CEDR Working Paper | WP #2016-4

CEDR
Center for Education Data & Research

What's in a Teacher Test? Assessing the Relationship Between Teacher Test Scores and Student Secondary STEM Achievement.

Dan Goldhaber, American Institutes for Research

Trevor Gratz, University of Washington Bothell

Roddy Theobald, American Institutes for Research

The Center for Education Data & Research
University of Washington Bothell
3876 Bridge Way N. Ste. 201 Seattle, WA 98103
(206) 547-5585 cedr@u.washington.edu
www.cedr.us

Abstract: We investigate the predictive validity of teacher credential test scores for student performance in secondary STEM classrooms in Washington state. After replicating earlier findings that teacher basic skills licensure test scores are a modest and statistically significant predictor of student math test score gains in elementary grades, we focus on three subject/grade combinations—middle school math, ninth-grade algebra and geometry, and ninth-grade biology—in which both current and prior year subject-area test scores are available and estimate value-added models that provide within-subject estimates of the relationship between teacher licensure test scores and student achievement gains. We find that basic skills tests are modestly predictive of student achievement in middle and high school math and highly predictive of student achievement in high school biology. On the other hand, subject-specific tests are a statistically significant predictor of student achievement only in high school biology.

Acknowledgements: This work was produced under financial support from the National Center for Analysis of Longitudinal Data in Education Research (CALDER) funded through grant #R305C120008 to American Institutes for Research from the Institute of Education Sciences, U.S. Department of Education and the National Science Foundation (grant #1555678). We wish to thank Gerhard Ottehenning, Vanessa Quince, and Melanie Rucinski for research assistance, as well as James Cowan and participants at the 2016 NSF STEM PI Convening and 2016 AERA Conference for helpful comments.

Suggested citation:
Goldhaber, D., Gratz, T. and Theobald, R. (2016). What's in a Teacher Test? Assessing the Relationship Between Teacher Test Scores and Student Secondary STEM Achievement. CEDR Working Paper 2016-4. University of Washington, Seattle, WA.

An educated, innovative, motivated workforce—human capital—is the most precious resource of any country in this new, flat world. Yet there is widespread concern about our K–12 science and mathematics education system, the foundation of that human capital in today's global economy (National Academies of Sciences, 2007).

Introduction

There is significant policy focus on the human capital of the nation's STEM teachers, motivated by concern over the need to improve STEM outcomes for students in K–12 schools and college (e.g., President's Council of Advisors on Science and Technology, 2010) and the vast body of empirical evidence showing the importance of teacher quality for student achievement (Aaronson et al., 2007; Goldhaber & Hansen, 2013; Rivkin et al., 2005).[1] One way that states try to ensure a high-quality teacher workforce is by requiring teacher candidates to pass licensure tests[2], often of both their basic skills and content knowledge, as a requirement for receiving a teaching license. Although several studies (e.g., Clotfelter et al., 2007; Goldhaber & Hansen, 2013; Goldhaber, 2007) find modest positive correlations between teacher performance on licensure exams and student math achievement gains in elementary grades, there is little evidence on whether licensure tests provide a useful "signal" of the future quality of secondary STEM teachers.

In this paper we use data from Washington state to investigate whether STEM teacher candidates who score better on licensure tests are also more effective at improving student performance once they enter the teaching workforce. We focus on three subject/grade combinations—middle school (seventh–eighth grade) math, ninth-grade algebra and geometry, and ninth-grade

[1] This focus on the human capital of STEM teachers is not new. In fact, there exists an extensive body of literature tracking the progress that the nation is (or is not) making toward having a high-capacity STEM teacher workforce. Unfortunately, the indicators often used to evaluate this progress—e.g., teacher credentials and degree type—have not been found to be highly predictive of student achievement (e.g., Wilson et al., 2001).
[2] We use the terms "licensure test" and "credential test" interchangeably.

biology—in which both current and prior year subject-area test scores are available, and estimate value-added models that provide within-subject estimates of the relationship between teacher licensure test scores and student achievement gains. This is the first paper to use traditional value-added methods to investigate the predictive validity of teacher licensure test scores in secondary math classrooms and the first to consider the predictive validity of teacher licensure test scores in any science classrooms.

We find that *basic skills* credential test scores are modestly predictive of student achievement in middle and high school math (though only statistically significant in middle school math) and highly predictive of student achievement in high school biology. The relationships between teacher candidate performance on *subject-specific* credential test scores and student performance are similar in magnitude to the relationships for basic skills tests, though statistically significant only in high school biology. The relationships that exist are most pronounced when teachers who score in the top quartile of these tests are compared to teachers who scored in the bottom quartile, and there is some evidence that the relationships between a teacher's subject-specific credential test scores and student performance are more pronounced for students in either advanced or remedial courses.

The paper proceeds as follows. In section II, we provide background and context for this study. We introduce our data and discuss summary statistics in section III, outline our analytic models in section IV, and describe our results in section V. We then offer some concluding thoughts in section VI.

Background

Teacher quality has long and repeatedly been shown to be one of the most important school-related influences on student achievement (Aaronson et al., 2007; Coleman et al., 1966; Rivkin et al., 2005; Rockoff, 2004), yet many proposed indicators of STEM teacher content knowledge have only face validity at the secondary level (Wilson et al., 2001). A number of studies have investigated the "predictive validity" of various preservice indicators of teacher content area preparation (i.e., the extent to which these indicators are predictive of student performance in teachers' classrooms). However, evidence about the predictive validity of commonly used indicators such as teacher degree type and degree level is mixed.[3]

Although a teacher's mathematical content knowledge has been shown to be predictive of student learning gains at the elementary level (Hill et al., 2005), evidence relating the factors that may influence a teacher's mathematical content knowledge—such as the number and type of courses prospective teachers take in college—to student achievement at any level is more mixed. Monk and King (1994) found that the number of undergraduate mathematics and physical science courses a teacher takes is positively related with how well students perform on math and science tests, respectively. Similarly, Boyd et al. (2009) found that first-year elementary teachers from teacher education programs that require mathematics courses are more effective in teaching math. However, when Harris and Sass (2011) included measures of pre-college ability to account for sorting by college major, they found no significant relationship between the number of courses taken in different areas and student achievement in secondary math courses. The same is true at the elementary level, with the exception that the number of math credits is negatively correlated with student math achievement.

Another potential preservice measure of teacher content knowledge is performance on licensure tests. Indeed, all but one state require teachers to pass various licensure tests to participate in the public school labor market. Licensure tests have a long history, dating back to the 1930s when the first national licensure exam, the National Teacher Examination, was developed (Ravitch, 2003).[4] Throughout their history, teacher licensure exams have been viewed primarily as an important quality

[3] For example, some studies find no relationship between generic teacher degree type (e.g., masters vs. bachelor's) and student achievement in mathematics (Monk & King, 1994; Aaronson et al., 2007), while others find that a bachelors or masters degree *in mathematics* is positively correlated with student achievement when teachers with this degree are teaching a mathematics course (Goldhaber & Brewer, 1997, 2000).
[4] This test was replaced in the 1990s by the Praxis exam series (ETS, 2016).

screen needed to professionalize teaching; advocates for national licensure exams often compared them to tests taken by lawyers and doctors before they are certified to practice (Maero, 1985). In recent years, reformers have also pushed for a more rigorous licensure exam incorporating not only written tests of content knowledge and pedagogy, but also a live teaching component, with the hope of raising standards for entry into the teaching profession (Baker, 2012). But licensure tests also have a disparate racial/ethnic impact on eligibility to teach so negatively impact efforts to diversify the teacher workforce (Goldhaber & Hansen, 2010). Hence, public debates about teacher licensure often center on the extent to which traditional licensure exams are a useful signal as opposed to an inefficient barrier to teachers with a less traditional background (e.g., Barmore, 2016).

Despite their widespread use in teacher licensing (as a “pass/fail” signal), teacher licensure test scores are typically not used for any additional personnel decisions (e.g., hiring or professional development). Indeed, test developers actively discourage the use of licensure tests for decisions other than licensure itself, despite the fact that test scores may be predictive of teacher quality away from the high-stakes cut-point used to determine employment eligibility.[5] In fact, empirical evidence at the elementary level shows positive and significant relationships between teachers' performance on some licensure exams and student test scores throughout the teacher test score distribution (Clotfelter et al., 2007; Goldhaber & Hansen, 2013; Goldhaber, 2007; Hendricks, 2014). Goldhaber (2007), for instance, analyzes data from North Carolina and finds that having a teacher who passed the Praxis II tests rather than one who failed is correlated with an increase in a student's mathematics achievement of about 6% of a standard deviation, and that a one standard deviation increase in a teacher’s test score is predictive of an increase in student mathematics achievement of about 3% of a standard deviation. Most recently,

---

[5] The test developer (Pearson) for the WEST-B (a basic skills test used in Washington state), for instance, states: “The subtest scores indicated on this report are only for the purposes of admission to state-approved teacher preparation programs and for teacher certification. They are NOT intended to be used for employment decisions, other college admissions decisions, or any other purpose.” http://www.west.nesinc.com/Content/Docs/WESTB_ScoreReport_backer.pdf

Hendricks (2014) documented increases in student achievement associated with the movement of a teacher with a high licensure score into the student's grade and school.

Although most existing evidence is focused at the elementary level, there are both theoretical and empirical reasons to believe that the relationship between teacher licensure test scores and student achievement might be stronger at the secondary level than elementary level. Theoretically, the relative importance of teachers' content knowledge may rise as teachers are expected to teach increasingly complex material in higher grades (Appleton, 2013). And empirically, Clotfelter et al. (2010) provide evidence of a relatively strong relationship between teacher licensure test scores and student achievement in high school. To our knowledge, this is the only existing evidence about the predictive validity of teacher licensure test scores at the secondary level, but (due to data limitations) it is based on a very different methodology than prior work at the elementary level. Specifically, Clotfelter et al. (2010) estimate a student fixed-effects model that relies on *cross-subject comparisons* (e.g., they find that students in high school math classrooms score higher on math tests relative to tests in other subjects when they have a math teacher who has high credential test scores relative to their teachers in other subjects). In the next section, we describe the data that will allow us to build on this existing work and estimate models that rely on *within-subject comparisons* (e.g., do students in secondary math classrooms score higher on math tests, all else equal, when they have a math teacher who has high credential test scores than a math teacher with lower credential test scores?).

Data and Summary Statistics

Data

This study combines four databases, all maintained and supplied by the Washington State Office of the Superintendent of Public Instruction (OSPI), to construct one panel data set containing student-teacher-classroom-year observations. These databases are the Washington State Credentials Database, the Washington State S-275 personnel report, the Comprehensive Education Data and Research System (CEDARS)[6], and the State Testing database.

The Washington State Credentials Database contains a complete history of scores on the state's teacher credential tests. In this study, we focus on two tests that have been required for teacher licensing in Washington state in recent years. Since 2002, prospective teachers in Washington state have had to pass the Washington Educator Skills Test-Basic (WEST-B)—an assessment of basic skills in reading, writing, and mathematics—as a requirement for admission into teacher education programs. The test is designed to reflect knowledge and skills described in textbooks, the Washington Essential Academic Learning Requirements, curriculum guides, and licensure standards. Because Washington state accepts a number of alternative tests that meet the WEST-B testing requirement for receiving a teaching credential,[7] only 82% of new teachers from 2006 through 2015 have taken the WEST-B. For these individuals, we observe their scores on the math, reading, and writing subtests for each time they took the test.

From 2010 to 2014, all teacher education program graduates also had to pass the Washington Educator Skills Test-Endorsements (WEST-E), a subject knowledge test for individual teaching endorsements, as a requirement for receiving a teaching credential.[8] Different WEST-E exams were required for teachers to become certified in different subject areas and grade levels, but every credentialed teacher had to pass at least one of these tests as a requirement for licensure. For this

[6] We also use the precursor to CEDARS, the Core Student Records System (CSRS), in our replication study.
[7] Passing scores for Praxis I, California Basic Educational Skills Test (CBEST), or the Pearson NES Essential Academic Skills test, as well as scores on the SAT and ACT above certain cutoffs (e.g., 515 on the math SAT) can be submitted as alternatives to the WEST-B exam (RCW 28A.410.220 & WAC 181-01-002).
[8] Prior to the WEST-E, the state required a passing score on the Praxis-II tests. Beginning in September 2014, the state replaced some WEST-E tests with assessments from the National Evaluation Series (NES). For parsimony, we only consider WEST-E scores in this paper.

study, we focus on scores on four WEST-E tests observed most frequently for teachers in our sample: Mathematics, Middle Level Mathematics (MLM), Science, and Biology.[9]

The credential exam data set is linkable to the state's S-275 database, which 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, annual compensation, and demographic characteristics. The data set also includes highest degree earned and experience, which we consider as other potential predictors of teacher effectiveness.

