

Teaching Practices and Cognitive Skills[☆]

Jan Bietenbeck*

CEMFI, Casado del Alisal 5, 28014 Madrid, Spain

Abstract

National Teaching Standards by various educational organizations in the United States call for a decrease in the use of traditional teaching practices (such as learning by rote) and an increase in the use of modern teaching practices (such as working in small groups) in schools. Yet a small literature in economics has consistently found that traditional teaching raises test scores, while the effect of modern teaching appears to be small and sometimes even negative. This paper uses data from the Trends in International Mathematics and Science Study (TIMSS) to show that traditional and modern teaching practices promote different cognitive skills in students. In particular, traditional teaching practices increase students' factual knowledge and their competency in solving routine problems, but have no significant effect on their reasoning skills. The effects of modern teaching practices are exactly the opposite, with modern teaching fostering reasoning skills. I provide evidence that standardized tests do not measure reasoning skills well, which explains the finding of only small or negative effects of modern teaching on test scores in the literature. I discuss the implications of these results for the recommendations made by National Teaching Standards.

Keywords: teaching practices, cognitive skills, standardized tests, TIMSS

JEL classification: I21

1. Introduction

Researchers, teachers, and parents have long debated which teaching practices are best for student learning in schools. Traditionally, teachers have relied on lecturing and repetitive practice in order to teach students basic facts and procedures. Several reform movements during the twentieth century attempted to introduce a more student-centered approach to teaching into schools, in which small group work and discussion among students were supposed to take center stage. Despite these efforts, traditional teaching practices still dominated in American classrooms by the year 1990 (Cuban, 1993). Since then, however, student-centered teaching has gained considerable support

[☆]Final version, March 2014. I am greatly indebted to David Dorn for his constant guidance and support. I thank David Autor, Manuel Bagues, Samuel Bentolila, Stéphane Bonhomme, Michael Greenstone, Richard Murnane, Jesse Rothstein, seminar participants at CEMFI, MIT, the SOLE meetings in Boston, and the EALE conference in Turin, the editor, Michele Pellizzari, and two anonymous referees for their valuable suggestions which substantially improved the paper. Funding from the Spanish Ministry of Science and Innovation (ECO2010-16726) and from the UAM-Accenture Chair in Economics and Management of Innovation (Research Award 2011) is gratefully acknowledged.

*Email: bietenbeck@cemfi.es

with the release and implementation of National Teaching Standards by various educational organizations (e.g., [NCTM, 1989, 1991](#); [NRC, 1996](#)). These call for a shift from traditional towards modern, student-centered teaching in schools as a way to promote students' reasoning skills over mere factual knowledge and routine problem-solving skills.¹ This emphasis on reasoning skills is motivated by the perception that such skills are becoming increasingly important in the labor market.

A small literature in economics has examined the effects of teaching practices on student outcomes. These studies find sizable positive impacts of traditional teaching practices, such as lecturing and rote memorization, on test scores ([Lavy, 2011](#); [Schwerdt and Wuppermann, 2011](#)). In contrast, estimates of the effects of modern teaching practices, such as working in small groups and emphasizing real-life applications, are comparatively small ([Lavy, 2011](#)) and sometimes even negative ([Murnane and Phillips, 1981](#); [Goldhaber and Brewer, 1997](#)). The existing empirical evidence therefore seems to suggest that a decreased emphasis on traditional teaching and an increased emphasis on modern teaching in schools will lower student test scores. Does this mean that National Teaching Standards are wrong in recommending such a change?

In this paper, I explore a more nuanced interpretation of these results. My starting hypothesis is that traditional and modern teaching practices promote different cognitive skills in students. In particular, I claim that just as National Teaching Standards posit, modern teaching raises students' reasoning skills. However, these skills are not measured well in standardized tests. In contrast, traditional teaching fosters the knowledge of basic facts and procedures that has historically been emphasized in schools and that is primarily assessed in standardized tests. Such heterogeneity of the effects of traditional and modern teaching practices across cognitive skills, if it exists, could explain the sizable positive impact of traditional teaching and the smaller or negative effect of modern teaching on test scores found in the literature, and it would change the way in which these results have to be interpreted with regards to National Teaching Standards.

