.02)	(0.04)	(0.05)	(0.02)	(0.02)	(0.02)	(0.05)	(0.04)	(0.02)
Northwest Univ	0.007	0.007	-0.102	0.035	-0.024	0.023	-0.129	0.096	-0.068	-0.008
	(0.05)	(0.09)	(0.07)	(0.06)	(0.06)	(0.05)	(0.09)	(0.07)	(0.06)	(0.07)
Pacific Lutheran Univ	-0.053***	0.006	-0.022	-0.017	0.038*	-0.027*	0.073*	0.027	0.035	0.044*
	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)
St Martin's Univ	-0.023	-0.060	0.027	-0.101**	0.007	-0.009	0.139**	0.025	-0.026	0.022
	(0.02)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)	(0.05)	(0.05)	(0.03)	(0.03)
Seattle Pacific Univ	-0.054***	0.010	-0.019	-0.024	0.032	-0.020	0.092**	0.033	-0.041	0.041
	(0.01)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
Seattle Univ	-0.026	0.033	-0.029	-0.015	0.009	0.022	0.033	-0.031	-0.050*	0.004
	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)
Evergreen State	-0.062*	0.122	-0.055	-0.172***	-0.042	-0.105**	0.154*	-0.019	-0.131**	-0.040
	(0.03)	(0.07)	(0.05)	(0.04)	(0.04)	(0.03)	(0.07)	(0.05)	(0.05)	(0.05)
U of Puget Sound	0.016	-0.003	0.025	-0.011	0.036	0.001	-0.016	0.038	0.011	0.030
	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)	(0.02)
UW Seattle	-0.046***	0.039	-0.028	-0.013	-0.022	-0.023*	0.033	-0.016	-0.017	-0.012
	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
UW Bothell	-0.004	0.007	-0.031	0.024	-0.028	0.028	-0.017	-0.046	-0.032	-0.015
	(0.02)	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)	(0.04)	(0.04)	(0.03)	(0.03)
UW Tacoma	0.019	0.012	-0.001	-0.046	0.050	0.025	0.041	0.056	-0.061	0.040
	(0.02)	(0.04)	(0.02)	(0.03)	(0.03)	(0.02)	(0.05)	(0.03)	(0.03)	(0.04)
Walla Walla Univ	0.105*	-0.119	-0.172	0.049	0.097	0.052	-0.113	0.005	-0.038	0.041
	(0.05)	(0.08)	(0.17)	(0.11)	(0.07)	(0.05)	(0.09)	(0.19)	(0.12)	(0.07)
Washington State	-0.005	0.034*	-0.024	-0.020	0.018	0.000	0.019	-0.012	-0.037*	0.015
	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
Western Washington	-0.024***	0.007	-0.002	-0.004	-0.013	-0.005	0.021	0.019	0.000	-0.011
	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Whitworth	0.005	0.015	-0.029	-0.058	0.000	-0.028	0.042	-0.019	-0.031	0.017
	(0.01)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)
Number of Observations			384433					384433		
R^2			0.65					0.57		

Both math and reading models also include *student covariates* (i.e., student's previous scores on the math and reading WASLs; gender; race/ethnicity; migrant, homeless, home-based, free and reduced-price lunch, and disability statuses; home language; and whether he or she received gifted, LEP, or special education services), *teacher covariates* (i.e., gender, race/ethnicity, degree level, and experience), *classroom covariates* (i.e., aggregated student demographic characteristics), and *school* and *district covariates* (i.e., percentages of students who are female, of various racial/ethnic groups, receiving special education services, and/or eligible for free or reduced-price lunches; percent of teachers with at least a Master's degrees, average teacher experience, average class size, and total enrollment).

Statistical significance is denoted with: * if p<0.05, ** if p<0.01 **, and *** if p<0.001.