Since the 2009–10 school year, teachers can be linked to the students in their classrooms using a unique classroom ID in the state's CEDARS database.[10] For the 2009–10 through 2014–15 school years, the CEDARS database contains information on individual student background variables including gender, race/ethnicity, and free or reduced-priced lunch eligibility, as well as participation in the following programs: gifted/highly capable; limited English proficiency (LEP); and special education. These student-level variables are used as control variables in all our models. From this data set, we are also able to create an indicator for whether each course was designated as an "advanced/honors" course at the school, which allows us to control for student tracking into advanced courses.

Student test score data—the outcome in our analysis—comes from the State Testing database. The database contains annual student test scores on the Measures of Student Progress (MSP) exams for 2009–10 through 2013–14 in reading (Grades 3–8), math (Grades 3–8), and science (Grades 5 and 8), as well as high school End-of-Course (EOC) exams in Algebra, Geometry, and Biology.[11] For 2014–15 the

[9] We also consider the two Elementary WEST-E subtests in our replication study.
[10] CEDARS data includes fields designed to link students to their individual teachers, based on reported schedules. However, limitations of reporting standards and practices across the state may result in ambiguities or inaccuracies around these links. Our replication study also uses proctor of the state assessment as the teacher–student link from the CSRS data system. The "proctor" variable was not intended to be a link between students and their classroom teachers; so this link may not accurately identify those classroom teachers.
[11] Approximately one-third of Washington state schools serving Grades 3–8 participated in a pilot of the SBA in the 2013–2014 school year, and the state did not collect student test scores from these schools. Students from these

state transitioned to the Smarter Balance Assessment (SBA) for Grades 3–8 in both math and reading. As discussed in the introduction, our primary analysis focuses on seventh-grade math, eighth-grade math, ninth-grade algebra and geometry, and ninth-grade biology, all grades in which both current and same-subject prior-year test scores are available. The range of years we can consider varies across these different subject/year combinations. Because third–eighth grade math test scores are available for the entire range of years that students may be linked to teachers, 2009–10 through 2014–15, and scores from the predecessor to the MSP exam—the Washington Assessment of Student Learning (WASL)—are also available for the 2008–09 academic year (i.e., a prior-year math score for 2009–10), we can estimate models for middle school math in all years of available CEDARS data (2009–10 through 2014–15). On the other hand, the Algebra and Geometry EOC exams were introduced in the 2010–2011 academic year, and the Biology EOC exam started in the 2011–12 school year. Thus we can only estimate models for ninth-grade algebra and geometry for 2010–11 through 2014–15, and for ninth-grade biology for 2011–12 and 2014–15.

We make a number of additional restrictions to our final analytic data set. Specifically, we only include student/teacher/year combinations in which the student has valid current and prior-year test scores, received instruction from a single teacher in that subject and year, and (in the case of ninth-graders) was enrolled in the appropriate course for the EOC test. Likewise, for each combination of grade level and teacher credential test, we only consider student/teacher/year combinations in which the teacher has at least one valid credential test score. This results in eight different analytic samples for our primary analysis, which we discuss in the next subsection.

---

schools therefore are not included in the 2013–14 data (because they are missing current-year test scores) or the 2014–15 data (because they are missing prior-year test scores).

Summary Statistics

The analytic samples considered in this paper—consisting of students in tested secondary STEM grades and subjects between 2009–10 and 2014–15 whose teacher had a valid credential test score—vary considerably both in terms of the number and characteristics of the students and teachers. **Table 1** presents student/year-level summary statistics for each of the eight analytic samples for our primary analysis. The first column of Table 1, for example, provides summary statistics for all seventh and eighth-grade students in the analytic data set whose math teacher has at least one valid WEST-B Math score. We standardize all student test scores within grade and year, so the means in column 1 of Table 1 for "Lagged Math" and "Lagged Reading" mean that students in this sample scored about 10% of a standard deviation higher on last year's tests than the average student in the same grade and year. The other summary statistics in column 1 are broadly representative of the demographics of public school students in Washington state, about 50% of whom are eligible for free/reduced priced lunch and about 25% of whom are underrepresented minorities (American Indian, Black, or Hispanic).

The differences in summary statistics between columns 2 and 3 of Table 1 (and between column 6 and columns 7 and 8) highlight an important difference between the ninth-grade algebra/geometry sample and the ninth-grade biology sample. Specifically, students in ninth grade who were enrolled in biology and took the biology EOC test tend to be considerably more advantaged and higher performing than ninth-graders who were enrolled in algebra or geometry and took the algebra or geometry EOC tests. This is likely because low-performing students often wait to take biology (and the biology EOC) until 10th grade. An important caveat to the high school portion of our analysis, then, is that it is only generalizable to the population of students who take these courses in ninth grade.

We now turn to teacher/year-level summary statistics, reported in **Table 2**. Teachers across all samples are, on average, considerably less experienced than the average Washington state teacher because, as discussed in the previous subsection, the credential tests considered in this analysis have only been required since 2002 (for the WEST-B) or 2010 (for the WEST-E). This also explains why teachers in the WEST-E samples tend to be less experienced than teachers in the WEST-B samples. In Table 2 (and in the analytic models described in the next section), credential test scores come from the *first time* each teacher took the test and are standardized across all teacher candidates who have ever taken these tests. For example, the mean for "WEST-B Math" in column 1 of Table 2 means that the average teacher in the WEST-B Math middle school sample scored over 50% of a standard deviation higher on their first WEST-B Math test than the average teacher candidate who took this test.

Our decision to standardize credential test scores across all years of data is important because, as shown in **Figure 1,** average scores on all three WEST-B tests have been increasing steadily over time. These trends could be explained by the increased availability of test preparation materials, a drop in test difficulty, or an increase in the average qualifications of teachers. The first two explanations would suggest that we should only standardize teacher test scores within years (since the time trends would have nothing to do with the qualifications of different cohorts of teacher candidates), while the latter explanation would suggest that we should standardize teacher test scores across years (as the time trends would reflect differences in average qualifications across test cohorts).

We test these explanations directly by estimating predictive validity models (described in the next section) with and without test-year (or "cohort") fixed effects. The year in which candidates take the WEST-B is highly predictive of the performance of their students (F = 30.84), and there is no evidence that the within-cohort relationship between WEST-B scores is any different than the cross-cohort relationship (t = -1.23). This suggests that changes in average WEST-B scores over time do reflect true differences in teacher candidate quality. This is consistent with evidence from other studies showing that average SAT scores of prospective teachers have increased over the past two decades

(Goldhaber & Walch, 2014; Lankford et al., 2014),[12] recent cohorts of prospective teachers have higher undergraduate GPAs than their predecessors (Gitomer, 2007), and new teachers are now coming from more competitive undergraduate institutions than in past years (Lankford et al., 2014). Finally, the developer of the WEST-B and WEST-E (Pearson) describes the tests as "criterion-referenced," meaning that they are "designed to measure a candidate's knowledge and skills in relation to an established standard (a criterion), rather than in relation to the performance of other candidates."[13] For these reasons, we standardize credential test scores across all years in our primary analysis.[14]

Means of the teacher credential test scores in Table 2 permit some comparisons across different kinds of teachers, but to dig into these differences further, **Figure 2** displays kernel density plots of WEST-B scores (on the original scoring scale) for six mutually exclusive groups of test takers. The first four groups are considered in this study: elementary teachers (used in the replication study), middle school math teachers, ninth-grade algebra and geometry teachers, and ninth-grade biology teachers.[15] For comparison, we also include all other teachers (i.e., those who are in the workforce but not in one of our analytic samples), and all test takers who never become teachers. Figure 2 shows that ninth-grade teachers tend to score higher on all three WEST-B tests than middle school math teachers, and both groups of teachers tend to score dramatically higher on the WEST-B Math test than elementary teachers, other teachers, and test takers who never enter the teaching workforce.[16]

---

[12] The increase in SAT scores documented in Lankford et al. (2014) is 0.10 standard deviations from 2002 to 2010, which is not as dramatic as the 0.19 standard deviation increase in WEST-B scores over the same time period.
[13] https://www.west.nesinc.com/PageView.aspx?f=GEN_AboutTheTests.html
[14] We also experiment with models that consider test scores standardized within year, and the results are qualitatively similar (results available from authors upon request).
[15] For the purposes of this figure, teacher type was determined by the number of students in each subject–grade combination taught in the analytic sample or elementary sample.
[16] These figures show scores from the first time each individual took each test, but teacher candidates can take these tests as many times as necessary to receive a passing score on each subtest; for example, 79.4% of individuals who fail the WEST-B Math test the first time pass eventually.

Figure 3 shows similar kernel density plots for WEST-E tests. The first two panels of Figure 3 show that ninth-grade algebra and geometry teachers tend to score considerably higher than middle school math teachers on both WEST-E Math tests, though both groups perform better, on average, than test takers who never enter the workforce.[17] For the other WEST-E tests, teachers in our sample do not perform much better, on average, than other teachers or test takers who never enter the workforce.

The "Cut Score" line in each plot within Figures 2 and 3 illustrates that, while the passing score is nominally set to the same scale score (240) for all tests, some of these credential tests are much more difficult to pass than others. **Figures 4 and 5** show overall passing rates for these tests across all teacher candidates in Washington state and compares these passing rates to those in other states (California, Florida, and Michigan) that report these numbers. Generally speaking, the passing rates on the WEST-B tests are much higher than the passing rates for basic skills credential tests in these other states, while the passing rates on the WEST-E tests considered in our primary analysis are more in line with (and even lower than in some cases) the passing rates for subject-specific credential tests in these other states.

We can also directly compare the difficulty of different WEST-E tests by comparing the WEST-E performance of candidates who took different WEST-E tests but had similar scores on the WEST-B. We find that candidates tend to perform 16–20 points (or about one standard deviation) higher on the Elementary Education WEST-E tests than candidates with similar WEST-B scores perform on the Middle Level Math, Science, or Biology WEST-E exam, and 40 points (or about two standard deviations) higher than candidates with similar WEST-B scores perform on the Mathematics WEST-E test. These differences in test difficulty have important policy implications that we discuss in the conclusion.

[17] 39.6% of teacher candidates who fail the WEST-E Math on their first test administration eventually pass it, while another 31.8% eventually pass the WEST-E MLM test.

Analytic Approach

Our primary analytic approach can be situated within a larger literature that uses value-added models (VAMs) to separate the impact of various interventions (including teacher characteristics) from other variables that influence student test performance.[18] Following the existing literature about the predictive validity of teacher licensure tests at the elementary level (e.g., Clotfelter et al., 2007; Goldhaber & Hansen, 2013; Goldhaber, 2007), we estimate variants of the following VAM:

$$Y_{ijgst} = \beta_0 + \beta_1 Y'_{i,g-1,t-1} + \beta_2 X_{igt} + \beta_3 Z_{jt} + \beta_4 Score_j + \varepsilon_{ijgst} \qquad (1)$$

In equation (1), Y_{ijgst} is the test score (MSP, SBA, or EOC) of student *i* in grade *g*, subject *s*, and year *t*, while in teacher *j*'s classroom. $Y'_{i,g-1,t-1}$ is a vector of student *i*'s prior test scores in reading, mathematics, and (for ninth-graders) science. The student test scores in both Y_{ijgst} and $Y'_{i,g-1,t-1}$ are standardized by test, grade, and year across all test takers. Therefore, the units of the coefficients on the right hand side of equation 1 are standard deviations of student performance (relative to other scores on the same test in the same grade and year). X_{igt} is a vector of student covariates for student *i,* in grade *g,* and year *t*, which includes indicators for race/ethnicity, gender, free or reduced-priced lunch eligibility, gifted/highly capable, limited English proficiency (LEP), special education, and learning disabled (see Table 1). Z_{jt} is a vector of teacher covariates in year *t* that control for other potential measures of teacher effectiveness, and that (if not included) may confound the effect of a teacher's licensure test score. This vector includes indicators for teacher experience level in year *t* and an indicator for whether or not the teacher possesses an advanced degree in year *t* (see Table 2).

[18] In the case of individual teacher evaluation, estimates from VAMs have been shown to be unbiased despite the presence of student sorting (Chetty et al. 2014a; Kane & Staiger, 2008), and a recent review of the literature surrounding value-added methodologies concluded, "To date, the studies that have used the strongest research designs provide compelling evidence that estimates of teacher value-added from standard models are not meaningfully biased by student-teacher sorting along observed or unobserved dimensions" and that "there is not any direct counter evidence indicating that value-added estimates are substantially biased" (Koedel et al., 2015).

