Program on Education Policy and Governance Working Papers Series

Gold Standards?:
State Standards Reform and Student Achievement

Joshua Goodman
John F. Kennedy School of School of Government
Harvard University
Joshua.goodman@hks.harvard.edu

PEPG 12-05

Gold Standards?: State Standards Reform and Student Achievement*

Joshua Goodman
John F. Kennedy School of Government
Harvard University
joshua_goodman@hks.harvard.edu

July 26, 2012

Abstract

Proponents of the recent and widely adopted Common Core State Standards argue that high quality curricular standards are critical to students' educational success. Little clear evidence exists, however, linking the quality of such standards to student achievement. I remedy this by connecting data on state-level student achievement from 1994-2011 with measures of the quality of states' curricular standards as judged by two independent organizations at three different moments in time. I show that, within states, changes in the quality of standards have little impact on overall student achievement. Improved standards do, however, raise achievement of 8th graders in low-scoring states, particularly for low-scoring students. Given the known weaknesses of U.S. middle schools, this result suggests that standards may be beneficial in settings where pedagogy would otherwise be poor.

*For their very helpful comments, I am grateful to Paul Peterson, Martin West and Antonio Wendland, as well as participants in the PEPG July 2012 Conference. Heather Sarsons and Shelby Lin provided outstanding research assistance. All errors are my own.

1 Introduction

Over the last couple of decades, policymakers and educators have increasingly turned to standards-based reform to solve some of the nation's perceived educational challenges. Proponents of standards-based reform argue that students benefit when school systems lay out clear descriptions of what students are expected to know at each phase of their educational development. Most recently, nearly all states in the U.S. have agreed to join the Common Core State Standards Initiative, the goal of which is to better align a currently diverse set of state curricula.

Given the immense amount of time and money being spent on such efforts, it is surprising how little evidence policymakers and educators have on the impact of such standards on student achievement. Research is beginning to shed light on the impact of curriculum on student achievement and later life outcomes (Goodman 2012, Cortes and Nomi 2012). Little is known, however, about how the quality of written standards translates into improvements in curriculum, pedagogy and student achievement. The challenges are twofold. First, it is difficult to find measures of the quality of standards. Second, the quality of standards may be correlated with many other features of an educational system, confounding efforts to isolate the effect of standards themselves.

I remedy this by connecting data on state-level student achievement from 1994-2011 with measures of the quality of states' curricular standards as judged by two independent organizations at three different moments in time. I show that, within states, changes in the quality of standards have little impact on overall student achievement. Improved math standards do, however, raise the math achievement of 8th graders, particulary for low-scoring students. Given the known weaknesses of U.S. middle schools, this result suggests that standards may be beneficial in settings where achievement would otherwise be low.

2 Data and Empirical Strategy

The measures of the quality of the state standards come from two organizations, both of which collected primary source documents from the states and rated the documents on criteria that I describe further below. The first organization, the American Federation of Teachers (AFT), is the

second-largest labor union in the U.S., representing 1.5 million teachers and other education personnel. AFT issued three reports on the quality of state standards, "Making Standards Matter" in 1998 and 2001, and "Sizing Up State Standards" in 2008 (Glidden 2008, of Teachers 2001, of Teachers 2008). In that most recent report, AFT wrote that "common, coherent content standards ensure that all children, regardless of neighborhood, are exposed to rich, well-sequenced content and skills, starting in kindergarten or before" (of Teachers 2008). For each academic subject, AFT rated each state as meeting or failing its criteria for high quality standards at the elementary, middle and high school levels.

The second organization, the Thomas B. Fordham Institute (TFI), is a non-profit education policy think-tank based in Washington, D.C. and Dayton, Ohio with the mission of "advancing educational excellence for every child through quality research, analysis, and commentary, as well as on-the-ground action and advocacy in Ohio" (www.edexcellence.net). TFI issued three reports on the quality of state standards, in 1998, 2005 and 2010, all titled "The State of State Standards" (Finn and Petrilli 1998, Klein 2005, Carmichael and Wilson 2010). In that most recent report, TFI described standards as "the destination: what we want our students to know and be able to do by the end of their K-12 experience, and the benchmarks they should reach along the way" (Carmichael and Wilson 2010). For each academic subject, TFI rated each state's overall standards on an A to F scale.

For each year from 1990 to 2011, I generate for each state both AFT and TFI quality measures for English and math standards. AFT measures are generated from the mean of three indicators for meeting AFT criteria in elementary, middle and high school. TFI measures are generated by converting letter grades to a 4.0 scale and then dividing by four. Both sets of quality measures thus assign 0 to the lowest quality standards and 1 to the highest quality standards. I assign these quality measures based on the year of the state's publication of its most recent standards document. Years preceding the earliest such document are assigned the earliest observed quality measure, while years after the latest such document are assigned the latest observed quality measure.

For the years between the state's earliest and latest publication of a standards document, I assign the quality measure from the most recent AFT or TFI report unless the scoring in the sub-

sequent report indicates that a new standards document was issued on or before that year. For example, Alabama's English standards received a score of four in the 2005 TFI report based on a 1999 state standards document. I therefore assign a score of four for Alabama's 2006 standards. The 2010 TFI report reviews, however, a 2007 Alabama English standards document and assigns a score of three. I thus assign a three to Alabama for 2007 through 2010. If, as occurs in a small number of cases, the same standards document was reviewed in subsequent reports but assigned a different quality measure, the new measure is assigned to the report year and the relevant following years. To account for the fact that each organization's grading standards may have changed between reports, I also record for each state, year and subject the year of the AFT or TFI report from which the quality measure is derived.