I test the hypothesis that traditional and modern teaching practices promote different cognitive skills using data from the 2007 wave of the Trends in International Mathematics and Science Study (TIMSS) for the United States. The data contains test scores for eighth-grade students' overall achievement in math and science as well as sub-scores measuring performance on three segments of the test which assessed distinct cognitive skills. One of these skills is reasoning, and the other two are factual knowledge and competency in solving routine problems. The data also includes information on teaching practices from a student questionnaire, which asked students to rate how often they engaged in a range of different classroom activities in a particular subject. Referring to National Teaching Standards, I classify activities as reflecting either a traditional or a

¹National Teaching Standards categorize teaching practices as “to be decreased” or “to be increased.” In line with the previous literature (e.g., [Lavy, 2011](#); [Schwerdt and Wuppermann, 2011](#)), I adopt the terminology “traditional” and “modern” teaching practices here.

modern teaching practice, and I use information on the frequency of these activities to define two class-level indices for use of modern and use of traditional teaching practices.

I begin my analysis of the effects of teaching practices on cognitive skills by relating the traditional and modern teaching indices to students' overall test scores in math and science. The empirical model exploits the fact that each student is observed twice in the data - once in math, and once in science - in order to include student fixed effects. This means that the impacts of teaching practices on test scores are identified from the variation of teaching practices between the two subjects for each student. The student fixed effects net out most potential confounding factors, such as the sorting of students to teachers and teachers' adjustment of teaching practices to their students' academic abilities. Moreover, the inclusion of a rich set of teacher-level control variables in the regression model mitigates the concern that overall teacher quality, rather than the teaching practices themselves, is driving the results. In line with the previous literature, my results show that traditional teaching has a positive and significant effect on students' overall math and science test scores, while the impact of modern teaching is close to zero and not statistically significant.

I then estimate separately the effects of traditional and modern teaching practices on each of the three cognitive skills for which sub-scores are available in the data. There is a positive impact of the traditional teaching index on students' factual knowledge and on their competency in solving routine problems, but no significant effect on students' reasoning skills. Conversely, the impact of the modern teaching index on students' factual knowledge and on their routine problem-solving skills is close to zero, while its estimated impact on reasoning is positive and significant. This positive effect of modern teaching on reasoning skills is masked in the overall test score regression because standardized tests, both in TIMSS and elsewhere, contain relatively few questions measuring these skills. Taken together, the results are in line with my initial hypothesis, and they suggest that an increased emphasis on modern teaching practices and a decreased emphasis on traditional teaching practices will lower students' overall test scores but promote their reasoning skills, the latter of which is the stated aim of National Teaching Standards.

This paper makes three important contributions relative to the small previous literature on the effects of teaching practices on student outcomes. First, it provides the first comprehensive analysis of the impacts of traditional and modern teaching practices on student test scores for the United States.² The only prior study that defines teaching

²Most previous studies based on United States data have focused either on teaching practices that would be classified as modern by National Teaching Standards (Murnane and Phillips, 1981; Goldhaber and Brewer, 1997) or on practices that do not fit into the framework of traditional versus modern teaching (Kane et al., 2011; some practices in Murnane and Phillips, 1981). Schwerdt and Wuppermann (2011) compare the impact of lecturing - an unambiguously traditional practice - with that of in-class problem solving. While they regard the latter as a modern teaching practice, this is not necessarily how National Teaching Standards would classify it (see the discussion in footnote 10). In a cross-national context, Algan et al. (2013) examine the impacts of traditional and modern teaching practices on students' social capital.