Nearly all of the differences between the models estimated in the next section relate to the specification of teacher licensure test scores, $Score_j$, and the presence or absence of various fixed effects. In our first specification of the model in equation 1, $Score_j$ is the credential test score of teacher *j* standardized across all years of test takers. The coefficient β_4 in these specifications can be interpreted as the extent to which continuous credential test scores provide a "signal" of future teacher effectiveness (i.e., the expected increase in student performance associated with a one standard deviation increase in the credential test score of teacher *j*). For some credential tests (mostly WEST-E tests), enough candidates failed the test on their first attempt to identify an additional specification of the model in equation 1 in which $Score_j$ is an indicator of whether teacher *j* passed the credential test on the first attempt. In these specifications, the coefficient β_4 can be interpreted as the extent to which there is predictive validity around the established "cut point" of the test (i.e., the expected increase in student performance associated with having a teacher who passed the test on the first attempt relative to having a teacher who failed). Finally, we can mitigate concerns about nonlinearities and ceiling effects in test scores (see Figure 2) by replacing $Score_j$ with a vector of indicators for the quartile of the distribution of test scores *for teachers in that sample* (Q2, Q3, or Q4, with the reference category being Q1) that the test score of teacher *j* falls into.[19] In these specifications, β_4 is actually a vector of coefficients, each of which represents the expected increase in a student's test score associated with having a teacher with a test score in the second, third, or fourth quartile (respectively), relative to having a teacher with a test score in the lowest quartile.[20]

To provide a complete picture of the relationships between teacher credential test scores and student performance, we estimate each of these specifications three times—once with no fixed effects

[19] We calculate quartiles within each sample because very few teachers in the analytic sample scored in the bottom quartile of the overall distribution of WEST-B Math scores.
[20] As a further check for nonlinearities, we also estimate models that replace the credential scores with a teacher fixed effect and plot the resulting value-added estimates against teacher credential scores.

(so teachers are compared to all other teachers in the sample), once with school fixed effects (so teachers are compared to other teachers in the sample in the same school), and once with school-by-year fixed effects (so teachers are compared to other teachers in the same school and year).[21] We estimate equation (1) by ordinary least squares (OLS) and cluster the error terms ε_{ijgst} at the teacher level to account for correlation between the errors of students taught by the same teacher. Finally, to explore whether teacher credential test scores appear to have a differential impact for different student subgroups, some models include terms that interact student characteristics (e.g., prior performance or participation in an advanced class) with the licensure exam score.

We conclude this section by discussing four potential sources of bias in the estimates from these models. First, a nonrandom subset of teacher candidates in Washington state has been required to take the WEST-B since its introduction in 2002, which could lead to bias if the relationship between credential test scores and effectiveness for the group of test takers is different than it *would have been* for non-test takers. Second, teacher candidates who take these tests are nonrandomly selected into the public teaching workforce, which could lead to "sample selection bias" if teacher candidates with a given credential score who enter the workforce are not representative of all teacher candidate with that score. Third, teacher candidates who enter the teaching workforce are nonrandomly sorted into different teaching positions, which could lead to bias if there are unobserved variables *unrelated to the teacher* correlated both with a teacher's credential score and the performance of the teacher's students. And fourth, teachers nonrandomly leave the teaching workforce, which could lead to "attrition bias" if teacher candidates with a given credential score who remain in the workforce are not representative of all teachers with that score.

[21] We also experiment with student fixed effects models in the middle school math sample, but due to sample size, limitations are not reported.

Although some researchers (e.g., Rothstein, 2010, 2014) have raised concerns about the third potential source of bias, our reading of the broader literature on value-added teacher effects (e.g., Bacher-Hicks et al., 2014; Chetty et al., 2014a; Jackson, 2014; Kane & Staiger, 2008; Kane et al., 2013; Koedel et al., 2015) suggests that the VAMs described above are sufficient to account for nonrandom sorting of teachers to classrooms and schools. That said, we describe some extensions in the next section that investigate the extent of this potential source of bias. On the other hand, we have no way to account for the first potential source of bias, so all results reported in this paper are only generalizable to the population of candidates required to take these licensure tests. As a check on the fourth source of bias, we estimate models predicting teacher attrition as a function of experience, degree level, prior estimated effectiveness, WEST-B scores, and an interaction between prior effectiveness and WEST-B scores. We do not find evidence that teachers with different WEST-B scores are any more or less likely to leave the workforce as a function of their prior estimated effectiveness.

However, an additional threat to validity that we cannot address directly is the second potential source of bias (sample selection bias). Unlike in some prior work (e.g., Goldhaber et al., 2014, 2016b), we do not have access to a convincing instrumental variable that is predictive of workforce entry for teacher candidates and that can be used to estimate a "Heckit model" (Heckman, 1979) that accounts for sample selection bias. Moreover, particularly in tests with lower passing rates (such as the WEST-E tests shown in Figure 3), it is quite plausible that teacher candidates who fail the test the first time are more likely to re-take the test and ultimately enter the workforce if they have a greater commitment to teaching. If these individuals become more effective teachers than teacher candidates with similar WEST-E scores but who did not enter the workforce *would have been* had they entered the workforce, this would cause a downward bias in the estimated relationship between WEST-E scores and student performance (i.e., the estimates reported in this paper represent lower bounds for the true relationship). We are less concerned about sample selection bias in the WEST-B results because so few teachers in the sample failed any of these tests.

Results

Before describing the primary results relating teacher licensure test scores to student achievement in secondary STEM subjects, we first provide some context for these findings. First, to relate this study to earlier work in elementary grades (e.g., Clotfelter et al., 2007; Goldhaber, 2007), we present estimates of the relationship between teacher credential test scores and student math performance in fourth and fifth grade in **Table 3**.[22] Panel A considers teacher WEST-B Math scores, and demonstrates that—similar to what has been reported in prior research at the elementary level—a one standard deviation increase in a teacher's WEST-B Math score is correlated with a .024 to .034 standard deviation increase in student math performance. The quartile results are shown graphically in the top-left panel of **Figure 6**. **Figure 7** simply plots estimated teacher value-added estimates and the teacher's WEST-B Math score and illustrates that, in elementary grades (top-left panel), the relationship looks relatively linear. Interestingly, when we consider the two WEST-E tests required for an elementary education credential in Panels B and C of Table 3, we see little evidence that either of these credential scores is predictive of student math performance (this is reflected in the first two panels of **Figures 8 and 9**).

As additional context for the magnitude of our findings, we note that our models predicted that students taught by a first-year teacher will score 0.0730 standard deviations lower in middle school math, 0.0529 standard deviations lower in high school math, and 0.0349 standard deviations lower in ninth-grade biology, all else equal, than students taught by other teachers.[23] Secondly, when we

[22] See Goldhaber et al. (2013) for details about the elementary data used to estimate these models.
[23] Note that these estimates are from the sample of teachers with WEST-B scores in each grade level, which is considerably less experienced than the overall teacher workforce (see Table 2).

estimate VAMs with a teacher fixed effect and calculate the standard deviation of teacher value-added (the teacher "effect size"), we find that the teacher effect size is 0.16 in middle school math, 0.27 in high school math, and 0.21 in ninth-grade biology.

Basic Skills Licensure Tests and Student Achievement

We now turn to the secondary STEM classrooms that are the primary focus of this analysis. **Table 4** shows the estimated relationships between WEST-B Math scores and student performance in middle school math (Panel A), ninth-grade algebra and geometry (Panel B), and ninth-grade biology (Panel C)[24]; note that we do not estimate models that consider whether candidates passed the WEST-B Math test on first attempt because so few candidates in the sample failed this test. The results in middle school math and ninth-grade algebra and geometry are broadly consistent with the results at the elementary level; a one standard deviation increase in a teacher's WEST-B Math score is correlated with a .014-.033 standard deviation increase in student math performance (though these coefficients are somewhat imprecisely estimated so are only statistically significant in middle school math). Thus, the expected increase in student performance associated with a one standard deviation increase in the teacher's WEST-B score is roughly equivalent to one-fifth to one-half of the expected increase in student performance associated with having a non-novice teacher relative to a first-year teacher.

When we allow for nonlinearities in the relationship between WEST-B scores and student math performance (columns 4-6 of Table 4), the expected difference in student performance associated with having a teacher who scored in the top quartile of the WEST-B Math relative to the bottom quartile is .045 to .054 standard deviations of student performance in middle school math. This is roughly one-thirdof a standard deviation of teacher performance in these grades. On the other hand, the comparable

[24] We also estimate models that consider other WEST-B tests, separately and jointly, the mean WEST-B score across subtests, and the maximum WEST-B score rather than the first WEST-B score. These results are available from the authors upon request.

different in ninth-grade algebra and geometry is just .006 to .018 standard deviations of student performance (see Figure 6).[25]

Finally, and perhaps most surprisingly, the relationships between WEST-B Math scores and student performance in ninth-grade biology are considerably stronger than in other grade levels (see Figures 6 and 7); a one standard deviation increase in a teacher's WEST-B Math score is correlated with a .079 to .155 standard deviation increase in student biology performance, and the expected difference in student performance associated with having a teacher who scored in the top quartile of the WEST-B Math relative to the bottom quartile is .125 to .197 standard deviations of student performance. To put this in context, this means that the expected difference in student biology performance associated with having a teacher in the top quartile of the WEST-B Math distribution relative to the bottom quartile is about a standard deviation of teacher effectiveness in ninth-grade biology, or roughly equivalent to the expected difference associated with having a teacher at the 83rd percentile of the ninth-grade biology value-added distribution relative to an average teacher.

Subject-Specific Licensure Tests and Student Achievement

Table 5 shows the estimated relationships between WEST-E scores and student performance in middle and high school math.[26] The estimates in Panel A of Table 5 give somewhat mixed evidence about the relationship between WEST-E Middle-Level Math (MLM) scores and student performance in middle school math. Specifically, the relationships between WEST-E MLM scores—either continuous scores (columns 1–3), a passing indicator (columns 4–6), and quartiles of performance (columns 7–9)—tend to be statistically significant (and comparable in magnitude to the WEST-B estimates from Table 4) when comparisons are made within schools, but not in the models without school or school-by-year

[25] Estimates from a student fixed-effects model in middle school math are broadly consistent with these results (available from the authors upon request).
[26] We also estimate models that control for WEST-B scores and find that the coefficients on WEST-E scores are generally positive but not statistically significant. These results are available from the authors upon request.

fixed effects. On the other hand, the estimates in Panels B and C of Table 5 give little evidence that WEST-E Math scores are predictive of student performance in middle school math or ninth-grade algebra and geometry, although the magnitude of the cross-school estimates for ninth-grade algebra and geometry (columns 1 and 4 of Panel C and the lower-left panel of Figure 8) are positive, relatively large, and marginally statistically significant.[27]

Finally, **Table 6** presents estimates of the relationships between each of the WEST-E tests that teachers can pass to teach high school biology (the Science and Biology tests) and student biology performance in ninth grade. Echoing the results for the WEST-B Math, the relationships between these test scores and student performance in ninth-grade biology tend to be large and statistically significant. The magnitudes of these coefficients are striking; for example, the expected difference in student performance in ninth-grade biology associated with having a teacher who passed the WEST-E Biology exam relative to a teacher who failed it on the first test administration is .267, or over a standard deviation of teacher effectiveness in ninth-grade biology (.21). Figure 8 reinforces that, as for the WEST-B Math, the WEST-E tests are a much stronger predictor of student performance in ninth-grade biology than in the other grade levels we consider. Figure 9 suggests that some of these results may be driven by influential observations (such as the outlier in the lower left corner of the two ninth-grade biology figures), but even when we drop influential observations (with leverage greater than 0.1) we still find statistically significant relationships (with virtually the same magnitudes).

Extensions and Robustness Checks

Given that the results for the subject-specific WEST-E tests (section B) are quite similar to the results for the basic skills WEST-B tests (section A), a natural question is whether WEST-E test scores provide any more signal about future teacher effectiveness than is already contained in the WEST-B test

[27] We do not present results for the WEST-E MLM test in ninth-grade algebra because of small sample sizes.

scores. To investigate this, **Table 7** presents estimates of the relationships between WEST-E scores and student performance in middle and high school math controlling for each teacher's WEST-B scores.[28] In middle school math, estimates from models based on within-school comparisons (columns 2, 3, 5, and 6) suggest that WEST-E MLM and WEST-E Math test scores do provide additional signal about future teacher effectiveness beyond WEST-B scores. That said, this does not appear to be the case in high school math, and perhaps more surprisingly, it does not appear to be the case when we investigate relationships between WEST-E scores and student performance in ninth-grade biology controlling for each teacher's WEST-B scores in **Table 8**. This suggests that the large and statistically significant relationships between WEST-E scores and student performance in ninth-grade biology shown in Table 6 can largely be explained by the relationships with WEST-B scores shown in Table 4.

Another natural question about these results, particularly given evidence about the extent of student tracking in secondary grades (Jackson, 2014), is whether the estimates are biased by higher performing students (along unobserved dimensions) being disproportionately assigned to teachers with higher credential test scores. We investigate the potential extent of this source of bias in two ways. First, we estimate some models that control for student covariates aggregated to the teacher level and find that the inclusion of these controls does not substantively change any of the primary results.[29] Second, following the procedure described in Clotfelter et al. (2006), we estimate models in middle school math that are restricted to schools in which students are distributed relatively equitably across classrooms. The coefficients of interest in these models are smaller in magnitude but still statistically significant. Both sets of results suggest that the primary findings are not driven solely by the nonrandom sorting of students to classrooms.

[28] In each model, we control for continuous WEST-B scores in math, reading, and writing.
[29] Results available from authors upon request.

In another extension of the results in Tables 4–6, we consider models that interact teacher credential test scores with different student characteristics (e.g., prior performance, participation in FRL, student URM indicator) to test whether credential test scores are differentially predictive of student performance for different types of students.[30] We find little evidence of differential effects by student prior performance or demographics.