Figures 1 and 2 show the mean AFT and TFI measures of standards quality in a given year for math and English respectively. The mean is taken as an unweighted average across all states, though the use of weights changes none of the overall patterns seen here. The data spans the years 1994-2011 and vertical lines are placed at years in which AFT or TFI issued its report.

Panel (A) of figure 1 shows that, according to AFT, the average quality of state standards in math was a remarkably high 80% in 1994. This rose slightly by the second report in 2001, then declined substantially by the third report in 2008. This decline may reflect real changes in the quality of standards or may indicate that the authors of the 2008 report were harsher judges of state standards. As described further below, my regression analysis will attempt to account for such potential changes in judgment between reports. Panel (B) of figure 1 shows that TFI assessed math standards more harshly than AFT, with the average quality under 40% in 1994. This remained relatively steady by the second report in 2005, then rose to over 50% by the third report in 2010.

Comparison of the two panels in figure 1 reveals that AFT and TFI differed greatly both on the level of quality they reported and on the changes over time in such quality. Such differences are less apparent for the English standards, as seen in panels (A) and (B) of figure 2. Both AFT and TFI see improvements in such standards after 1998 and both see some subsequent decline in later years, though the magnitude of that decline is much larger according to AFT. These two figures

suggest relatively little agreement between the two organizations over how to measure the quality of standards, a fact I formalize in the results section below.

Outcomes data come from the National Assessment of Educational Progress (NAEP), a set of exams in math and reading given to representative samples of public school 4th and 8th graders in each state. The NAEP is administered by the National Center for Education Statistics within the U.S. Department of Education. Scores are available for roughly every two years starting in the early 1990s through 2011, the most recent assessment. I use each state's mean score, as well as its 10th, 25th, 50th, 75th and 90th percentile scores. I standardize each score by the student-level national mean and standard deviation of that assessment, so that results can be interpreted as student-level standard deviation impacts. In some specifications I characterize each state by the mean of its 4th and 8th grade scores, while in other specifications I explore scores separately by grade.

Prior studies of the relationship between state standards and student achievement have relied on cross-sectional comparisons that using measures of standards quality captured at one moment in time (Whitehurst 2009). Such estimates are likely confounded by other state-level factors correlated with both standards quality and achievement, such as teacher quality or educational culture. I improve on this by comparing states to themselves over time, estimating how within-state changes in standards quality are related to within-state changes in student achievement.

I do so by running fixed effects regressions of the following form:

$$Score_{st} = \beta_0 + \beta_1 Quality_{st} + \lambda_s + \mu_t + ReportYear_{st} + \epsilon_{st} \quad (1)$$

where $Score$ is a test score for state s in year t and $Quality$ is the AFT or TFI measure of the quality of standards in that state and year. State fixed effects λ control for any factors that are constant within a given state over time and thus implement the within-state comparison at the heart of this analysis. Year fixed effects μ control for any factors that are constant within a given year across all states, such as national trends in standards quality or achievement. To account for the fact that AFT and TFI may have changed their grading standards over time, I also include $ReportYear$, a set of indicators for the year of the AFT or TFI report from which the quality measure was taken.

Standard errors ϵ are clustered by state to account for serial correlation in the error term. For these regressions, I standardize each quality measure within year. Reported estimates can thus be interpreted as the impact of a one (state-level) standard deviation improvement in standards quality on the average student's achievement.

3 Results

Before exploring the impact of standards quality on achievement, I first explore the extent to which the quality measures generated by AFT and TFI relate to each other. Table 1 regresses TFI's quality measures on AFT's quality measures. Column (1) includes no additional controls. If the two organizations' measures were identical, the regression coefficients would have a magnitude of one, given that the two are both standardized to have the same scale. Instead, the coefficients for math and English have a magnitude of less than 0.3. Though statistically significant, these coefficients suggest that the two organizations agree only somewhat on which states have the highest quality standards in any given year. Even more striking is column (2), in which the addition of state and year fixed effects reduces both coefficients to magnitudes close zero and statistical insignificance. This suggests that the two organizations disagree almost entirely over the magnitude of within-state changes in standards quality. Overall, these results confirms what figures 1 and 2 suggested, namely that the two organizations are measuring fairly different aspects of state standards. I therefore report all subsequent results separately for the two organizations.

Table 2 shows the relationship between standards quality and student achievement for both subjects and both organizations' quality measures separately. The outcome used here is the average of 4th and 8th grade test scores in the given subject. Column (1) is OLS with no controls. The negative and statistically significant coefficients suggest that states with one standard deviation higher quality standards have test scores that are 0.06 standard deviations lower. This small and negative relationship is unlikely to be causal given the many factors other than standards quality that vary between states.

Column (2) controls for these other factors by including state fixed effects, so that the estimated coefficients relate within-state changes in standards quality to within-state changes in student

achievement. Inclusion of such fixed effects reveals that within-state changes in standards quality have no statistically significant relationship to student achievement. These results are unchanged by inclusion of year fixed effects and fixed effects for the publication year of the AFT or TFI report from which the quality measure was derived. In math and reading, the 95% confidence interval rules out positive impacts of a one standard deviation improvement in standards quality of more than 0.02 standard deviations. These results provide little evidence of an overall relationship between standards quality and achievement for American 4th and 8th graders.

One possible explanation for the lack of an observed relationship in table 2 is that significant time may pass between the publication of state standards and the subsequent changes in pedagogy that might improve student achievement. Columns (5)-(7) explore this by relating student achievement to lagged measures of standards quality. Each replicates column (4) but lags the quality measures by 1-3 years, allowing such newly adopted standards time to take effect. The lagged results are nearly identical to the contemporaneous ones, suggesting little evidence that the lack of observed relationship is due to the time it takes standards to translate into classroom practice.