practices in a similar way as this paper does is the analysis by [Lavy \(2011\)](#) for Israel. Like in this current paper, the author uses information from a student survey on the frequency of use of a range of classroom activities in order to define two class-level indices of traditional and modern teaching. [Lavy \(2011\)](#) finds that both traditional and modern teaching are positively related to student achievement, but that the impact of traditional teaching is larger. Second, to the best of my knowledge, this is the first study to analyze whether traditional and modern teaching practices affect different cognitive skills in different ways. Third, in an extension of my analysis, I exploit the international dimension of the TIMSS database in order to estimate the effects of traditional and modern teaching practices across a large set of European and Asian countries. The similarity between the estimates from these regressions and those obtained for the United States lends credibility to my headline results and is evidence of their external validity.

The remainder of the paper is organized as follows. Section 2 describes the nature and content of National Teaching Standards in more detail. Section 3 presents the data and discusses the measurement of teaching practices and cognitive skills. Section 4 explains the empirical strategy. Section 5 presents the headline estimates as well as results from several robustness checks. Section 6 extends the analysis to other countries. Section 7 concludes.

2. National Teaching Standards

In its influential 1983 report *A Nation At Risk*, President Ronald Reagan’s National Commission on Excellence in Education painted a grim picture of the state of the education system in the United States. Citing falling SAT scores and disappointing results of American students in international tests, it warned of a “rising tide of mediocrity that threatens our very future as a Nation and a people” ([NCEE, 1983](#), p.5). The perception of the Commission was that the United States were falling behind other nations in terms of economic competitiveness, and that flaws in the education system were one of the principal reasons for this development. Consequently, the report called for large-scale educational reform that would lead to the excellence in education needed for the country to keep its competitive edge in global markets. One of the key elements of such reform was supposed to be an improvement in the quality of teaching in schools.

As a response to *A Nation At Risk*, the National Council of Teachers of Mathematics published the *Curriculum and Evaluation Standards for School Mathematics* ([NCTM, 1989](#)). Commonly referred to as the NCTM Standards, they set out a framework of mathematical skills that students should master at different grade levels. In doing so, they placed a strong emphasis on reasoning skills relative to mere factual knowledge and routine problem-solving skills. This was meant to reflect the demands of the modern labor market, in which “businesses no longer seek workers with [...] ‘shopkeeper’ arithmetic skills [...] [but] the ability to work with others on problems, [and] the ability to see the applicability of mathematical ideas to common and complex problems” ([NCTM,](#)

1989, p.3). According to the NCTM Standards, one of the key reforms that was needed in order to achieve this mathematical literacy based on reasoning was a change in the way in which mathematics was taught in schools. In particular, teachers should increase the use of small group work, student discussions, and real-life applications of mathematics in the classroom. In contrast, they should decrease the use of lecturing, routine problem-solving, and rote memorization of facts and procedures by students.³

The NCTM Standards received considerable attention by policy makers and the media, with the initial reactions being predominantly positive. This led the U.S. Department of Education to commission other professional education bodies to develop similar standards for other school subjects. Among these, the National Research Council's *National Science Education Standards* (NRC, 1996) - the science counterpart of the NCTM Standards - stand out as particularly relevant to this paper. Importantly, there is a wide agreement among standards across all subjects about how teachers should teach (Zemelman et al., 2005). In particular, they have in common the recommendation of a shift from traditional, teacher-centered teaching towards modern, student-centered teaching in order to promote students' reasoning skills.

National Teaching Standards, as the collection of the various subject standards for teaching is often referred to, have had a considerable influence on state curricula as well as on teacher education and professional development. Indeed, during the 1990s, the National Science Foundation provided funding for states and districts to implement math and science reforms that promoted instruction consistent with National Teaching Standards under its Systemic Initiatives program. In-service teacher training, in which teachers were taught to emphasize modern teaching practices in their classrooms, was a key ingredient of many of these reforms (Hamilton et al., 2003). The ensuing introduction of more modern teaching into schools has often been met with criticism. In particular, many stakeholders in education fear that with the focus on student interactions rather than on rote memorization and practice, students will fail to learn basic facts and procedures.⁴ In conclusion, the issue of the relative merits of traditional versus modern teaching practices for raising student learning remains highly controversial.