Table 6. Main and Interaction Effects Between Program Indicators and Graduation Date

	MATH			READING		
	(1)	(2)	(3)	(4)	(5)	(6)
	Main Effects	2000-2004	2005-2009	Main Effects	2000-2004	2005-2009
Antioch	0.031	-0.088**	-0.057	0.049*	-0.058*	-0.051
	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
Central Washington	-0.005	-0.015*	0.037***	0.003	0.005	0.017
	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
City U	0.003	-0.009	0.003	0.010	-0.020	0.003
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Eastern Washington	-0.014**	-0.004	0.020	-0.024***	0.002	0.021
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Heritage	-0.020	0.015	0.051**	0.001	-0.026	0.011
	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Pacific Lutheran	0.007	0.020	-0.018	-0.009	-0.012	-0.019
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
Seattle U	-0.058***	0.084***	0.097***	-0.033**	0.073***	0.047*
	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
UW Seattle	0.012*	0.044**	0.016	0.009	-0.020	-0.011
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
UW Bothell	-0.031	0.052	0.073*	-0.028	0.075*	0.082*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Washington State	0.008	0.004	-0.028**	0.008	0.004	0.005
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Western Washington	0.003	0.012	0.017	0.011*	0.008	0.006
	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
Whitworth	-0.015	0.045**	0.026	-0.037***	0.052**	0.060**
	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
All Other Programs	0.000	-0.010	0.016	-0.002	-0.005	-0.013
	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
Number of Observations		384433			384433	
R^2		0.65			0.57	

NOTE: Interaction effects are only reported for the twelve institutions with at least 30 graduates from both time periods. All other programs are collapsed and their combined effects are reported together in the "other" category. Coefficients for the 2000-2004 and 2005-2009 interactions should be interpreted relative to the main effect (representing graduates before 2000).

Both math and reading models also include *student covariates* (i.e., student's previous scores on the math and reading WASLs; gender; race/ethnicity; migrant, homeless, home-based, free and reduced-price lunch, and disability statuses; home language; and whether he or she received gifted, LEP, or special education services), *teacher covariates* (i.e., gender, race/ethnicity, degree level, and experience), *classroom covariates* (i.e., aggregated student demographic characteristics), and *school* and *district covariates* (i.e., percentages of students who are female, of various racial/ethnic groups, receiving special education services, and/or eligible for free or reduced-price lunches; percent of teachers with at least a Master's degrees, average teacher experience, average class size, and total enrollment).

Statistical significance is denoted with: * if p<0.05, ** if p<0.01 **, and *** if p<0.001.

Appendix

Table A1. Mean Teacher Characteristics of Teachers from Different Training Institutions