Students of different ages and skill levels may be differently affected by the quality of state standards. Tables 3 and 4 analyze 4th and 8th graders separately, exploring impacts on both mean achievement and other percentiles of the achievement distribution. I also divide the sample into low- and high-scoring states, defined by a state being above or below the median state's test score in 2003, the first year that all states were administered the NAEP. I interact these indicators with standards quality to see whether improvements in such quality have differential impacts depending on the initial achievement level of the state as a whole.

Table 3 shows little evidence of a relation between standards quality and achievement for 4th graders at any point in the skill distribution and in any type of state. Table 4 suggests, however, that 8th graders in low-scoring states do benefit from improved standards. In such states, according to AFT's quality measure, a one standard deviation improvement in math standards leads to an improvement of 0.027 standard deviations in the mean student's math achievement. This impact is felt across the achievement distribution, but is twice as large for the lowest scoring students as it is for the highest scoring students. There is also marginally significant evidence that a one

standard deviation improvement in TFI's measure of English standards quality leads to a 0.026 standard deviation improvements in English scores. This effect is significant and pronounced for those at the low end of the skill distribution and diminishes the further up the distribution the student lies.

Tables 5 and 6 show the impact of standards quality on the test scores of demographic subgroups of students, as reported by NAEP. Columns (1) and (2) show the test scores of non-poor and poor students, where poverty is determined by participation in the federal school lunch program. Columns (3) and (4) shows the test scores of white and minority students respectively, where minority scores are the mean of available scores for black and Hispanic students. Fourth grade test scores show little discernible movement in response to changed standards for any of the subgroups in question. Low-scoring states' eighth grade test scores react particularly strongly to improve standards in math for black and Hispanic students and in English for poor students. Given existing correlations between poverty, race and academic achievement, the estimates here are thus consistent with the prior finding that standards quality matters most for low-scoring students in low-scoring states.

4 Conclusion

Given the current transition by nearly all U.S. states to adopt Common Core State Standards, researchers and policymakers should be thinking quite carefully about the role that standards play in influencing student achievement. The results presented in this paper suggest that, over the last couple of decades, changes in the quality of state standards have had little impact on overall student achievement.

There are two possible explanations for this lack of observed relationship. The first is that such a relationship does exist but that our existing measures of standards quality are too poor to detect it in the data. This is plausible given that the two organizations studied here themselves seemed to have little agreement over which states were improving standards over time. This suggests that policymakers, educators and researchers should think more carefully about clearly defining quality when it comes to educational standards.

The second possibility is that high quality standards do not ultimately translate into the pedagogical changes necessary to influence student achievement. Teachers and administrators may be unaware of or unresponsive to the standards, which suggests the need for better communication between schools and the states. Or schools may use their own better standards, which suggests that state intervention may be unnecessary in many cases. This possibility suggests that educators should figure out why improved standards do not ultimately impact classroom performance in measurably beneficial ways.

The clearest positive result presented here is that high quality standards raise achievement of eighth graders in low-scoring states, particularly for the lowest-scoring students. Recent research has highlighted the extent to which the transition to middle school is remarkably damaging to student achievement, particularly in mathematics (Schwerdt and West 2011). Standards may therefore have an important role to play in settings where pedagogy would otherwise be poor and students would be struggling.

References

Carmichael, S.B., M. G. P.-M. K. and W. Wilson (2010). The state of state standards and the common core in 2010. Technical report, The Fordham Foundation.

Cortes, K., G. J. and T. Nomi (2012). Doubling up: Intensive math education and educational attainment. Technical report, Harvard Kennedy School.

Finn, C.E., J. L. and M. Petrilli (1998). The state of state standards. Technical report, The Fordham Foundation.

Glidden, H. (2008). Making standards matter 1998. an annual fifty-state report on efforts to raise american standards. Technical report, American Federation of Teachers.

Goodman, J. (2012). The labor of division: Labor market returns to compulsory math coursework. Technical report, Harvard Kennedy School.

Klein, D. (2005). The state of state math standards. Technical report, The Fordham Foundation.

of Teachers, A. F. (2001). Making standards matter 2001. a fifty-state report on efforts to implement a standards-based system. Technical report, American Federation of Teachers.

of Teachers, A. F. (2008). Sizing up state standards 2008. Technical report, American Federation of Teachers.

Schwerdt, G. and M. West (2011). The impact of alternative grade configuratons on student outcomes through middle and high school. Working Paper 11-02, Program on Policy and Governance.

Whitehurst, G. (2009). *Don't forget curriculum*. Brookings Institution.

A Appendix

Below are excerpts from the 1998 AFT and TFI reports describing the criteria by which those organizations judged the quality of state standards.

A.1 AFT math and English criteria

1. Standards must define in every grade, or for selected clusters of grades, the common content and skills students should learn in each of the core subjects. No matter how clear and specific standards may be, if they do not indicate the various grades or levels at which students are expected to master specific material, they are not very useful. A document that merely states what is to be accomplished by the end of schooling is not very helpful for ensuring a common core curriculum in early and middle grades. Nor can it provide sufficient guidance to curriculum designers or test developers so that teachers know if their students are on track for meeting the standards at the end of their schooling.