3. Data

The empirical analysis uses data from TIMSS, an international assessment of the math and science knowledge of fourth- and eighth-grade students. It was first carried out by the International Association for the Evaluation of Educational Achievement (IEA)

³The proposed changes in teaching practices were later outlined in more detail in the *Professional Standards for Teaching Mathematics* (NCTM, 1991). The NCTM Standards have since been updated several times; however, their main recommendations regarding teaching practices remain unchanged.

⁴In mathematics, the implementation of curricula based on the NCTM Standards has led to the so-called "math wars." For an illustrative example of this conflict between supporters of traditional and of modern teaching practices, see Hartocollis (2000).

in 1995 and has since been repeated every four years with a new sample of students. A total of 63 countries participated in TIMSS across the five waves to 2011. In this paper, I focus my attention on the nationally representative sample of United States eighth-grade students assessed in 2007. This is the only wave that contains separate test scores measuring achievement on the three cognitive skill dimensions used in my analysis.⁵

TIMSS collects its data in a two-stage clustered sampling design. Schools are chosen in the first stage, and one or two math classes are randomly sampled within each of these schools in the second stage. All students in the selected classes are administered standardized tests in math and science, and background information is obtained from students and their teachers in both subjects via questionnaires. The sampling design thus implies that all students are observed twice in the data - once in math, and once in science - while teachers are usually observed only with one class. Note that students that are in the same math class do not necessarily attend the same science class. In particular, in 37% of schools in the data, math classes split up into several science classes, which in turn may contain students from other (potentially not sampled) math classes in the same school. This is advantageous in terms of the identification strategy used in this paper because it implies a greater variation in teaching practices across subjects. However, it also means that only a small fraction of the students in a particular science class is observed in the data if this class contains a large number of students from non-sampled math classes. In order to account for this complex sampling design, student sampling weights provided with the TIMSS database are used throughout the analysis.

3.1. Sample Selection

The full sample consists of 7,377 eighth-grade students in 532 math classes and 687 science classes in 239 schools. I exclude from this sample 25 students who cannot be linked to their science teacher as well as 270 students who have more than one teacher in math or in science. Furthermore, I drop 653 students in unusually small or large classes (teacher-reported class size smaller than 10 or greater than 50) because the interactions between teachers and students in these classes are potentially very different from those in a class at the median of the class size distribution (24 students). Finally, as a consequence of the sampling design used by TIMSS, very few students are observed in some of the science classes in the sample. To guarantee a minimum of precision in the measurement of class-level teaching practices based on student reports, I drop 372 students who are observed in classes where less than five students provided information

⁵My focus on eighth-grade students is explained by the fact that the data contains much more detailed information on teaching practices for eighth-grade students than for fourth-grade students. Moreover, fourth-grade students are typically taught by the same teacher in all subjects, which means that the between-teacher and -subject variation in teaching practices used for identification in this paper does not exist for them.

on teaching practices.⁶ The final sample thus consists of 6,057 students in 425 math classes and 462 science classes in 221 schools.⁷

The regressions in the later parts of this paper include as controls a rich set of teacher and class variables drawn from the TIMSS teacher questionnaire. The teacher variables are indicators for female teachers, age ranges 30-39, 40-49, and 50 or more years, having majored in the subject taught, having a postgraduate (Master’s or Ph.D.) degree, having a teaching certificate, and having 2, 3-5, or 6 or more years of teaching experience.⁸ The class variables are class size and teaching time in minutes per week. Table 1 shows the means and standard deviations of these variables. In order not to reduce the sample size any further, missing values in these variables are set to zero, and dummies for missing values for each control variable are included in the regressions. Results are however robust to dropping all observations with a missing value in any of the control variables, which reduces the sample size by 36%. The results from this reduced sample are available upon request.