	Out of State	Antioch University	Central Washington University	City University	Eastern Washington University	Evergreen State College	Gonzaga University
Number of Teachers	1885	82	1182	532	751	49	119
Female (%)	81.1	72.0	78.4	68.8	74.2	71.4	84.9
American Indian (%)	0.5	0.0	0.4	0.9	0.3	2.0	0.8
Asian Pacific Islander (%)	3.0	11.0	1.8	3.4	1.5	6.1	3.4
Black (%)	1.3	8.5	0.8	3.9	0.1	0.0	0.0
Hispanic (%)	1.6	1.2	2.5	2.3	2.8	0.0	0.8
White (%)	93.6	79.3	94.5	89.5	95.3	89.8	95.0
Masters Degree or higher (%)	66.7	85.4	55.8	88.7	69.9	100.0	75.6
Average Teaching Experience (yrs)	14.8	5.3	13.9	5.7	14.9	8.0	13.0
1 to 2 Years Teaching Experience (%)	5.8	19.5	9.5	23.7	6.0	18.4	6.7
3 to 5 Years Teaching Experience (%)	12.4	35.4	12.7	32.5	10.9	16.3	17.6
6 to 12 Years Teaching Experience (%)	27.3	40.2	28.4	34.0	27.8	34.7	30.3
13+ Years Teaching Experience (%)	54.5	4.9	49.4	9.8	55.3	30.6	45.4
	Heritage University	Northwest University	Pacific Lutheran University	Saint Martin's University	Seattle Pacific University	Seattle University	University of Puget Sound
Number of Teachers	180	24	346	147	200	163	164
Female (%)	77.8	70.8	81.5	78.9	79.5	75.5	75.0
American Indian (%)	1.1	0.0	0.9	1.4	0.0	1.2	1.2
Asian Pacific Islander (%)	2.2	0.0	2.9	1.4	1.5	6.1	1.8
Black (%)	2.8	0.0	0.3	2.7	1.5	2.5	0.6
Hispanic (%)	23.3	0.0	1.4	1.4	2.0	1.2	3.0
White (%)	70.6	100.0	94.5	93.2	95.0	89.0	93.3
Masters Degree or higher (%)	68.9	54.2	67.6	57.1	55.5	87.1	75.0
Average Teaching Experience (yrs)	6.8	8.6	12.9	10.5	13.9	12.3	17.2
1 to 2 Years Teaching Experience (%)	18.9	25.0	11.0	13.6	11.0	12.9	9.1
3 to 5 Years Teaching Experience (%)	25.6	16.7	13.0	8.8	5.5	18.4	4.9
6 to 12 Years Teaching Experience (%)	39.4	25.0	30.6	40.8	32.0	25.8	18.9
13+ Years Teaching Experience (%)	16.1	33.3	45.4	36.7	51.5	42.9	67.1
	University of Washington Seattle	University of Washington Bothell	University of Washington Tacoma	Walla Walla University	Washington State University	Western Washington University	Whitworth University
Number of Teachers	474	105	88	14	910	1046	235
Female (%)	76.8	81.0	79.5	71.4	81.9	79.5	74.9
American Indian (%)	1.1	1.0	0.0	0.0	0.7	0.9	0.9
Asian Pacific Islander (%)	7.6	2.9	5.7	0.0	1.2	2.0	3.4
Black (%)	1.1	1.0	3.4	0.0	0.9	0.9	0.9
Hispanic (%)	2.7	5.7	1.1	0.0	3.0	0.9	1.3
White (%)	87.6	89.5	89.8	100.0	94.3	95.4	93.6
Masters Degree or higher (%)	64.1	35.2	28.4	85.7	63.4	55.4	73.6
Average Teaching Experience (yrs)	17.6	4.8	5.6	12.4	12.3	13.2	12.2
1 to 2 Years Teaching Experience (%)	6.8	29.5	23.9	7.1	12.6	8.1	11.5
3 to 5 Years Teaching Experience (%)	8.9	28.6	25.0	0.0	18.8	13.6	17.4
6 to 12 Years Teaching Experience (%)	19.4	41.9	48.9	50.0	30.5	35.0	29.8
13+ Years Teaching Experience (%)	65.0	0.0	2.3	42.9	38.0	43.3	41.3

Table A2. Mean Student Characteristics of Teachers from Different Training Institutions

	Out of State	Antioch University	Central Washington University	City University	Eastern Washington University	Evergreen State College	Gonzaga University
Number of Students	82335	3480	52283	22957	34675	1959	5959
Math score (mean)	406.3	401.0	400.7	405.1	404.4	401.6	407.5
Math score (std dev)	41.5	40.3	40.7	41.1	40.8	42.2	41.7
Reading score (mean)	411.2	408.8	408.9	410.5	409.7	409.3	410.5
Reading score (std dev)	23.68	23.51	23.40	23.80	24.66	25.00	24.68
Female (%)	49.2	48.8	48.9	49.0	49.0	49.1	48.1
American Indian (%)	2.3	1.8	2.4	2.2	3.1	3.7	3.1
Asian Pacific Islander (%)	8.8	18.0	7.0	10.5	4.6	11.8	5.3
Black (%)	5.3	12.6	5.2	7.1	3.8	9.9	3.4
Hispanic (%)	13.2	13.9	23.6	11.5	13.8	11.1	9.5
White (%)	67.6	51.1	59.0	65.2	71.8	61.6	74.9
Multiracial (%)	2.5	2.3	2.2	3.1	2.7	1.3	3.6
Free Lunch (%)	35.6	40.8	42.9	33.8	43.3	37.8	38.8
Learning Disabled (%)	7.9	8.7	8.1	8.4	7.8	9.2	8.3
Gifted (%)	4.8	2.2	4.0	4.0	4.3	4.9	4.4
Limited English Proficency (%)	5.2	6.9	7.6	4.5	4.9	4.4	3.5
Special Education (%)	11.7	12.8	11.5	12.3	11.7	13.1	12.8