Finally, to test whether the predictive power of subject-specific licensure tests might matter more depending on the nature of the course taught, we estimate models that interact teacher credential test scores with an indicator for whether the course is designated as an "advanced course" or a "remedial course". Due to sample size limitations, we were able to estimate WEST-E advanced course models only for middle and high school math classes. The interaction term between the WEST-E MLM test and the advanced course indicator was positive and significant for seventh- and eighth-grade math courses, while the interaction term between the WEST-E Math test and the remedial course indicator was positive and significant for the same grade levels. This suggests that the subject-specific knowledge that teachers possess (as measured by these subject tests) may be more important for teaching in these "focused" classrooms than in other classrooms.[31]

Conclusion

The results from this study suggest several broad conclusions and directions for future research. First, the findings from elementary and middle school about the modest, positive relationship between WEST-B Math scores and student math performance reinforce conclusions from the existing literature

[30] These estimates are available from the authors upon request.
[31] Due to sample size limitations, we were able to estimate WEST-E advanced course models for seventh and eighth-grade math classes as well as ninth-grade algebra and geometry classes. An interaction term between the WEST-E MLM test and an advanced course indicator was positive and significant for seventh- and eighth-grade math courses. We found no evidence of a significant interaction between the WEST-E Math test and an advanced course interaction.

(e.g., Clotfelter et al., 2007; Goldhaber, 2007) that basic skills credential test scores provide a significant, if modest, signal about future math teacher effectiveness. Given the very limited evidence about preservice predictors of future teacher effectiveness (e.g., Harris & Sass, 2011), this suggests that basic skills test scores could be used for reasons beyond the pass/fail requirement for initial teacher credentialing (for example, as a measure of content knowledge for teaching for hiring and other personnel decisions).

The second broad conclusion is that subject-specific credential test scores provide some additional signal about teacher effectiveness in some subjects, although the relationships are not always statistically significant. The key policy question, then, is whether these results justify the barrier to entry they represent to potential STEM teachers. Our preliminary analysis in Section III suggests that the WEST-E tests in STEM fields are much more difficult to pass than the WEST-E tests in other fields like elementary education. Moreover, teachers who fail the WEST-E the first time they take it are about 10 percentage points less likely to enter the workforce, and teacher candidates of color tend to be more likely to fail these tests than White teacher candidates (Goldhaber & Hansen, 2013), so are disproportionately impacted by this barrier to entry. These trends could be particularly problematic given the well-documented difficulty of school districts, and districts in Washington state in particular, to attract STEM teachers and teachers of color (Goldhaber et al., 2015a, 2015b).

The final broad conclusion, and a unique contribution of this paper, relates to our investigation of the impact of teachers on science test scores and, specifically, the finding that relationships between credential test scores and student performance in ninth-grade biology are considerably stronger than in math classrooms. Given that ninth-graders who take biology in Washington state are a high-performing subgroup of all ninth-graders (see Table 1), this may be partially a function of credential test scores being more predictive of the performance of more advanced students. However, a more intriguing explanation is that teacher content knowledge (as measured by credential tests) is simply more important to student performance in science than in math. We caution against such a broad interpretation based on the relatively small ninth-grade biology sample sizes in this paper, but this possibility is certainly worthy of future investigation.

We conclude by suggesting a future line of research in this area. Given empirical evidence about the influence of teachers on non-tested outcomes (e.g., Jackson, 2012), the results described in this paper suggest that STEM teachers with high credential test performance may influence other STEM outcomes we care about (e.g., future STEM course taking and performance, majoring in STEM fields, and employment in STEM industries). Specifically, the development of P-20 data warehouses across the country might allow researchers to investigate the role of STEM teachers in influencing each of these important outcomes.

References

Aaronson, D., Barrow, L., and Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1):95-135.

Appleton, K. (2013). *Elementary science teacher education: International perspectives on contemporary issues and practice*. Routledge.

Bacher-Hicks, A., Kane, T. J., & Staiger, D. O. (2014). Validating Teacher Effect Estimates using Changes in Teacher Assignments in Los Angeles (No. 20657). Cambridge, MA: National Bureau of Economic Research.

Baker, A. (2012). To earn classroom certification, more teaching and less testing. *The New York Times*, July 30, 2012.

Barmore, P. (2016). Will controversial new tests for teachers make the profession even more overwhelmingly white? *The Hechinger Report*, April 4, 2016. Accessed from http://hechingerreport.org/will-controversial-new-tests-for-teachers-make-the-profession-even-more-overwhelmingly-white/

Betts, J. R., Zau, A., and Rice, L. (2003). Determinants of student achievement: New evidence from San Diego. Public Policy Institute of California San Francisco.

Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., and Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis*, 31(4):416-440.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9), 2593–2632.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633–2679.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, 41(4), 778-820.

Clotfelter, C. T., Ladd, H. F., and Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, 26(6): 673-682.

Clotfelter, C., Ladd H. F., and Vigdor, J. (2010). Teacher Credentials and Student Achievement in High School a Cross-Subject Analysis with Student Fixed Effects. *Journal of Human Resources,* 45 (3):655-681.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., and York, R. (1966). Equality of Educational Opportunity. *Washington, DC*, pages1066–5684.

Gitomer, D. H. (2007). Teacher Quality in a Changing Policy Landscape: Improvements in the Teacher Pool. *Education Testing Service*.

Goldhaber, D. (2007). Everyone's Doing It, but What Does Teacher Testing Tell Us About Teacher Effectiveness? *Journal of Human Resources* 42 (4): 765–94.

Goldhaber, D., and Brewer, D. J. (1997). Why Don't Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity. *Journal of Human Resources,* 505–23.

Goldhaber, D., and Brewer, D. J. (2000). Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement. *Educational Evaluation and Policy Analysis* 22 (2): 129–45.

Goldhaber, D., and Hansen, M. (2013). Is It Just a Bad Class? Assessing the Long-Term Stability of Estimated Teacher Performance. *Economica* 80 (319): 589–612.

Goldhaber, D., and Walch, J. (2014). Gains in Teacher Quality: Academic Capabilities of the US Teaching Force Are on the Rise. *Education Next* 14 (1): 38.

Goldhaber, D., Cowan, J., and Theobald, R. (2016a). Evaluating prospective teachers: Testing the predictive validity of the edTPA. CALDER Working Paper 157.

Goldhaber, D., Liddle, S., and Theobald, R. (2013). The gateway to the profession: Evaluating teacher preparation programs based on student achievement. *Economics of Education* Review 34: 29-44.

Goldhaber, D., Grout, C., & Huntington-Klein, N. (2014a). Screen twice, cut once: Assessing the predictive validity of teacher selection tools. CEDR Working Paper 2014-9. University of Washington, Seattle, WA.

Goldhaber, D., Krieg, J., Theobald, R., and Brown, N. (2015a). Refueling the STEM and special education teacher pipelines. *Phi Delta Kappan* 97, pp. 56-62.

Goldhaber, D., Theobald, R., and Tien, C. (2015b). Educator and student diversity in Washington State: Gaps and historical trends. CEDR Policy Brief 2015-10.

Goldhaber, D., Krieg, J., and Theobald, R. (2016b). Does the match matter? Exploring whether student teaching experiences affect teacher effectiveness and career paths. CALDER Working Paper 149.

Harris, D. N., and Sass, T. R. (2011). Teacher Training, Teacher Quality and Student Achievement. *Journal of Public Economics* 95 (7): 798–812.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153–161.

Hendricks, M. D. (2014). Public Schools Are Hemorrhaging Talented Teachers. Can Higher Salaries Function as a Tourniquet? Association for Education Finance and Policy.

Hill, H.C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American educational research journal,* 42(2), 371-406.

Jackson, C. K. (2012). Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina (No. w18624). National Bureau of Economic Research.

Jackson, C. K. (2014). Teacher quality at the high school level: The importance of accounting for tracks. *Journal of Labor Economics*, 32(4), 645–684.

Kane, T. J, and Staiger, D. O. (2008). *Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation*. Technical report, National Bureau of Economic Research.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). Have We Identified Effective Teachers? Seattle, WA: Bill and Melinda Gates Foundation.

Koedel, C., Mihaly, K., and Rockoff, J. E. (2015). Value-Added Modeling: A Review. *Economics of Education Review*.

Lankford, H., Loeb, S., McEachin, A., Miller, L. C., and Wycoff, J. (2014). Who Enters Teaching? Encouraging Evidence That the Status of Teaching Is Improving. *Educational Researcher* 43 (9): 444–53.

Meskauskas, J.A. (1976) Evaluation models for criterion-referenced testing: Views regarding mastery and standard-setting. *Review of Educational Research,* 46(1), 133-158.

Monk, D. H., and King, J. A. (1994). Multilevel Teacher Resource Effects in Pupil Performance in Secondary Mathematics and Science. The Case of Teacher Subject Matter Preparation. *Choices and Consequences: Contemporary Policy Issues in Education*. Pages 29–58.

President's Council of Advisors on Science and Technology (US). (2010). *Prepare and Inspire: K-12 Education in Science, Technology, Engineering, and Math (STEM) for America's Future*. Executive Office of the President.

Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica* 73 (2): 417–58.

Rockoff, J. E. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review*: 247–52.

Wilson, S. M., Floden, R. E., and Ferrini-Mundy, J. (2001). *Teacher Preparation Research: Current Knowledge, Gaps, and Recommendations: A Research Report Prepared for the US Department of Education and the Office for Educational Research and Improvement, February 2001*. Center for the Study of Teaching and Policy

Tables

Table 1. Student-Year Level Summary Statistics

	1	2	3	4	5	6	7	8

Credential Test	WEST-B Math	WEST-B Math	WEST-B Math	WEST-E MLM	WEST-E Math	WEST-E Math	WEST-E Science	WEST-E Biology
Grade(s)	7th, 8th	9th	9th	7th, 8th	7th, 8th	9th	9th	9th
Subject	Math	Alg/Geo	Biology	Math	Math	Alg/Geo	Biology	Biology
Lagged Math	0.105 (0.928)	-0.018 (0.808)	0.425 (0.988)	0.066 (0.905)	0.268 (0.941)	-0.094 (0.793)	0.318 (0.987)	0.388 (0.954)
Lagged Reading	0.095 (0.920)	0.031 (0.860)	0.356 (0.914)	0.072 (0.912)	0.234 (0.906)	-0.033 (0.856)	0.264 (0.925)	0.318 (0.910)
Lagged Science		-0.009 (0.862)	0.378 (0.970)			-0.087 (0.859)	0.241 (0.969)	0.335 (0.933)
Female	0.496 (0.500)	0.501 (0.500)	0.516 (0.500)	0.496 (0.500)	0.498 (0.500)	0.495 (0.500)	0.507 (0.500)	0.506 (0.500)
Multi-racial	0.048 (0.214)	0.044 (0.204)	0.043 (0.203)	0.049 (0.216)	0.047 (0.212)	0.042 (0.201)	0.049 (0.215)	0.046 (0.210)
American Indian	0.017 (0.129)	0.018 (0.133)	0.017 (0.130)	0.018 (0.132)	0.017 (0.130)	0.018 (0.134)	0.024 (0.153)	0.020 (0.141)
Asian/ Pacific Isl.	0.109 (0.312)	0.090 (0.286)	0.132 (0.339)	0.105 (0.306)	0.147 (0.355)	0.089 (0.284)	0.151 (0.358)	0.167 (0.373)
Black	0.059 (0.236)	0.060 (0.238)	0.052 (0.221)	0.063 (0.243)	0.066 (0.249)	0.069 (0.253)	0.068 (0.252)	0.069 (0.254)
Hispanic	0.213 (0.410)	0.216 (0.411)	0.160 (0.366)	0.204 (0.403)	0.171 (0.376)	0.246 (0.431)	0.192 (0.394)	0.163 (0.369)
Gifted	0.074 (0.262)	0.027 (0.163)	0.075 (0.263)	0.074 (0.261)	0.113 (0.317)	0.026 (0.158)	0.060 (0.237)	0.093 (0.291)
LEP	0.050 (0.217)	0.044 (0.205)	0.023 (0.149)	0.052 (0.222)	0.039 (0.193)	0.053 (0.225)	0.032 (0.177)	0.025 (0.156)
Spec. Ed.	0.061 (0.240)	0.051 (0.220)	0.058 (0.234)	0.062 (0.242)	0.057 (0.232)	0.060 (0.237)	0.079 (0.270)	0.070 (0.255)
FRL	0.483 (0.500)	0.486 (0.500)	0.376 (0.484)	0.497 (0.500)	0.428 (0.495)	0.523 (0.499)	0.415 (0.493)	0.384 (0.486)
Learning Disability	0.033 (0.177)	0.028 (0.165)	0.033 (0.177)	0.032 (0.177)	0.028 (0.164)	0.033 (0.180)	0.044 (0.204)	0.035 (0.184)
Seventh Grade	0.522 (0.500)			0.541 (0.498)	0.434 (0.496)			
Eighth Grade	0.478 (0.500)			0.459 (0.498)	0.566 (0.496)			
Observations	135079	54512	15116	50764	37009	24832	5141	6046