3.2. Measuring Teaching Practices

I use information from the TIMSS student questionnaire in order to measure teaching practices. The questionnaire asked students to rate on a four-point scale how often they engaged in a range of different activities in the classroom. Importantly, all students responded separately for math and for science, such that a student’s cognitive skills in, say, science can be related to the teaching practices of her teacher in that same subject. This is a notable improvement over the measurement of teaching practices in [Lavy \(2011\)](#), where students indicated which share of their teachers across all subjects employed a particular teaching practice. I assign a value of 0 to the answer “never”, 0.25 to “some lessons”, 0.5 to “about half the lessons”, and 1 to “every or almost every lesson.” These values aim to reflect the frequency scale that is inherent in the answer categories, and they let me interpret students’ responses as the percentage of lessons in which a particular activity was used.

I refer to National Teaching Standards ([NCTM, 1989, 1991](#); [NRC, 1996](#); [Zemelman et al., 2005](#)) in order to select those classroom activities from the student questionnaire that can be categorized as reflecting either a traditional or a modern teaching practice. I should note here that the list of activities that students are asked about differs somewhat between math and science. In this paper, I focus on the three traditional and three modern teaching practices that are available for both subjects. This ensures that results

⁶The response rate to the teaching practice questions was 93%.

⁷Results are robust to all of the sample restrictions discussed in this paragraph. Estimates for the unrestricted sample are available upon request.

⁸I choose this functional form for teaching experience because prior research ([Rockoff, 2004](#); [Clotfelter et al., 2010](#)) has shown that the positive impact of teaching experience on student test scores is fully accounted for by the first five years, with the first two years explaining the majority of the effect. Results are however robust to alternative functional form assumptions such as a quadratic in experience.

are not driven by mechanical differences in the treatment variables between math and science.⁹ The traditional practices are listening to the teacher lecture, memorizing facts, formulas and procedures, and working problems. The modern practices are working in small groups, giving explanations, and relating what is learned to students' daily lives.¹⁰ Table A1 presents the exact wording of the six teaching practices in the student questionnaire separately for math and for science.

In order to gain precision in the measurement of teaching practices, I aggregate students' answers to the class level as follows. First, for each of the six teaching practices, I calculate the mean of students' answers at the class level while leaving out each student's own answer.¹¹ In a second step, I then take the average of these class-level means across the three traditional and across the three modern teaching practices. The resulting class-level indices of traditional and modern teaching measure the frequency, expressed in percentage of lessons, with which a teacher employs the teaching practices included in them. Table 2 shows the class-level means and standard deviations of the six individual teaching practices and of the traditional and modern teaching indices. It also presents the distribution of student responses across the four answer categories for each teaching practice. The numbers in the table suggest that by 2007, both traditional and modern teaching practices played an important role in American classrooms.

It is important to note that the frequency scale underlying students' answers implies that the two teaching practice indices do not stand in a mechanical trade-off to each other: a ten percent increase in the traditional teaching index does not necessarily imply that the modern teaching index decreases by the same amount. For example, a teacher that uses a variety of traditional and modern teaching practices in all of her lessons will score highly on both of the indices. Indeed, I find that the two indices are weakly positively correlated with a correlation coefficient of 22%.¹² This correlation partly reflects the fact that there is no clear time budget constraint for some of the individual teaching practices considered here. For example, relating the content of the lesson to students' daily lives can be done in a variety of ways, and it is not obvious that this

⁹Results are qualitatively and quantitatively similar if one further modern teaching practice in math (using a variety of solution paths to solve complex problems) and one further traditional teaching practice in science (reading textbooks) that are available in the student questionnaire are included in the analysis. Furthermore, results are robust to controlling for four other classroom activities that do not fit into the broad categories of traditional and modern teaching as defined by National Teaching Standards (reviewing homework, beginning homework in class, having a quiz or test, and using computers).