Documents that simply repeat the same standard from cluster to cluster or grade to grade are nearly as ineffective as those with no grade breakdowns because they do not indicate the development expected of students as they move from grade to grade. Standards that are the same from grade to grade or cluster to cluster but assert "student work will reflect a grade-appropriate level of quality and complexity," without defining "grade-appropriate" in any of the documents, are also judged to be inadequate. Strong standards should show how knowledge and skills build over the years by clearly defining the specific expectations of progress or development for each grade or grade cluster. Otherwise, experience tells us that teachers, parents, students, and curriculum and assessment developers are likely to interpret "grade-appropriate" differently, jeopardizing the implementation of a common core curriculum.

2. Standards must be detailed, explicit, and firmly rooted in the content of the subject areas to lead to a common core curriculum. Strong standards must provide clear guidance to teachers, curriculum and assessment developers, textbook publishers, and others, so that one person's interpretation of the core knowledge and skills students should learn in a particular grade level or cluster of grades won't be very different from someone else's. If the standards are unclear, the curriculum across schools and districts can vary widely, and the integrity of any assessments based on the standards may be compromised. Teachers, students, parents, and others will be left to guess the academic content and expectations for mastery; and if they guess wrong, student achievement will suffer.

In this report, we do not attempt to judge the overall quality or rigor of the content covered in each state's subject-matter standards. We do not try to determine, for example, whether the ninth-grade algebra standards in a given state contain the most salient content for ninth graders. But, the content must be defined. It is not enough for standards to emphasize the skills students should learn, but leave the content to local discretion. For instance, a standard that asks students to "edit their own work to reflect correct grammar and mechanics" is inadequate according to our criteria. What level of grammar and mechanics is expected at different levels? The grammar expected from a fourth grader is different and less sophisticated than the grammar expected of an eighth grader. The standards should reflect this difference. It is also not enough to make a laundry list of concepts and skills in order to "cover" everything. That approach will result in an unmanageable and often fragmented set of expectations that fails to define the content most important for students to learn.

3. For each of the four core curriculum areas, particular content must be present. In our 1996 and 1997 reports, we highlighted obvious "holes" or weaknesses in each subject–for example, a lack of history in the social studies standards. This year, we are more explicit about particular content that must be present in each of the four subject-matter areas. We identified that content by reviewing numerous documents and reports to determine where there was consensus on the content that all students should learn in each subject-matter area. Appendix C lists the materials reviewed, which include the national subject-matter standards documents, the National Assessment of Educational Progress (NAEP) frameworks, and the TIMSS framework. Having conducted the reviews, we concluded that even if standards documents were clear and specific, they would be judged insufficient if they did not include the following content at each education level:

English: The consensus in the documents on English language arts indicates that English standards should address the basic skills and knowledge that are the foundations of learning how to read (e.g., letter-sound recognition, decoding skills, vocabulary), reading comprehension (e.g., exposure to a variety of literary genres), writing conventions (e.g., spelling, writing mechanics), and writing forms (e.g., narrative, persuasive, expository). In laying out these standards, it is important for a state to indicate in which grades or clusters key elements will be taught.

Math: Based on the math documents reviewed, math standards should include number sense and operations, measurement, geometry, data analysis and probability, and algebra and functions at each level. It is necessary for the standards to provide guidance on the specific mathematical concepts students should learn at each level.

4. Standards must provide attention to both content and skills. It is not enough for standards to emphasize the skills students should learn but leave the content to local discretion. It is also not enough for standards to emphasize subject knowledge with no discussion of the skills needed to apply that knowledge. Skills isolated from content, and context or content items isolated from applications, are meaningless and impossible to teach or assess. To lead a common core of learning across the state, it is imperative that the standards pursue process and application skills through the specific content of the subject areas.

For example, it is not enough for standards to simply name the "U.S. Revolutionary War" but provide no elaboration. Do students need to know the dates of the Revolutionary War, or should they analyze its causes and effects? Without some guidance on what students should be able to do with the knowledge, the quality and complexity of the student work will differ substantially across the state. Also, curriculum designers and assessment developers will be forced to make their own determination of what content to teach and how to asses students' understanding. Some students may be grossly unprepared for the tests through no fault of their own or their teachers, because the standards were not clear about the application skills students needed to be able to do.

A.2 TFI English criteria

A. Purpose, audience, expectations, and assumptions of the standards document(s) 1. The document is written in clear English prose, for the general public as well as for educators. 2. It assumes that English is the language to be used in English language-arts classes, and the only language to be used. 3. It expects all students to demonstrate use of standard English, orally and in writing. 4. It acknowledges the existence of a corpus of literary works called American literature, however diverse its origins and the social groups it portrays. 5. It expects students to become literate American citizens. 6. It expects explicit and systematic instruction in decoding skills in the primary grades as well as the use of meaningful reading materials. 7. It expects students to do

regular independent reading through the grades, suggesting how much reading students should do per year as a minimum, with some guidance about its quality. 8. It expects the standards to serve as the basis for clear and reliable statewide assessments.

B. Organization of the standards 1. They are presented grade by grade or in clusters of no more than 3 to 4 grade levels. 2. They are grouped in categories reflecting coherent bodies of scholarship or research in the English language arts. 3. They distinguish higher-order knowledge and skills from lower-order skills, if lower-level skills are mentioned.