Table 2. Teacher-Year Level Summary Statistics

	1	2	3	4	5	6	7	8
Credential Test	WEST-B Math	WEST-B Math	WEST-B Math	WEST-E MLM	WEST-E Math	WEST-E Math	WEST-E Science	WEST-E Biology
Grade(s)	7th, 8th	9th	9th	7th, 8th	7th, 8th	9th	9th	9th
Subject	Math	Alg/Geo	Biology	Math	Math	Alg/Geo	Biology	Biology
Exp: 1 Year	0.123 (0.329)	0.125 (0.331)	0.100 (0.300)	0.180 (0.385)	0.160 (0.367)	0.234 (0.424)	0.203 (0.404)	0.207 (0.406)
Exp: 2 Year	0.126 (0.331)	0.106 (0.308)	0.100 (0.300)	0.138 (0.346)	0.132 (0.339)	0.149 (0.356)	0.188 (0.392)	0.193 (0.396)
Exp: 3 Year	0.129 (0.335)	0.111 (0.314)	0.097 (0.297)	0.106 (0.308)	0.095 (0.293)	0.102 (0.303)	0.109 (0.313)	0.124 (0.331)
Exp: 4 Year	0.135 (0.342)	0.122 (0.328)	0.114 (0.319)	0.090 (0.287)	0.082 (0.274)	0.059 (0.235)	0.063 (0.243)	0.062 (0.242)
Exp: 5 plus	0.394 (0.489)	0.406 (0.491)	0.466 (0.500)	0.345 (0.476)	0.347 (0.476)	0.192 (0.394)	0.141 (0.349)	0.138 (0.346)
Advanced Degree	0.556 (0.497)	0.568 (0.495)	0.703 (0.458)	0.570 (0.495)	0.579 (0.494)	0.532 (0.499)	0.688 (0.465)	0.703 (0.458)
WEST-B Math	0.543 (0.595)	0.709 (0.519)	0.653 (0.461)					
WEST-B Reading	0.179 (0.802)	0.234 (0.847)	0.547 (0.657)					
WEST-B Writing	0.193 (0.870)	0.225 (0.869)	0.544 (0.706)					
WEST-E MLM				0.280 (0.786)				
WEST-E Math					-0.068 (0.832)	0.266 (0.724)		
WEST-E Science							0.184 (0.893)	
WEST-E Biology								0.219 (0.898)
Observations	2118	1646	350	809	539	765	128	145
Unique Tch	914	773	185	387	256	427	90	92

Table 3. Value-Added Model (VAM) of Student Math Achievement at the Elementary Level

Panel A: Elementary Math, WEST-B Math						
WEST-B Math Standardized Score	0.027*** (0.005)	0.024*** (0.005)	0.034*** (0.005)			
WEST-B Math Q2				0.034** (0.012)	0.033** (0.011)	0.038** (0.012)
WEST-B Math Q3				0.047*** (0.012)	0.038*** (0.011)	0.048*** (0.011)
WEST-B Math Q4				0.054*** (0.013)	0.047*** (0.012)	0.069*** (0.012)
School fixed effects		X			X	
School-by-year fixed effects			X			X
Number unique teachers	2966	2788	2393	2966	2788	2393
Number students	158,459	149,269	113,137	158,459	149,269	113,137
Panel B: Elementary Math, WEST-E Elementary Subtest I (Elem I)						
WEST-E Elem I Standardized Score	0.002 (0.009)	-0.001 (0.009)	0.001 (0.009)			
WEST-E Elem I Q2				0.033 (0.020)	0.017 (0.023)	-0.014 (0.024)
WEST-E Elem I Q3				0.056* (0.022)	0.034 (0.022)	0.031 (0.023)
WEST-E Elem I Q4				0.014 (0.023)	0.001 (0.024)	-0.003 (0.025)
School fixed effects		X			X	
School-by-year fixed effects			X			X
Number unique teachers	1048	801	659	1048	801	659
Number students	37,426	29,748	21,796	37,426	29,748	21,796
Panel C: Elementary Math, WEST-E Elementary Subtest II (Elem II)						
WEST-E Elem II Standardized Score	0.003 (0.009)	-0.002 (0.009)	0.005 (0.009)			
WEST-E Elem II Q2				0.024 (0.021)	-0.002 (0.022)	0.001 (0.022)
WEST-E Elem II Q3				0.013 (0.023)	-0.014 (0.024)	0.013 (0.023)
WEST-E Elem II Q4				0.023 (0.024)	0.002 (0.023)	0.015 (0.023)
School fixed effects		X			X	
School-by-year fixed effects			X			X
Number unique teachers	1048	801	659	1048	801	659
Number students	37,426	29,748	21,796	37,426	29,748	21,796

NOTE: p-values from two-sided t-test: +p<0.1, *p<0.05, **p<0.01, ***p<0.001. All models control for prior scores interacted by year and student demographics (see Table 1) and teacher experience and degree level (see Table 2). "Number unique teachers" refers to teachers who identify the model, and "Number students" is the number of students for those teachers. Standard errors are clustered at the teacher level.

Table 4. VAM of Student Math or Biology Achievement at the Middle or High School Level, Basic Skills Test

Panel A: Middle School Math, WEST-B Math						
WEST-B Math Standardized Score	0.025* (0.012)	0.029** (0.010)	0.023* (0.010)			
WEST-B Math Q2				0.013 (0.017)	0.027+ (0.015)	0.017 (0.016)
WEST-B Math Q3				0.007 (0.019)	0.015 (0.016)	0.012 (0.016)
WEST-B Math Q4				0.049** (0.019)	0.054** (0.017)	0.045* (0.018)
School fixed effects		X			X	
School-by-year fixed effects			X			X
Number unique teachers	914	820	701	914	820	701
Number students	135,079	123,888	95,823	135,079	123,888	95,823
Panel B: Ninth Grade Algebra and Geometry, WEST-B Math						
WEST-B Math Standardized Score	0.033 (0.023)	0.014 (0.017)	0.014 (0.015)			
WEST-B Math Q2				0.047 (0.031)	0.011 (0.019)	0.010 (0.021)
WEST-B Math Q3				0.037 (0.034)	0.015 (0.022)	0.017 (0.021)
WEST-B Math Q4				0.013 (0.038)	0.006 (0.028)	0.018 (0.033)
School fixed effects		X			X	
School-by-year fixed effects			X			X
Number unique teachers	773	694	601	773	694	601
Number students	54,354	49,173	39,375	54,354	49,173	39,375
Panel C: Ninth Grade Biology, WEST-B Math						
WEST-B Math Standardized Score	0.155*** (0.033)	0.068* (0.028)	0.079*** (0.018)			
WEST-B Math Q2				0.155** (0.049)	0.070* (0.033)	0.053+ (0.028)
WEST-B Math Q3				0.159* (0.062)	0.015 (0.041)	-0.022 (0.042)
WEST-B Math Q4				0.197*** (0.053)	0.130*** (0.034)	0.125*** (0.032)
School fixed effects		X			X	
School-by-year fixed effects			X			X
Number unique teachers	185	141	113	185	141	113
Number students	15,144	11,417	8318	15,144	11,417	8318

NOTE: p-values from two-sided t-test: +p<0.1, *p<0.05, **p<0.01, ***p<0.001. All models control for prior scores interacted by year and student demographics (see Table 1) and teacher experience and degree level (see Table 2). “Number unique teachers” refers to teachers who identify the model, and “Number students” is the number of students for those teachers. Standard errors are clustered at the teacher level.

Table 5. VAM of Student Math Achievement at the Middle or High School Level, Subject-Specific Tests

Panel A: Middle School Math, WEST-E Middle Level Math (MLM)									
WEST-E MLM Standardized Score	0.017 (0.011)	0.032** (0.011)	0.038** (0.014)						
WEST-E MLM Passing Score				0.001 (0.019)	0.040+ (0.023)	0.046 (0.029)			
WEST-E MLM Q2							-0.043+ (0.025)	-0.021 (0.027)	0.003 (0.033)
WEST-E MLM Q3							-0.002 (0.024)	0.055* (0.025)	0.083** (0.028)
WEST-E MLM Q4							0.042+ (0.025)	0.053+ (0.029)	0.078* (0.036)
School FEs		X			X			X	
School-by-year FEs			X			X			X
Unique teachers	387	285	223	387	285	223	387	285	223
Number students	50,764	36,456	21,773	50,764	36,456	21,773	50,764	36,456	21,773
Panel B: Middle School Math, WEST-E Math									
WEST-E Math Standardized Score	-0.002 (0.013)	0.005 (0.012)	-0.001 (0.013)						
WEST-E Math Passing Score				-0.018 (0.021)	-0.013 (0.022)	-0.039 (0.027)			
WEST-E Math Q2							-0.031 (0.030)	0.023 (0.043)	0.030 (0.034)
WEST-E Math Q3							-0.041 (0.032)	-0.015 (0.040)	-0.035 (0.047)
WEST-E Math Q4							-0.008 (0.031)	0.033 (0.037)	0.000 (0.033)
School FEs		X			X			X	
School-by-year FEs			X			X			X
Unique teachers	256	161	106	256	161	106	256	161	106
Number students	37,009	23,044	12,501	37,009	23,044	12,501	37,009	23,044	12,501
Panel C: Ninth Grade Algebra and Geometry, WEST-E Math									
WEST-E Math Standardized Score	0.040+ (0.022)	0.011 (0.013)	0.010 (0.014)						
WEST-E Math Passing Score				0.039 (0.034)	-0.002 (0.023)	0.001 (0.023)			
WEST-E Math Q2							0.048 (0.038)	-0.009 (0.028)	0.004 (0.027)
WEST-E Math Q3							0.024 (0.042)	-0.018 (0.035)	-0.081* (0.034)
WEST-E Math Q4							0.065 (0.045)	0.015 (0.030)	0.017 (0.030)
School FEs		X			X			X	
School-by-year FEs			X			X			X
Unique teachers	427	331	249	427	331	249	427	331	249
Number students	24,771	19,737	12,473	24,771	19,737	12,473	24,771	19,737	12,473

NOTE: p-values from two-sided t-test: +p<0.1, *p<0.05, **p<0.01, ***p<0.001. All models control for prior scores interacted by year and student demographics (see Table 1) and teacher experience and degree level (see Table 2). "Number unique teachers" refers to teachers who identify the model, and "Number students" is the number of students for those teachers. Standard errors are clustered at the teacher level.

Table 6. Value-Added Model of Student Biology Achievement at the High School Level, Subject-Specific Tests

Panel A: Ninth Grade Biology, WEST-E Science

WEST-E Science Standardized Score	0.100** (0.033)		
WEST-E Science Passing		0.147 (0.092)	
WEST-E Science Q2			0.158 (0.100)
WEST-E Science Q3			0.208* (0.083)
WEST-E Science Q4			0.194* (0.084)
School fixed effects			
School-by-year fixed effects			
Number unique teachers	90	90	90
Number students	5148	5148	5148
Panel B: Ninth Grade Biology, WEST-E Biology			
WEST-E Biology Standardized Score	0.067+ (0.040)		
WEST-E Passing		0.267* (0.103)	
WEST-E Biology Q2			0.032 (0.077)
WEST-E Biology Q3			0.043 (0.111)
WEST-E Biology Q4			0.085 (0.070)
School fixed effects			
School-by-year fixed effects			
Number unique teachers	92	92	92
Number students	6061	6061	6061

NOTE: p-values from two-sided t-test: *p<0.05, **p<0.01, ***p<0.001. All models control for prior scores interacted by year and student demographics (see Table 1) and teacher experience and degree level (see Table 2). "Number unique teachers" refers to teachers who identify the model, and "Number students" is the number of students for those teachers. Standard errors are clustered at the teacher level.