¹⁰It is not immediately clear whether working problems should be considered a traditional or a modern teaching practice. While National Teaching Standards call for a reduction in the working of routine problems and drill worksheets, they encourage the use of complex problems which require students to reason. I decide to categorize working problems as traditional here based on its relatively high correlation with the other traditional teaching practices (average correlation coefficient of 31%, compared to 5% for the modern teaching practices). In Section 5.3, I show that my results do not depend on the inclusion of this practice in either category. In the remainder of the paper, I will refer to working problems as "working routine problems" so as to make clear that it is considered a traditional teaching practice.

¹¹Results are robust to including each student's own answer in the class-level mean.

¹²Lavy (2011) reports a correlation coefficient of 81% for his modern and traditional teaching indices.

necessarily reduces the time that a teacher can spend on, say, lecturing.

Recall that the interest of this paper is to estimate the effects of traditional and modern teaching practices on cognitive skills. The slight positive correlation between the two teaching practice indices does not prevent this, but it does raise the question how the results should be interpreted. In my preferred specification below, I include the traditional and modern teaching indices as multiple treatments in the same regression. This means that the estimated coefficients reflect the effect of a rise in the traditional (modern) teaching index on test scores, holding constant the modern (traditional) teaching index. This counterfactual corresponds to an increase in the variety of teaching practices that are used during a lesson, or an increase in traditional or modern teaching practices at the expense of such practices as reviewing homework, which is neither traditional nor modern. Section 5 also presents results from an alternative regression setup that uses the traditional teaching index net of the modern teaching index as treatment. These specifications measure the impact of a change in the relative frequency of use of traditional teaching practices compared to modern teaching practices. Results from these regressions are qualitatively similar to those obtained from the headline model.

3.3. *Measuring Cognitive Skills*

The standardized tests in TIMSS assess students' knowledge of the eighth-grade math and science curricula using both multiple-choice and open-response questions. The focus on eighth-grade curriculum knowledge rather than students' overall knowledge of math and science is important here as it ensures that teaching practices during eighth grade can meaningfully influence student achievement on these tests. Two studies by the National Center for Education Statistics (Smith Neidorf et al., 2006a,b) compare the standardized tests in TIMSS to the eighth-grade math and science tests used in the National Assessment of Educational Progress (NAEP), the largest nationally representative assessment of American students. These studies find that the tests from the two assessments are indeed very similar in terms of both content covered and cognitive skills measured. This confirms the validity of using TIMSS test scores for measuring eighth-grade curriculum knowledge of students in the United States.

The TIMSS math and science tests are organized around three so-called cognitive domains, which measure distinct cognitive skills. The *knowing* domain focuses on students' ability to recall definitions and facts and to recognize known characteristics, for example shapes of objects in math and tools and materials in science. The *applying* domain measures students' competency in solving routine problems, which will typically have been repetitively practiced in classroom exercises. The *reasoning* domain assesses students' capacity for logical, systematic thinking by confronting them with complex problems set in unfamiliar contexts. Each question on the tests belongs to one of these three domains and gives a certain number of score points if answered correctly. The relative weight of each domain in the tests, as measured by the share of score points

attributed to it, is determined by education experts ahead of the assessment. The shares of score points in the knowing, applying and reasoning domains are 35%, 41%, and 24%, respectively, in math, and 37%, 41%, and 22%, respectively, in science (Ruddock et al., 2008).

From the description in the previous paragraph, it is clear that the knowing and applying domains measure the skills that schools have traditionally focused on, whereas the reasoning domain measures precisely the skill that National Teaching Standards want to promote. Note that questions assessing reasoning skills have the smallest weight, namely less than a quarter, in the computation of the overall math and science scores. Importantly, this low emphasis on reasoning is not an idiosyncratic characteristic of the TIMSS assessment, but a common feature of many standardized tests. Indeed, the share of items measuring reasoning skills in the NAEP eighth-grade math and science tests is very similar to the one reported here (Smith Neidorf et al., 2006a,b). Also, already the National Teaching Standards expressed concerns about the fact that commonly used standardized tests do not measure reasoning skills well. After all, teachers have little incentive to use modern teaching practices and to thus raise students' reasoning skills if this does not translate into higher test scores (see, for example, Chira, 1992).