C. Disciplinary coverage of the standards 1. The standards clearly address listening and speaking. They include use of various discussion purposes and roles, how to participate in discussion, desirable qualities in formal speaking, and use of established as well as peer-generated or personal criteria for evaluating formal and informal speech. 2. The standards clearly address reading (and viewing) to understand and use information through the grades. They include progressive development of reading skills and a reading vocabulary, and knowledge and use of a variety of textual features, genres, and reading strategies for academic, occupational, and civic purposes. 3. The standards clearly address the reading (or viewing), interpretation, and critical evaluation of literature. They include knowledge of diverse literary elements and genres, different kinds of literary responses, and use of a variety of interpretive and critical lenses. They also specify those key authors, works, and literary traditions in American literature and in the literary and civic heritage of English-speaking people that all students should study because of their literary quality and cultural significance. 4. The standards clearly address writing for communication and personal expression. They require familiarity with writing processes, established as well as peer-generated or personal evaluation criteria, and various rhetorical elements, strategies, genres, and modes of organization. 5. The standards clearly address oral and written language conventions. They require the use of standard English conventions for sentence structure, spelling, usage, penmanship, capitalization, and punctuation. 6. The standards clearly address the nature, dynamics, and history of the English language. They cover the nature of its vocabulary, its structure (grammar), the evolution of its oral and written forms, and the distinction between the variability of its oral forms and the relative permanence of its written form today. 7. The standards clearly address research processes, including developing questions and locating, understanding, evaluating, synthesizing, and using various sources of information for reading, writing, and speaking assignments. These sources include dictionaries, thesauruses, other reference materials, observations of empirical phenomena, interviews with informants, and computer data bases.

D. Quality of the standards 1. They are clear. 2. They are specific 3. They are measurable (i.e., they can lead to observable, comparable results across students and schools). 4. They are comprehensive. 5. They are demanding: a. They are of increasing intellectual difficulty at each higher educational level and cover all important indices of learning in the area they address. b. They index or illustrate growth through the grades for reading by referring to specific reading levels or to titles of specific literary or academic works as examples of a reading level. c. They illustrate growth through the grades for writing with writing samples. d. For other subdisciplines, they provide examples of specific reading, writing, or oral language features, activities, or assignments that clarify what is expected for each standard or benchmark. 6. Their overall contents are sufficiently specific, comprehensive, and demanding to lead to a common core of high academic expectations for all students in the state, no matter what school they attend.

E. Anti-Literary or Anti-Academic Requirements or Expectations: Negative Criteria 1. The document implies that the literary or popular culture of our or any other country is monolithic in

nature. 2. The reading/literature standards require students to relate what they read to their lived experiences.23 3. The reading/literature standards want reading materials to address contemporary social issues. 4. The document implies that all literary and nonliterary texts are susceptible of an infinite number of interpretations and that all points of view or interpretations are equally valid regardless of the logic, accuracy, and adequacy of the supporting evidence. 5. The examples of classroom activities or student writing offered are politically slanted or reflect an attempt to manipulate students' feelings, thinking, or behavior. 6. The standards teach moral or social dogma. 7. The document explicitly or implicitly recommends one instructional approach for all teachers to follow.

A.3 TFI math criteria

I. Clarity: the success the document has in achieving its own purpose. A. The words and sentences themselves must be understandable, syntactically unambiguous, and without needless jargon. B. What the language says should be mathematically and pedagogically definite, leaving no doubt of what the inner and outer boundaries are, of what is being asked of the student or teacher. C. Testability of the lessons as described.

II. Content: is the state asking K-12 instruction in mathematics to contain the right things, and in the right amount and pacing? A. Adequacy of Primary school content (K-6, approximately) B. Adequacy of Middle school content (grades 7-9, approximately) C. Adequacy of Secondary school content (grades 10-12, approximately)

III. Mathematical Reasoning: do the standards as a whole and throughout demand attention to the structural organization by which the parts of mathematics are connected to each other?

IV. Negative Qualities: the presence of unfortunate features of the document that injure its intent or alienate the reader to no good purpose or, if taken seriously, will tend to cause that reader to deviate from what otherwise good, clear advice the document contains.

A. False Doctrine: Demands in the standards that are injurious to the correct transmission of mathematical information, including: excessive reliance on calculators, excessive emphases on "real-world problems," the fashionable notion that a mathematical question may have a multitude of different valid answers, as well as the occurrence of plain mathematical error. B. Inflation: Bloated or pretentious prose, repetitiousness, evidence of mathematical ignorance, bureaucratic jargon, empty pronouncements, and other irrelevancies.