Table 7. VAM of Student Achievement at Middle or High School level, Subject-Specific Tests Controlling for Basic-Skills Tests

Panel A: Middle School Math, WEST-E Middle Level Math (MLM) Controlling for WEST-B Scores									
WEST-E MLM Standardized Score	0.026 (0.017)	0.037+ (0.020)	0.086** (0.026)						
				0.016	0.033	0.062			

WEST-E MLM Passing Score				(0.029)	(0.033)	(0.043)			
WEST-E MLM Q2							-0.029 (0.031)	-0.013 (0.034)	0.002 (0.042)
WEST-E MLM Q3							0.030 (0.030)	0.013 (0.035)	0.092* (0.039)
WEST-E MLM Q4							0.059+ (0.033)	0.139*** (0.036)	0.175*** (0.049)
School FEs		X			X			X	
School-by-year FEs			X			X			X
Unique teachers	275	190	143	275	190	143	275	190	143
Number students	32336	22106	11406	32336	22106	11406	32336	22106	11406
Panel B: Middle School Math, WEST-E Math Controlling for WEST-B Scores									
WEST-E Math Standardized Score	0.011 (0.018)	0.047* (0.023)	0.041* (0.018)						
WEST-E Math Passing Score				0.016 (0.031)	0.042 (0.038)	0.011 (0.037)			
WEST-E Math Q2							-0.045 (0.037)	0.176*** (0.051)	0.170*** (0.047)
WEST-E Math Q3							-0.004 (0.037)	0.145** (0.055)	0.083 (0.059)
WEST-E Math Q4							0.017 (0.041)	0.170** (0.059)	0.116** (0.044)
School FEs		X			X			X	
School-by-year FEs			X			X			X
Unique teachers	190	97	62	190	97	62	190	97	62
Number students	25582	12218	7217	25582	12218	7217	25582	12218	7217
Panel C: Ninth Grade Algebra and Geometry, WEST-E Math Controlling for WEST-B Scores									
WEST-E Math Standardized Score	0.024 (0.027)	-0.002 (0.019)	-0.021 (0.021)						
WEST-E Math Passing Score				0.004 (0.040)	-0.026 (0.028)	-0.040 (0.030)			
WEST-E Math Q2							0.030 (0.044)	-0.016 (0.030)	-0.029 (0.035)
WEST-E Math Q3							-0.019 (0.051)	-0.040 (0.040)	-0.098* (0.043)
WEST-E Math Q4							0.013 (0.053)	-0.044 (0.039)	-0.053 (0.043)
School FEs		X			X			X	
School-by-year FEs			X			X			X
Unique teachers	339	245	178	339	245	178	339	245	178
Number students	19949	14645	8614	19949	14645	8614	19949	14645	8614

NOTE: p-values from two-sided t-test: +p<0.1, *p<0.05, **p<0.01, ***p<0.001. All models control for prior scores interacted by year, student demographics (see Table 1), and teacher experience, degree level, and WEST-B scores (see Table 2). "Number unique teachers" refers to teachers who identify the model, and "Number students" is the number of students for those teachers. Standard errors are clustered at the teacher level.

Table 8. Value-Added Model of Student Biology Achievement at the High School Level, Subject-Specific Tests Controlling for Basic-Skills Tests

Panel A: Ninth Grade Biology, WEST-E Science Controlling for WEST-B Scores			
	0.015		

WEST-E Science Standardized Score	(0.027)		
WEST-E Science Passing		0.046 (0.061)	
WEST-E Science Q2			0.078 (0.056)
WEST-E Science Q3			0.019 (0.070)
WEST-E Science Q4			-0.010 (0.076)
School fixed effects			
School-by-year fixed effects			
Number unique teachers	71	71	71
Number students	4321	4321	4321
Panel B: Ninth Grade Biology, WEST-E Biology Controlling for WEST-B Scores			
WEST-E Biology Standardized Score	0.013 (0.043)		
WEST-E Passing		0.117 (0.103)	
WEST-E Biology Q2			0.014 (0.049)
WEST-E Biology Q3			-0.008 (0.086)
WEST-E Biology Q4			-0.058 (0.070)
School fixed effects			
School-by-year fixed effects			
Number unique teachers	76	76	76
Number students	5105	5105	5105

NOTE: p-values from two-sided t-test: *p<0.05, **p<0.01, ***p<0.001. All models control for prior scores interacted by year and student demographics (see Table 1) and teacher experience, degree level, and WEST-B scores (see Table 2). “Number unique teachers” refers to teachers who identify the model, and “Number students” is the number of students for those teachers. Standard errors are clustered at the teacher level.