	Heritage University	Northwest University	Pacific Lutheran University	Saint Martin's University	Seattle Pacific University	Seattle University	University of Puget Sound
Number of Students	7001	966	15944	6599	8971	6938	8231
Math score (mean)	389.7	404.7	404.9	399.4	410.2	410.9	406.4
Math score (std dev)	40.4	42.4	40.2	39.5	42.8	42.5	40.3
Reading score (mean)	403.0	410.5	410.5	407.5	412.9	413.3	411.5
Reading score (std dev)	23.2	23.5	23.3	23.6	23.4	23.2	23.4
Female (%)	49.3	49.3	49.7	49.7	48.9	49.0	49.0
American Indian (%)	4.6	1.6	2.3	3.0	1.6	1.1	2.6
Asian Pacific Islander (%)	2.8	14.7	9.5	6.5	11.6	17.0	10.3
Black (%)	3.1	6.2	8.4	7.2	5.9	7.6	9.7
Hispanic (%)	56.4	13.0	10.8	12.0	13.3	12.5	10.9
White (%)	32.4	60.2	66.5	68.4	64.6	58.8	63.6
Multiracial (%)	0.8	4.1	2.2	2.2	2.6	2.9	2.4
Free Lunch (%)	67.2	31.7	33.6	44.2	31.3	31.0	35.1
Learning Disabled (%)	7.8	8.1	7.0	7.9	8.0	9.0	7.7
Gifted (%)	2.4	6.0	3.8	3.7	6.8	5.6	4.3
Limited English Proficency (%)	17.1	5.4	3.2	3.2	4.7	5.8	3.5
Special Education (%)	9.7	10.8	11.1	12.2	12.0	12.3	11.5

	University of Washington Seattle	University of Washington Bothell	University of Washington Tacoma	Walla Walla University	Washington State University	Western Washington University	Whitworth University
Number of Students	22559	5062	3918	682	41183	49551	10669
Math score (mean)	412.8	410.1	397.7	394.2	404.8	405.6	405.6
Math score (std dev)	43.1	43.5	39.2	40.7	41.2	40.4	40.9
Reading score (mean)	414.1	413.1	407.8	406.5	410.7	411.2	409.7
Reading score (std dev)	23.6	24.4	22.5	23.7	23.9	23.6	25.3
Female (%)	49.4	49.4	47.9	51.4	49.2	49.1	49.4
American Indian (%)	1.7	1.8	1.6	1.9	1.9	3.0	2.9
Asian Pacific Islander (%)	13.9	15.8	12.1	4.4	6.8	7.9	4.3
Black (%)	6.8	5.2	19.3	2.6	3.9	4.2	4.1
Hispanic (%)	12.3	13.5	12.7	25.6	17.6	12.9	9.8
White (%)	62.3	59.6	52.3	65.3	67.1	69.6	75.5
Multiracial (%)	2.8	4.1	1.7	0.1	2.5	2.2	3.1
Free Lunch (%)	29.0	31.1	46.3	51.2	38.7	35.2	42.0
Learning Disabled (%)	8.3	8.8	7.6	9.5	7.7	8.4	8.2
Gifted (%)	5.5	4.9	2.9	3.1	3.8	3.5	4.6
Limited English Proficency (%)	4.6	6.3	4.5	12.6	6.5	4.8	3.8
Special Education (%)	12.3	12.4	10.7	13.5	11.4	12.5	12.4