The TIMSS database contains test scores for students' overall achievement as well as sub-scores for achievement on each of the three cognitive domains for both math and science. Scores are reported in the form of five so-called plausible values, or imputed values. This is due to the fact that like many other large-scale educational assessments, TIMSS uses an incomplete-booklet design for its tests. That is, each individual student in the assessment only completes a subset of items from a larger pool of questions. IEA then applies Item Response Theory to estimate a test score distribution for each student, and the five plausible values are random draws from this distribution. Regressions in the later parts of this paper account for the uncertainty regarding a student's true test score introduced by this design feature (see Section 5.1 for details). All test scores are standardized to have mean zero and standard deviation one in the full sample.

Perhaps not surprisingly, sub-scores for the three cognitive domains are highly correlated at the student level. In particular, the correlation coefficient between the knowing score and the applying score (the reasoning score) is 0.96 (0.93), and the correlation coefficient between the applying score and the reasoning score is 0.93.¹³ A likely explanation for this correlation is that students' academic ability is an important determinant of all three cognitive skills. Notably, the competing explanation that the three cognitive domains measure essentially the same skills is not consistent with the result of heterogeneous impacts of teaching practices across them. Also, recall that the cognitive domains were defined with the purpose of measuring distinct cognitive skills in mind.

¹³The reported correlation coefficients refer to correlations between the means of the five plausible values for each cognitive domain and are averages across math and science.

3.4. *Trends in Teaching Practices in American Classrooms, 1993-2007*

In this subsection, I address the question whether teaching practices in American classrooms have actually changed with the introduction of National Teaching Standards. To the best of my knowledge, the only existing empirical evidence on this issue comes from a study by Smith et al. (2002), which relies on data from a repeated cross-sectional survey of mathematics and science teachers conducted in 1993 and 2000. In the survey, teachers were asked how frequently they employed a range of different teaching practices in their classrooms. Two of these practices are of particular interest here: lecturing, which is a traditional practice according to National Teaching Standards, and working in small groups, which is a modern practice. Smith et al. (2002) find that there was a modest decrease in the frequency of lecturing in science between 1993 and 2000. In contrast, there were no significant changes in the frequency of working in small groups in science, or in the frequency of use of either practice in mathematics.

One possible explanation for the small changes in teaching practices found by Smith et al. (2002) is that the period they consider might be too short to pick up longer-term trends. This is especially true if it takes time for curricula and textbooks to be updated to conform with National Teaching Standards and for teachers to undergo related training programs. The longer time horizon of TIMSS provides me with the opportunity to augment the existing empirical evidence on recent trends in teaching practices. Due to changes in the student questionnaire between the different waves of the study, comparable information on all teaching practices is not available in all years. It is however possible to track the frequency with which one modern teaching practice, working in small groups, was used across the four waves between 1995 and 2007. In 1995, 22% of students reported working in groups in every or almost every lesson, and this frequency rose to 23% in 1999, 26% in 2003, and 28% in 2007.¹⁴ While there is no single traditional teaching practice that appears in the student questionnaire in all of these four waves, the fraction of students that reported listening to the teacher lecture every or almost every lesson declined marginally from 44% to 43% between 2003 and 2007. In sum, while the evidence presented here is not conclusive, it is consistent with the view that there has been a slow shift from traditional towards modern teaching practices in American classrooms following the release of National Teaching Standards.