Figures

Figure 1. Average WEST-B Scores by Subtest and Testing Year

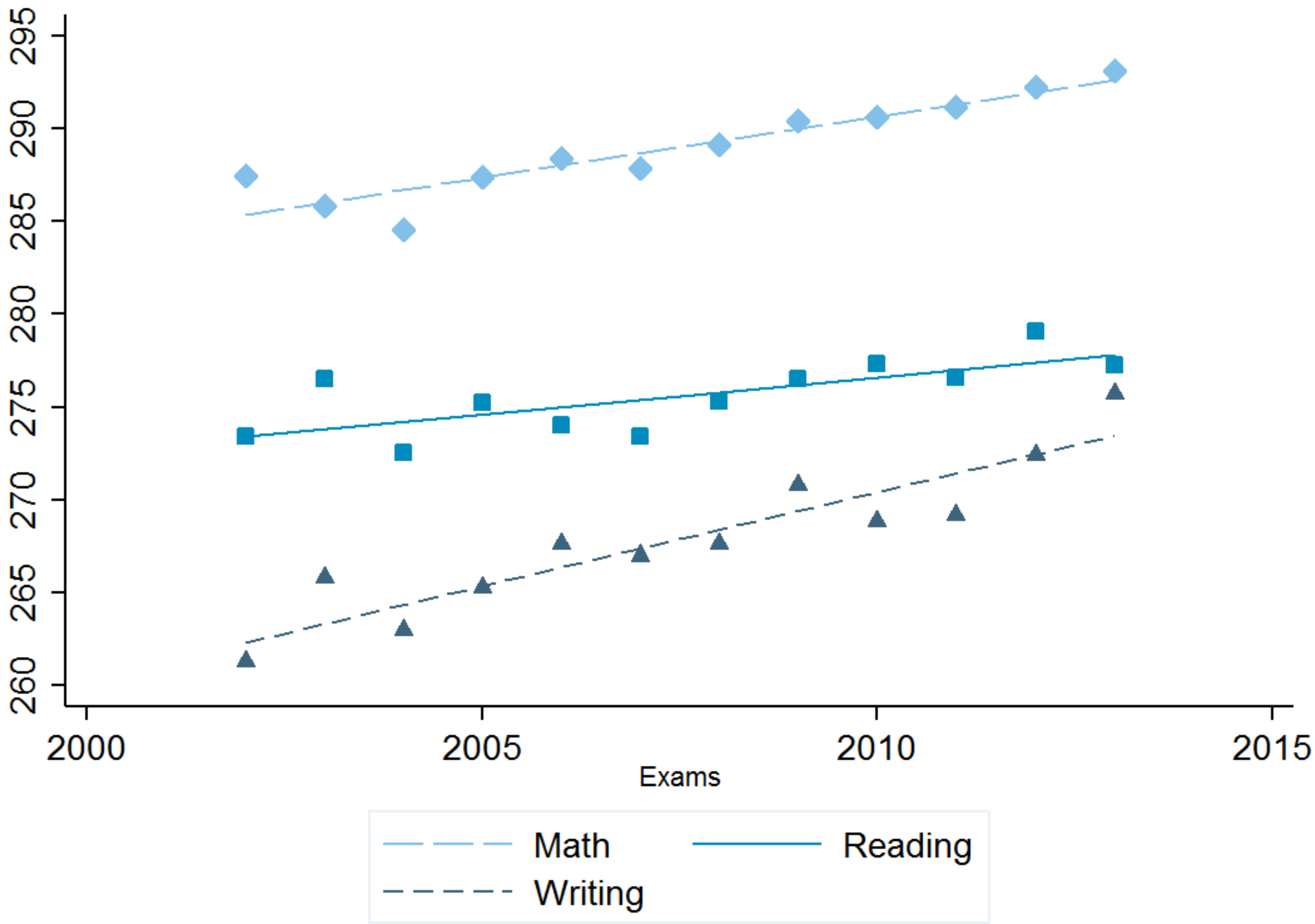


Figure 2. WEST-B Scores by Subtest and Analytic Sample

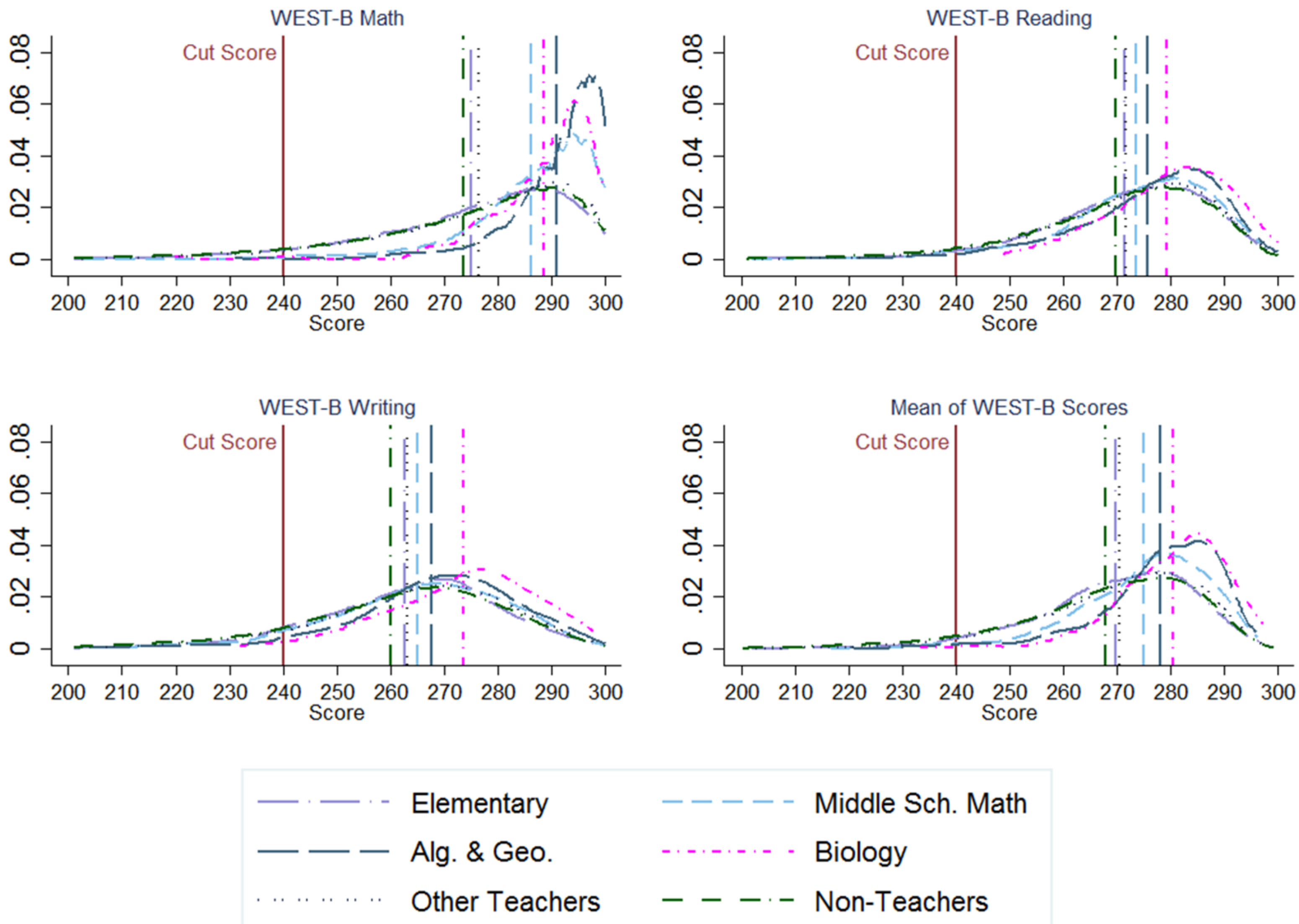


Figure 3. WEST-E Scores by Subtest and Analytic Sample

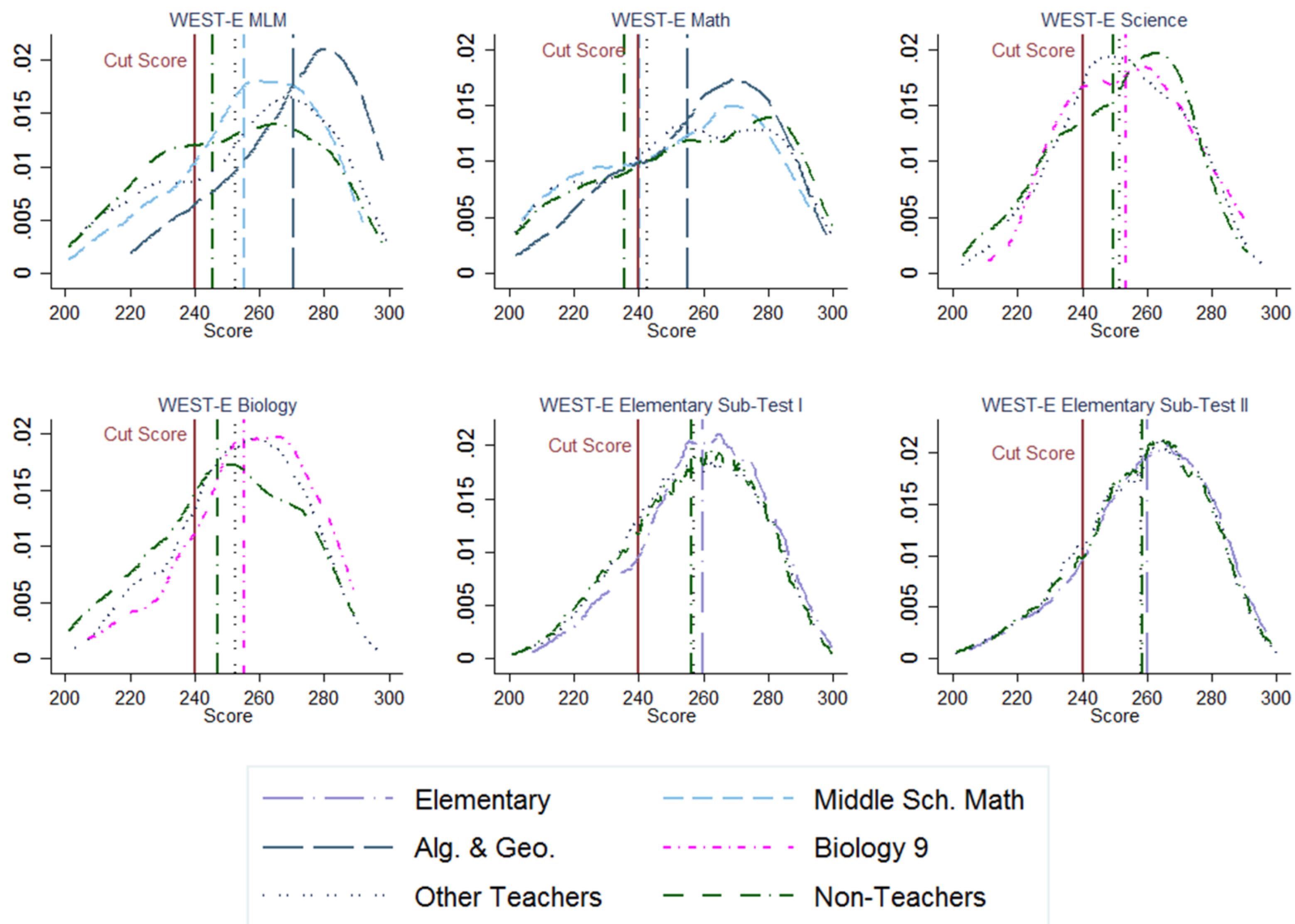


Figure 4. Basic Skills Credential Test Passing Rates by Subtest and State

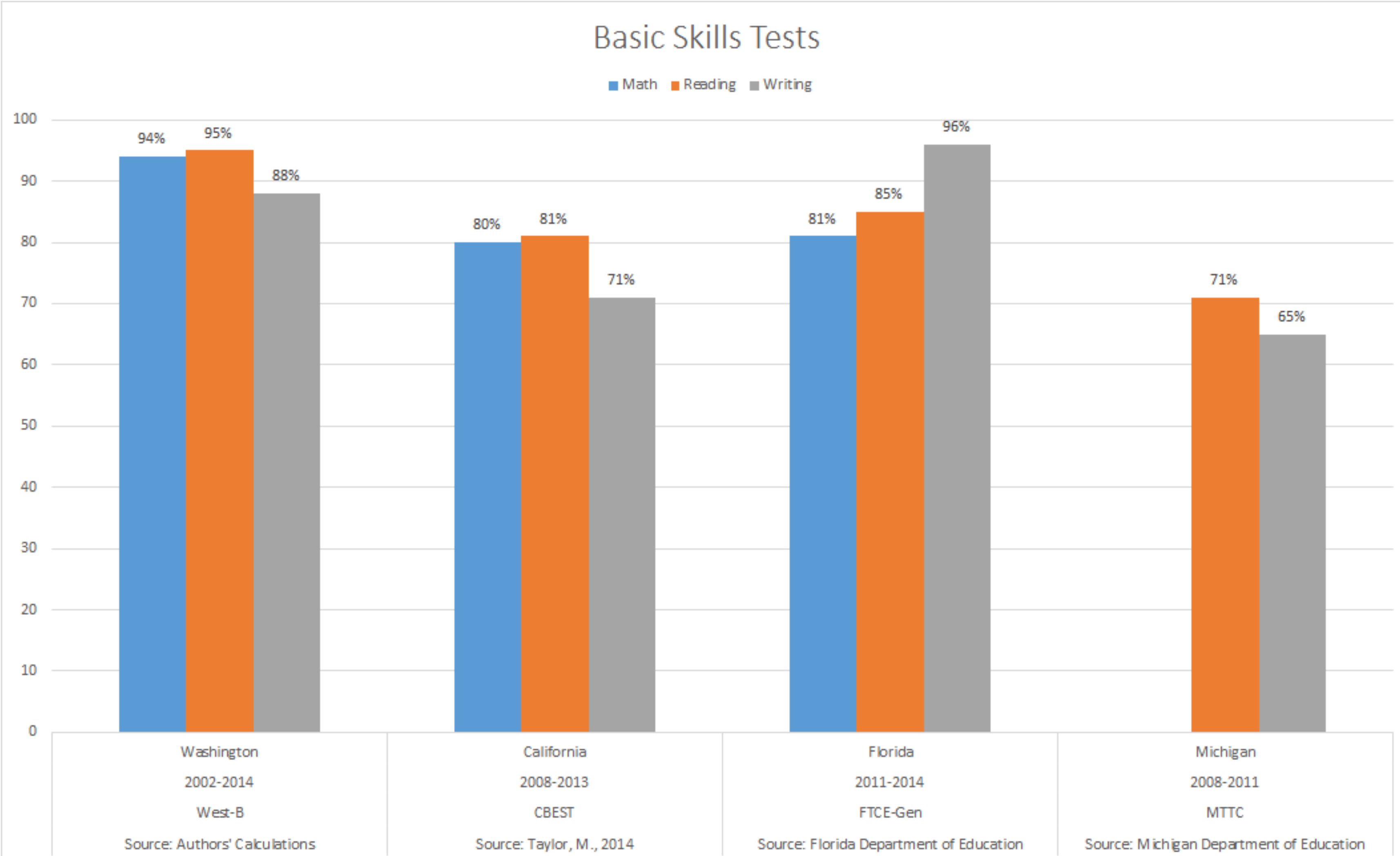

Basic Skills Tests
Math
Reading
Writing
100
90
80
70
60
50
40
30
20
10
0
94%
95%
88%
80%
81%
71%
81%
85%
96%
71%
65%
Washington
2002-2014
West-B
Source: Authors' Calculations
California
2008-2013
CBEST
Source: Taylor, M., 2014
Florida
2011-2014
FTCE-Gen
Source: Florida Department of Education
Michigan
2008-2011
MTTC
Source: Michigan Department of Education

Figure 5. Subject-Specific Credential Test Passing Rates by Subtest and State

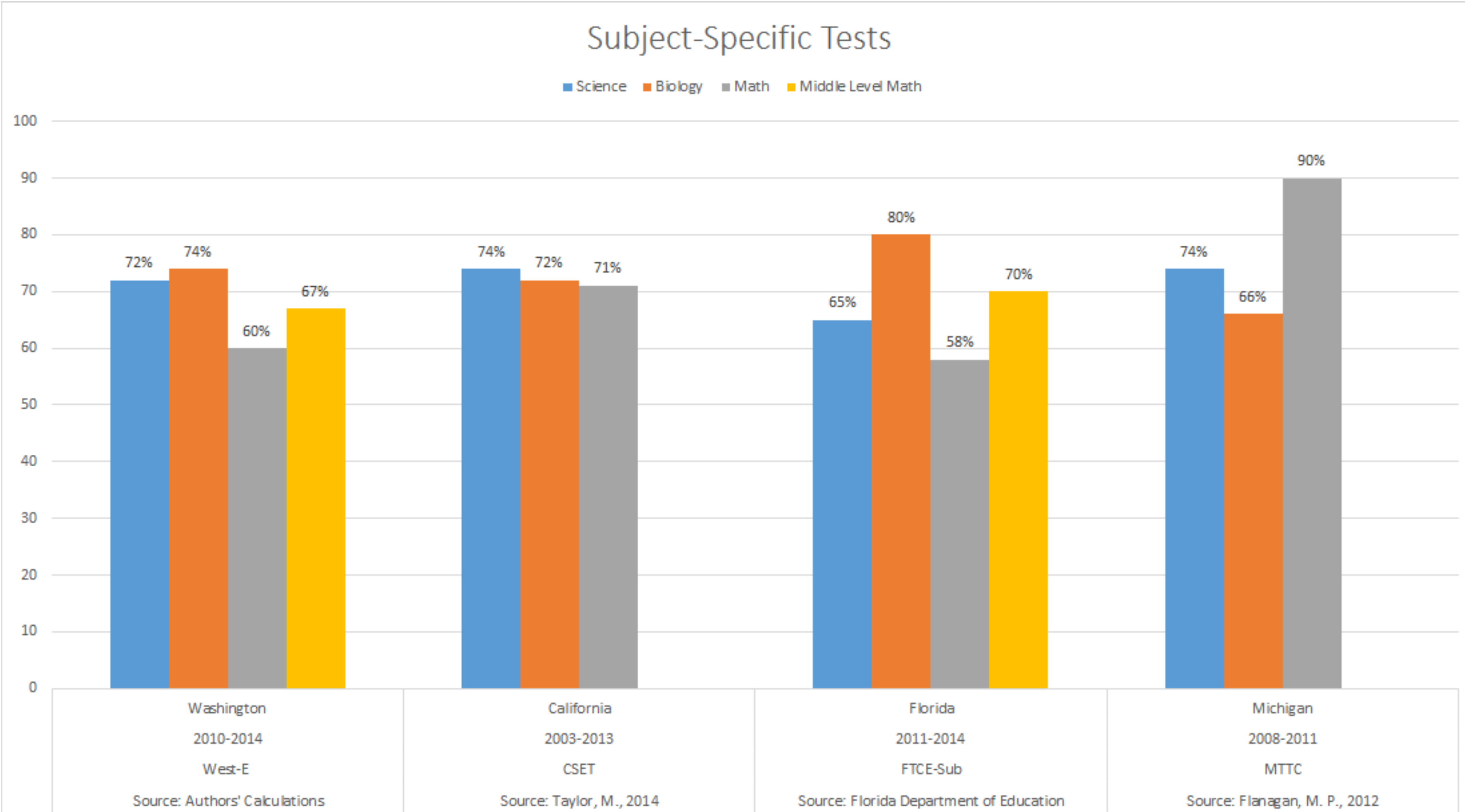


Figure 6. Estimated Effects on Student Achievement by Sample and Quartile of WEST-B Math Score