Table A3. Student-to-Teacher Matching by Subject, Grade, and Year

		MATH					READING				
		Grade 3	Grade 4	Grade 5	Grade 6	All Grades	Grade 3	Grade 4	Grade 5	Grade 6	All Grades
2005-06	Useable matches	65,595	64,275	64,849	48,030	242,749	64,879	64,047	65,379	47,741	242,046
		87.1%	86.2%	84.8%	62.5%	80.1%	86.2%	85.9%	85.5%	62.1%	79.8%
	Unuseable matches	2,174	2,209	1,802	6,363	12,548	2,705	2,270	2,096	6,192	13,263
		2.9%	3.0%	2.4%	8.3%	4.1%	3.6%	3.0%	2.7%	8.1%	4.4%
	Unmatched	7,528	8,104	9,782	22,434	47,848	7,713	8,271	8,958	22,894	47,836
		10.0%	10.9%	12.8%	29.2%	15.8%	10.2%	11.1%	11.7%	29.8%	15.8%
	No. of Students	75,297	74,588	76,433	76,827	303,145	75,297	74,588	76,433	76,827	303,145
2006-07	Useable matches	66,244	65,465	63,881	47,918	243,508	65,764	65,365	64,159	47,750	243,038
		87.1%	86.3%	84.8%	62.5%	80.1%	86.5%	86.2%	85.2%	62.2%	80.0%
	Unuseable matches	2,399	2,182	2,190	5,235	12,006	2,758	2,407	2,474	5,486	13,125
		3.2%	2.9%	2.9%	6.8%	4.0%	3.6%	3.2%	3.3%	7.2%	4.3%
	Unmatched	7,415	8,169	9,224	23,555	48,363	7,536	8,044	8,662	23,472	47,714
		9.7%	10.8%	12.3%	30.7%	15.9%	9.9%	10.6%	11.5%	30.6%	15.7%
	No. of Students	76,058	75,816	75,295	76,708	303,877	76,058	75,816	75,295	76,708	303,877
2007-08	Useable matches	65,226	64,648	63,829	44,156	237,859	65,152	64,940	63,941	43,878	237,911
		84.1%	83.5%	82.7%	57.8%	77.1%	84.0%	83.9%	82.9%	57.4%	77.1%
	Unuseable matches	2,374	2,536	2,502	5,002	12,414	2,460	2,499	2,429	5,197	12,585
		3.1%	3.3%	3.2%	6.5%	4.0%	3.2%	3.2%	3.1%	6.8%	4.1%
	Unmatched	9,990	10,221	10,806	27,275	58,292	9,978	9,966	10,767	27,358	58,069
		12.9%	13.2%	14.0%	35.7%	18.9%	12.9%	12.9%	14.0%	35.8%	18.8%
	No. of Students	77,590	77,405	77,137	76,433	308,565	77,590	77,405	77,137	76,433	308,565
2008-09	Useable matches	33,958	33,943	33,607	22,457	123,965	33,937	33,978	33,584	22,071	123,570
		43.3%	43.2%	42.9%	28.9%	39.6%	43.2%	43.2%	42.9%	28.4%	39.5%
	Unuseable matches	1,752	1,589	1,618	1,212	6,171	1,723	1,601	1,592	1,433	6,349
		2.2%	2.0%	2.1%	1.6%	2.0%	2.2%	2.0%	2.0%	1.8%	2.0%
	Unmatched	42,780	43,054	43,043	54,011	182,888	42,830	43,007	43,092	54,176	183,105
		54.5%	54.8%	55.0%	69.5%	58.4%	54.6%	54.7%	55.1%	69.7%	58.5%
	No. of Students	78,490	78,586	78,268	77,680	313,024	78,490	78,586	78,268	77,680	313,024
	Total Students	**307,435**	**306,395**	**307,133**	**307,648**	**1,228,611**	**307,435**	**306,395**	**307,133**	**307,648**	**1,228,611**

Note: Unuseable matches are for proctor names that matched more than one teacher name in a given school.

Table A4. Number of Teachers with WEST-B Scores by Initial Certificate Granting Program

Program Name	Count	Percent
Out-of-state	236	16.04
Antioch University	30	2.04
Central Washington University	156	10.61
City University	238	16.18
Eastern Washington University	62	4.21
Evergreen State College	10	0.68
Gonzaga University	12	0.82
Heritage University	59	4.01
Northwest University	9	0.61
Pacific Lutheran University	55	3.74
Seattle Pacific University	24	1.63
Seattle University	35	2.38
St. Martin's University	25	1.7
University of Puget Sound	22	1.5
UW Bothell	50	3.4
UW Seattle	47	3.2
UW Tacoma	36	2.45
Walla Walla University	1	0.07
Washington State University	192	13.05
Western Washington University	129	8.77
Whitworth University	43	2.92
Total	1471	100