Figure 1: Mean Quality of Math Standards by Year

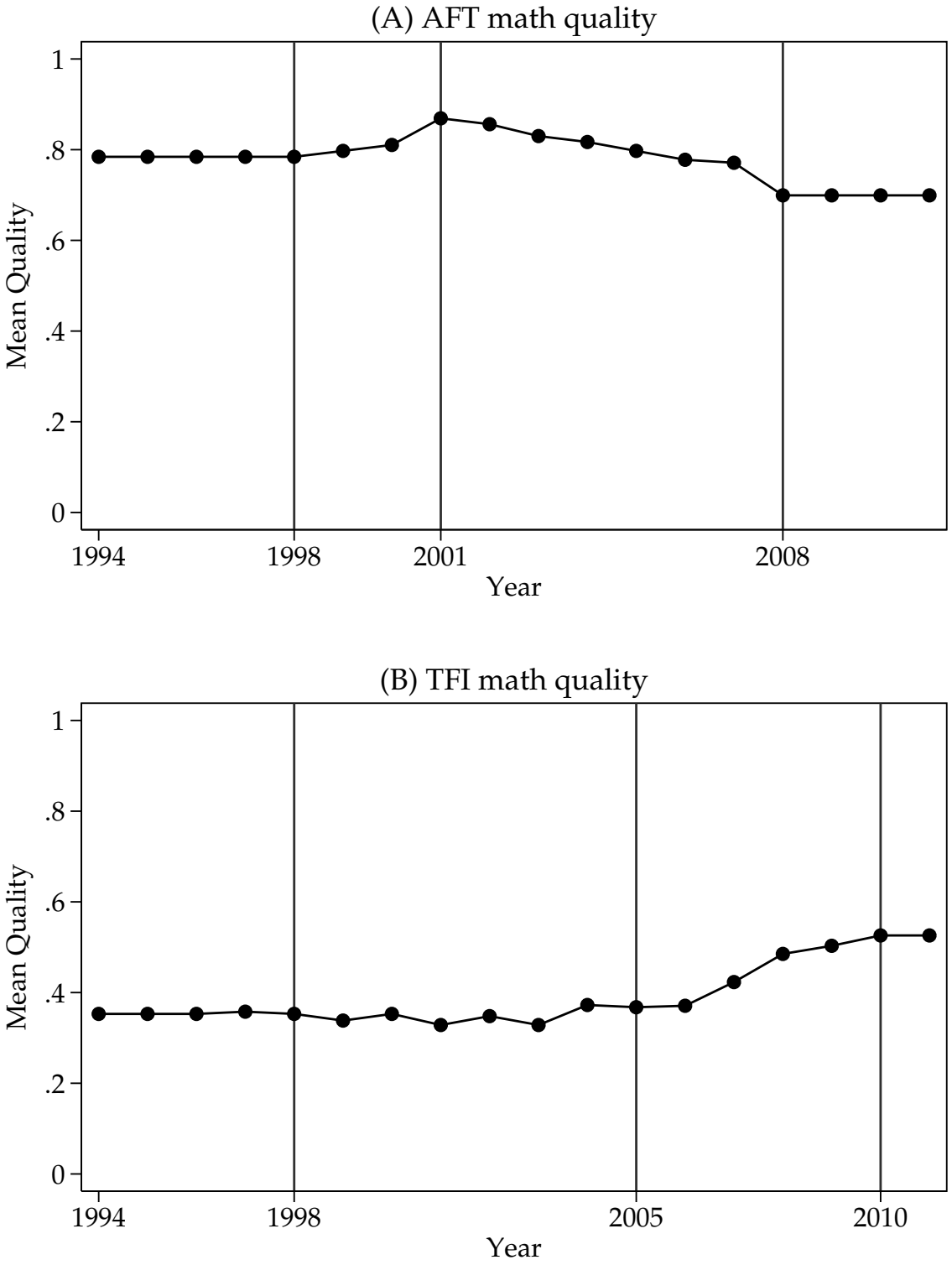

Figure 2: Mean Quality of English Standards by Year

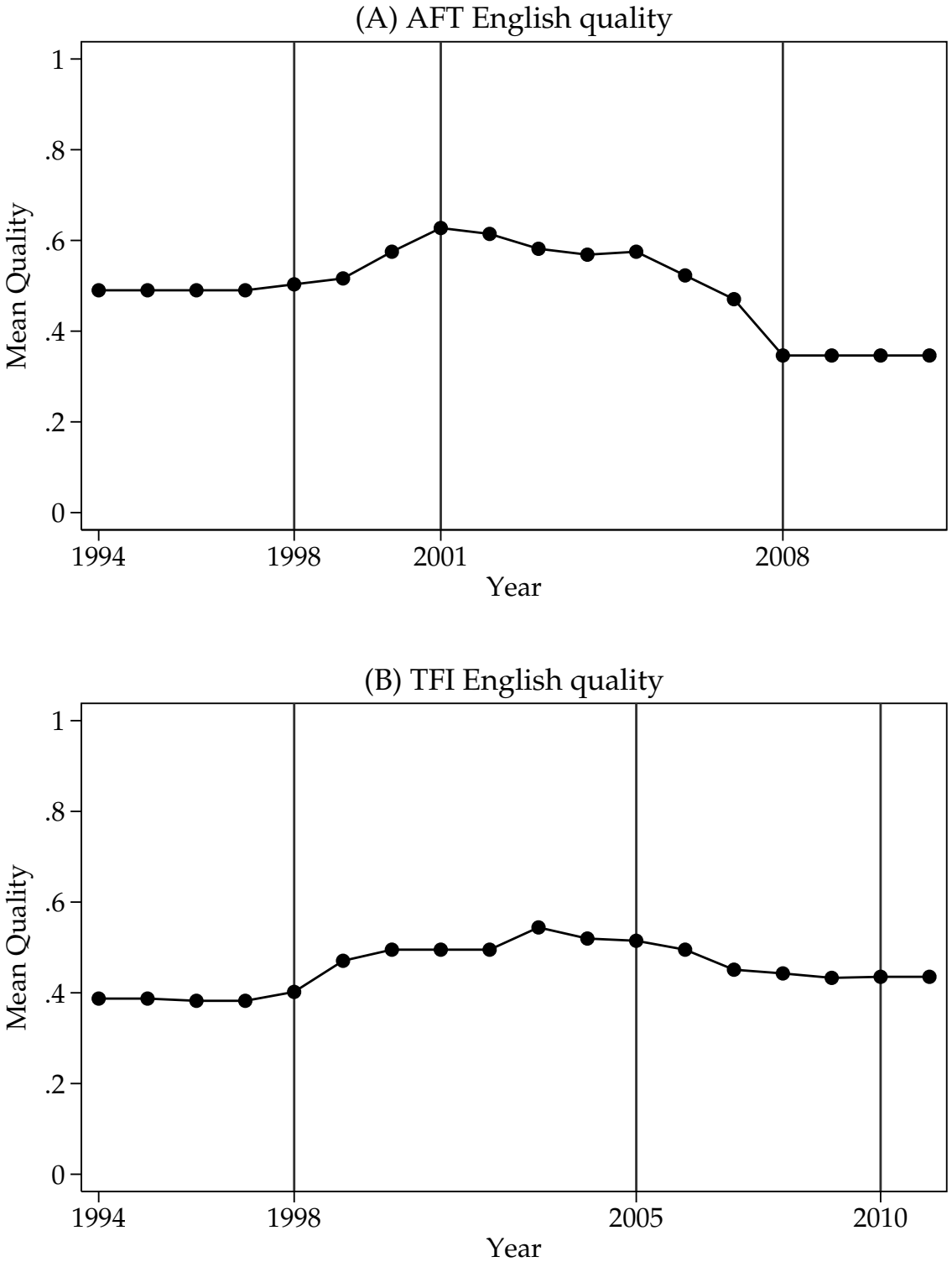

Table 1: AFT's vs. TFI's Quality Measures

	(1) OLS, no controls	(2) State and year fixed effects
(A) TFI math quality		
AFT math quality	0.297***	0.073
	(0.076)	(0.090)
R^2	0.088	0.554
N	918	918
(B) TFI English quality		
AFT English quality	0.271***	0.126
	(0.092)	(0.099)
R^2	0.074	0.595
N	918	918

Notes: Heteroskedasticity robust standard errors clustered by state are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient comes from a separate regression. Column (1) includes no controls. Column (2) includes state and year fixed effects.

Table 2: Overall Effects of State Standards Quality

	(1) OLS, no controls	(2) State fixed effects	(3) State and year fixed effects	(4) Report year fixed effects	(5) One year lag	(6) Two year lag	(7) Three year lag
(A) Math							
AFT math quality	-0.057***	0.001	0.003	0.004	0.004	0.003	0.003
	(0.019)	(0.008)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)
R^2	0.059	0.931	0.942	0.942	0.942	0.942	0.942
N	343	343	343	343	343	343	343
TFI math quality	-0.053**	-0.002	-0.001	-0.001	0.007	0.006	0.010
	(0.026)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.013)
R^2	0.051	0.931	0.942	0.942	0.942	0.942	0.943
N	343	343	343	343	343	343	343
(B) Reading							
AFT English quality	-0.055**	-0.009	-0.009	-0.009	-0.007	-0.007	-0.006
	(0.025)	(0.008)	(0.008)	(0.009)	(0.008)	(0.007)	(0.006)
R^2	0.073	0.923	0.924	0.924	0.924	0.924	0.924
N	379	379	379	379	379	379	379
TFI English quality	-0.060**	0.006	0.006	0.008	0.008	0.008	0.009
	(0.029)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
R^2	0.091	0.923	0.924	0.924	0.924	0.924	0.924
N	379	379	379	379	379	379	379