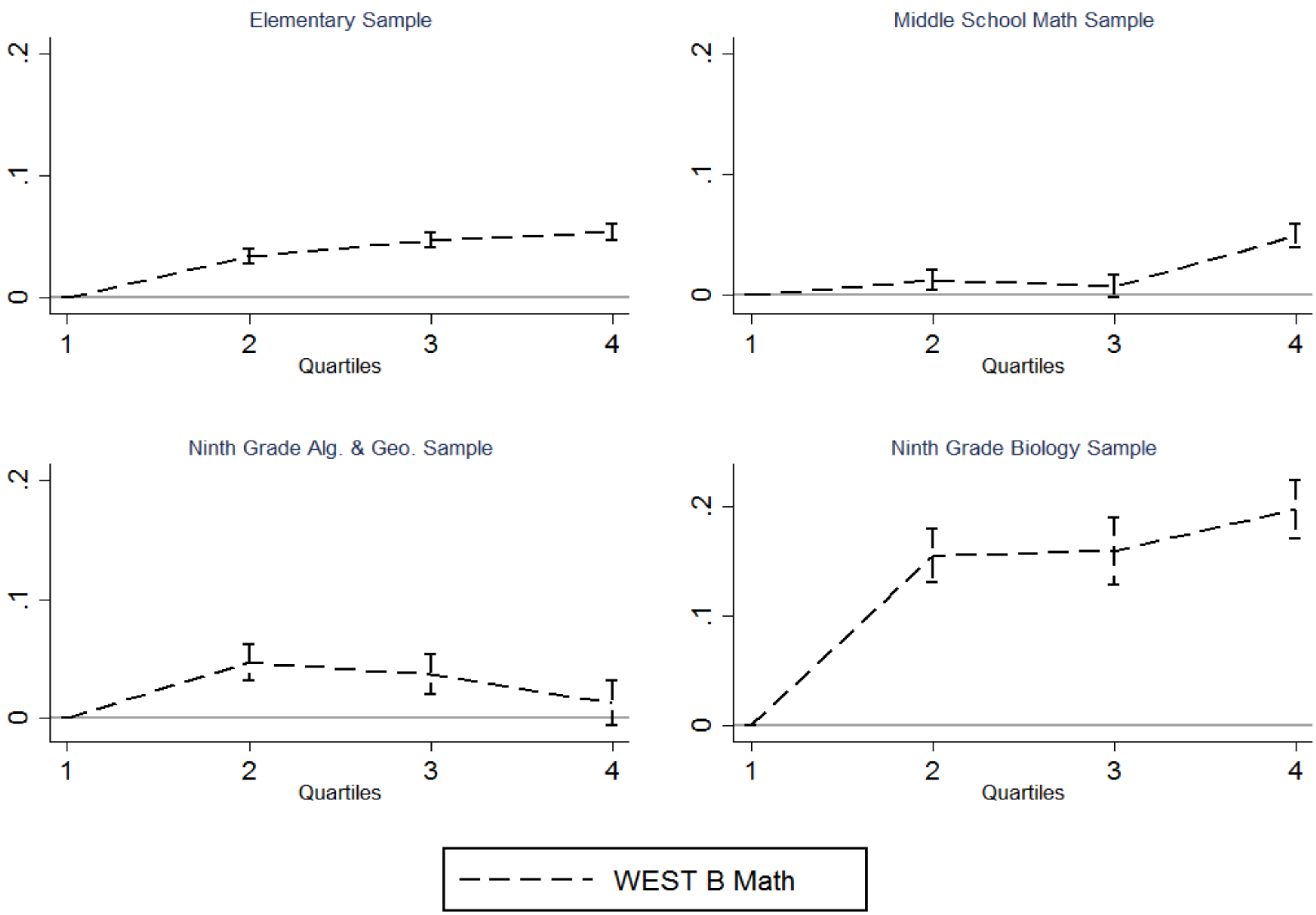
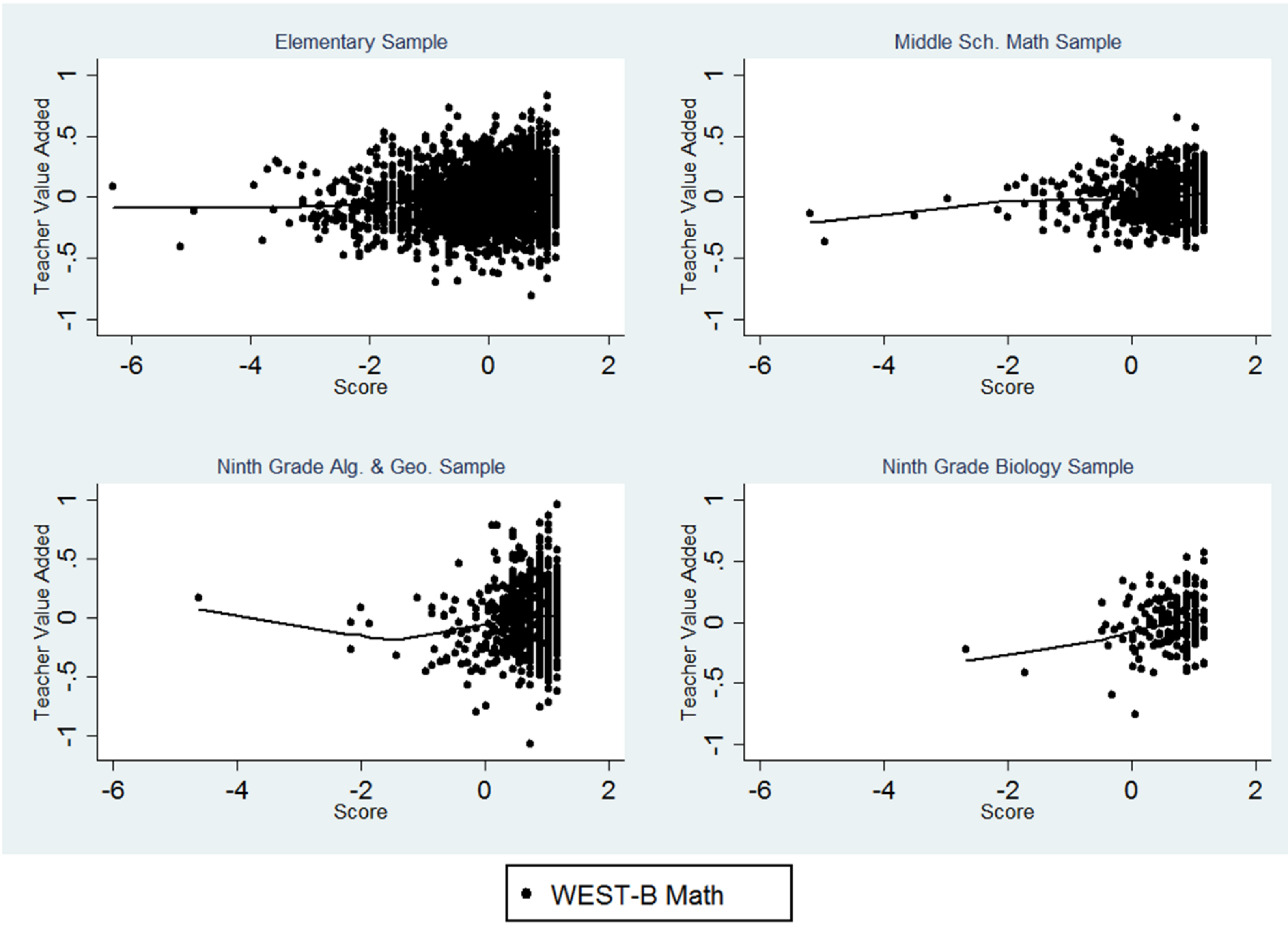


Figure 7. WEST-B Scores and Estimated Teacher Value-Added



Elementary Sample
Middle Sch. Math Sample
Ninth Grade Alg. & Geo. Sample
Ninth Grade Biology Sample
Teacher Value Added
Score
WEST-B Math

Figure 8. Estimated Effects on Student Achievement by Sample, Subtest, and Quartile of WEST-E Score

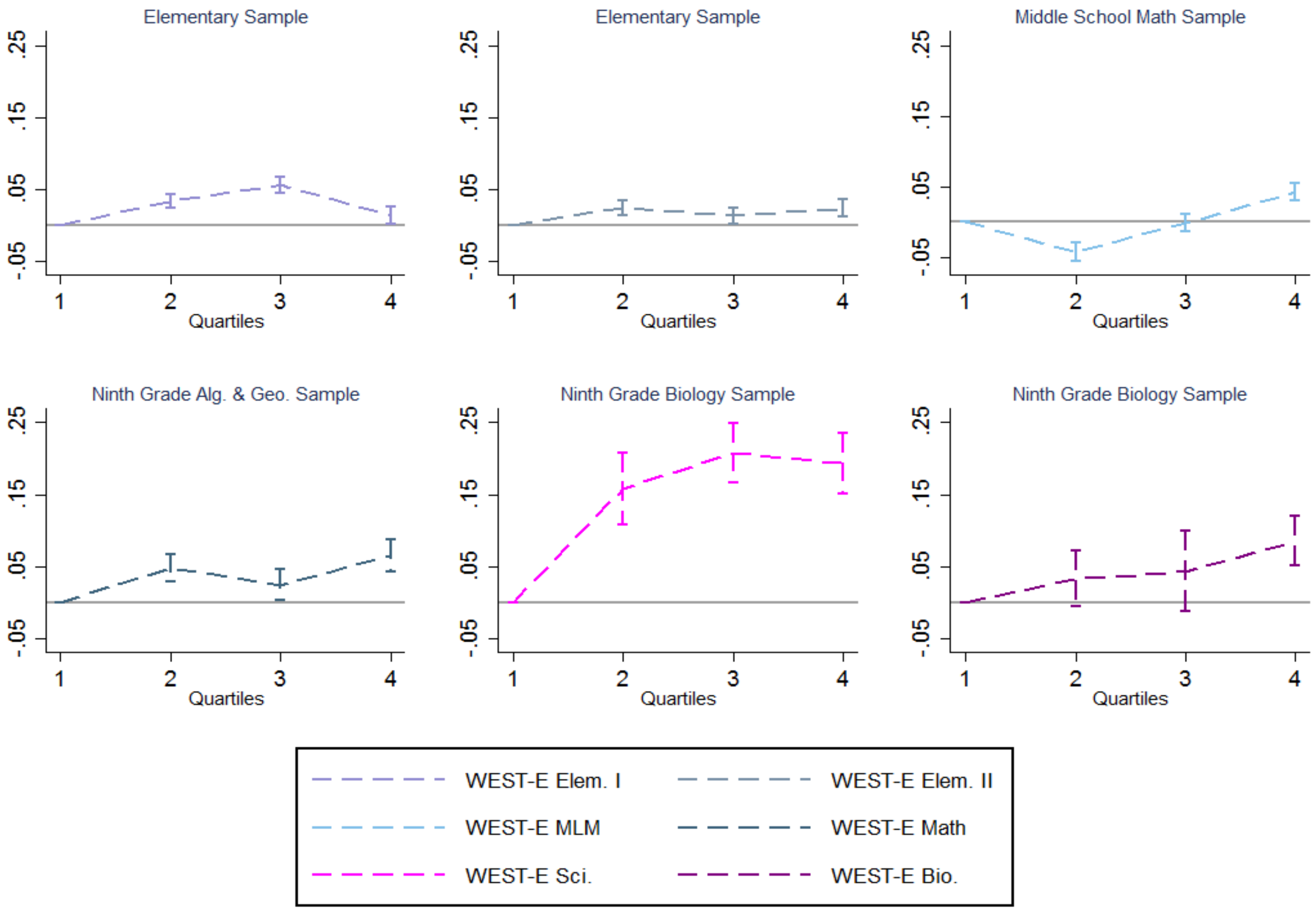
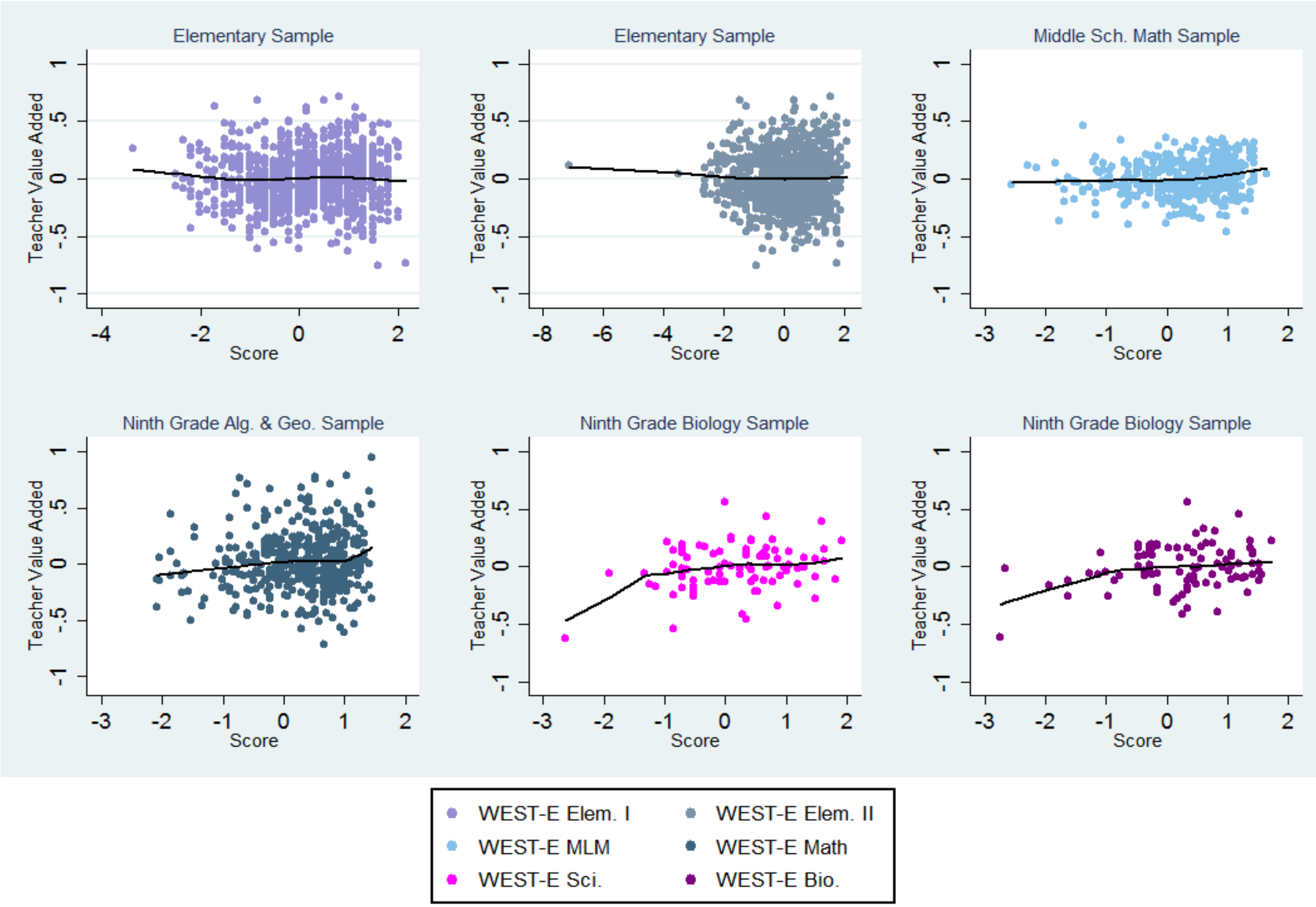


Figure 9. WEST-E Scores and Estimated Teacher Value-Added



Elementary Sample
Elementary Sample
Middle Sch. Math Sample
Ninth Grade Alg. & Geo. Sample
Ninth Grade Biology Sample
Ninth Grade Biology Sample
Teacher Value Added
Score
WEST-E Elem. I
WEST-E Elem. II
WEST-E MLM
WEST-E Math
WEST-E Sci.
WEST-E Bio.