Notes: Heteroskedasticity robust standard errors clustered by state are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient comes from separate regressions using the average of 4th and 8th grade test scores as outcomes. Column (1) includes no controls. Column (2) includes state fixed effects. Column (3) includes state and year fixed effects. Column (4) also includes fixed effects for the publication year of the AFT or TFI report from which the quality measure was derived. Columns (5)-(7) replicate column (4) but use lagged standards quality instead of the contemporaneous measure.

Table 3: Distributional Effects of State Standards Quality, Grade 4

	(1) Mean score	(2) 10th %ile	(3) 25th %ile	(4) 50th %ile	(5) 75th %ile	(6) 90th %ile
(A) Math, 4th grade						
AFT math quality * high state	-0.013	-0.036*	-0.017	-0.006	0.001	-0.003
	(0.008)	(0.018)	(0.013)	(0.007)	(0.007)	(0.010)
AFT math quality * low state	0.004	0.006	0.003	-0.000	0.002	-0.000
	(0.012)	(0.015)	(0.014)	(0.011)	(0.011)	(0.011)
N	337	337	337	337	337	337
TFI math quality * high state	-0.015	-0.017	-0.014	-0.014	-0.012	-0.009
	(0.012)	(0.016)	(0.014)	(0.013)	(0.012)	(0.012)
TFI math quality * low state	0.014	-0.000	0.006	0.015	0.020	0.022
	(0.020)	(0.019)	(0.020)	(0.020)	(0.023)	(0.023)
N	337	337	337	337	337	337
(B) Reading, 4th grade						
AFT English quality * high state	-0.011	-0.007	-0.007	-0.011	-0.011	-0.013
	(0.013)	(0.022)	(0.015)	(0.011)	(0.009)	(0.008)
AFT English quality * low state	-0.007	-0.013	-0.005	0.002	-0.002	-0.002
	(0.014)	(0.018)	(0.014)	(0.012)	(0.013)	(0.013)
N	379	379	379	379	379	379
TFI English quality * high state	0.004	-0.002	0.004	0.003	0.005	0.008
	(0.010)	(0.014)	(0.011)	(0.010)	(0.009)	(0.009)
TFI English quality * low state	0.014	0.041	0.021	0.011	0.005	-0.001
	(0.023)	(0.039)	(0.026)	(0.017)	(0.014)	(0.012)
N	379	379	379	379	379	379

Notes: Heteroskedasticity robust standard errors clustered by state are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient comes from separate regressions using 4th grade test scores as outcomes. Column (1) replicates column (3) from table 2, interacting standards quality with indicators for high- and low-scoring states. Columns (2)-(6) use identical specifications but with quantiles of the test score distribution as outcomes.

Table 4: Distributional Effects of State Standards Quality, Grade 8

	(1) Mean score	(2) 10th %ile	(3) 25th %ile	(4) 50th %ile	(5) 75th %ile	(6) 90th %ile
(A) Math, 8th grade						
AFT math quality * high state	0.005	-0.001	0.004	0.005	0.004	0.009
	(0.011)	(0.017)	(0.013)	(0.010)	(0.011)	(0.011)
AFT math quality * low state	0.027**	0.029	0.028*	0.026**	0.021**	0.015
	(0.012)	(0.019)	(0.014)	(0.011)	(0.010)	(0.010)
N	336	336	336	336	336	336
TFI math quality * high state	0.003	0.004	0.002	0.001	0.002	0.001
	(0.012)	(0.013)	(0.012)	(0.012)	(0.013)	(0.014)
TFI math quality * low state	-0.010	-0.014	-0.013	-0.012	-0.006	-0.009
	(0.019)	(0.023)	(0.022)	(0.019)	(0.018)	(0.017)
N	336	336	336	336	336	336
(B) Reading, 8th grade						
AFT English quality * high state	0.000	-0.002	0.003	0.005	0.007	0.004
	(0.011)	(0.016)	(0.013)	(0.011)	(0.011)	(0.011)
AFT English quality * low state	-0.009	-0.015	-0.018	-0.011	-0.011	-0.006
	(0.015)	(0.026)	(0.019)	(0.013)	(0.013)	(0.011)
N	334	334	334	334	334	334
TFI English quality * high state	-0.003	-0.008	-0.002	0.000	0.002	0.002
	(0.007)	(0.011)	(0.008)	(0.007)	(0.007)	(0.008)
TFI English quality * low state	0.026*	0.052**	0.034**	0.023**	0.012	0.008
	(0.013)	(0.024)	(0.015)	(0.011)	(0.011)	(0.011)
N	334	334	334	334	334	334

Notes: Heteroskedasticity robust standard errors clustered by state are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient comes from separate regressions using 8th grade test scores as outcomes. Column (1) replicates column (3) from table 2, interacting standards quality with indicators for high- and low-scoring states. Columns (2)-(6) use identical specifications but with quantiles of the test score distribution as outcomes.

Table 5: Subgroup Effects of State Standards Quality, Grade 4

	(1) Non-poor students	(2) Poor students	(3) White students	(4) Black/Hisp. students
(A) Math, 4th grade				
AFT math quality * high state	0.001	0.003	-0.000	-0.016
	(0.010)	(0.017)	(0.011)	(0.017)
AFT math quality * low state	0.001	0.003	0.000	0.013
	(0.013)	(0.012)	(0.010)	(0.020)
N	340	340	340	320
TFI math quality * high state	-0.009	-0.026**	-0.015	-0.010
	(0.012)	(0.010)	(0.014)	(0.011)
TFI math quality * low state	0.017	0.015	0.005	0.006
	(0.017)	(0.017)	(0.012)	(0.022)
N	340	340	340	320
(B) Reading, 4th grade				
AFT English quality * high state	-0.014	-0.006	-0.018	-0.010
	(0.009)	(0.018)	(0.011)	(0.020)
AFT English quality * low state	-0.003	-0.006	-0.018	0.001
	(0.017)	(0.015)	(0.013)	(0.025)
N	339	339	379	354
TFI English quality * high state	0.008	-0.009	0.011	-0.008
	(0.009)	(0.014)	(0.010)	(0.016)
TFI English quality * low state	0.008	0.012	0.011	0.008
	(0.015)	(0.026)	(0.016)	(0.030)
N	339	339	379	354

Notes: Heteroskedasticity robust standard errors clustered by state are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient comes from separate regressions using 4th grade test scores for the listed subgroup as outcomes. Column (1) replicates column (3) from table 2, interacting standards quality with indicators for high- and low-scoring states. Column (4) uses the average of black and Hispanic subscores as the outcome.

Table 6: Subgroup Effects of State Standards Quality, Grade 8

	(1) Non-poor students	(2) Poor students	(3) White students	(4) Black/Hisp. students
(A) Math, 8th grade				
AFT math quality * high state	0.004	0.006	0.001	0.009
	(0.010)	(0.017)	(0.009)	(0.023)
AFT math quality * low state	0.027**	0.026*	0.015	0.042**
	(0.013)	(0.015)	(0.011)	(0.018)
N	336	336	332	305
TFI math quality * high state	0.001	-0.007	0.003	-0.009
	(0.011)	(0.014)	(0.010)	(0.015)
TFI math quality * low state	-0.010	-0.014	-0.025	-0.007
	(0.016)	(0.019)	(0.018)	(0.020)
N	336	336	332	305
(B) Reading, 8th grade				
AFT English quality * high state	0.005	-0.002	0.000	0.018
	(0.010)	(0.013)	(0.010)	(0.015)
AFT English quality * low state	-0.006	-0.009	-0.020	-0.008
	(0.014)	(0.018)	(0.015)	(0.019)
N	334	334	329	304
TFI English quality * high state	-0.001	0.000	0.000	-0.003
	(0.006)	(0.009)	(0.007)	(0.012)
TFI English quality * low state	0.013	0.031**	0.009	0.022
	(0.011)	(0.015)	(0.013)	(0.016)
N	334	334	329	304

Notes: Heteroskedasticity robust standard errors clustered by state are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient comes from separate regressions using 8th grade test scores for the listed subgroup as outcomes. Column (1) replicates column (3) from table 2, interacting standards quality with indicators for high- and low-scoring states. Column (4) uses the average of black and Hispanic subscores as the outcome.