4. Empirical Strategy

The ideal experiment to estimate the effects of teaching practices on test scores would randomly vary teaching practices across students. There are two reasons why in practice

¹⁴The labeling of the answer categories changed between the 1999 and 2003 waves. The percentages reported here are the fraction of students reporting working in groups “almost always” in 1995 and 1999 and “every or almost every lesson” in 2003 and 2007. The figures are averages across math and science, and the trends in both subjects were similar.

the pairing of students and teaching practices will not be random. First, students sort into schools and classrooms according to their (or their parents’) preferences for particular teaching practices. For example, students with high unobserved academic ability might sort into schools that emphasize modern teaching practices. In this case, any naive estimate of the effect of modern teaching practices on test scores that does not account for this sorting pattern will be biased upward. Second, it is plausible that teachers adjust their teaching practices according to the students they face. If teaching practices are (partially) a function of student-level determinants of test scores that are not controlled for in the regression (such as students’ unobserved academic ability), this will again lead to a bias in the estimated coefficient of interest.

Previous studies have addressed these issues by including student fixed effects in the empirical model. This strategy accounts for the sorting to teaching practices across schools and classrooms based on fixed student characteristics such as academic ability. Moreover, under some assumptions, which are discussed in detail below, student fixed effects also account for the adjustment of a teacher’s teaching practices to her students. In order to include student fixed effects in the empirical model, one needs data that contains multiple observations per student either at different points in time or in different subjects at the same point in time. The TIMSS data with its two observations per student (one in math, and one in science) fulfills this requirement. I exploit this feature of the data and follow the literature in estimating a student fixed-effects model of the impacts of traditional and modern teaching practices on students’ cognitive skills. This means that I identify the effects of interest using the variation of teaching practices between the two subjects for each student.¹⁵

Section 5 presents estimates of the following empirical model:

$$A_{ijs} = \alpha + \beta_1 \text{TradTI}_{ijs} + \beta_2 \text{ModnTI}_{ijs} + \mathbf{X}_{js}\gamma + \lambda_i + \varepsilon_{ijs}, \quad (1)$$

where student i ’s test score in subject s taught by teacher j , A_{ijs} , is determined by the traditional and modern teaching practice indices, TradTI_{ijs} and ModnTI_{ijs} , and by a vector of other teacher and class characteristics, \mathbf{X}_{js} . λ_i is the student fixed effect, which controls for any subject-invariant student-level determinants of test scores, and ε_{ijs} is a student-by-subject specific error term. Note that because students are observed twice in the same school, the student fixed effect at the same time controls for any school characteristics that influence test scores. The following section will first present results from regressions in which A_{ijs} is the overall math or science score. Afterwards, A_{ijs} will then be replaced by the knowing, applying, and reasoning scores in order to estimate separately the effects of traditional and modern teaching practices on student achievement on each of these cognitive skills.

¹⁵Identification based on within-student between-subject variation has been used by [Dee \(2007\)](#), [Clotfelter et al. \(2010\)](#), and [Schwerdt and Wuppermann \(2011\)](#), among others.

The parameters of interest in Equation 1 are β_1 and β_2 . The identifying assumption is that the two teaching practice indices, $TradTI_{ijs}$ and $ModnTI_{ijs}$, are uncorrelated with the error term conditional on the other regressors. One way in which this assumption could be violated is if subject-specific unobservable determinants of student test scores were correlated with the teaching practice indices. That is, the student fixed effects in my model do not account for sorting of students to teaching practices across schools and classrooms based on subject-specific academic ability. While this issue cannot be definitely addressed with the data at hand, the fact that math and science are closely related subjects which arguably require very similar skills mitigates this concern here. Moreover, in a related study that also relies on between-subject variation for identification, [Clotfelter et al. \(2010\)](#) provide suggestive evidence based on tracking patterns that academic ability is indeed highly correlated across subjects. Finally, note that any sorting based on students' overall academic ability is of course accounted for by the estimation strategy.

Another way in which the identifying assumption underlying the model in Equation 1 could be violated is if teachers who emphasize certain teaching practices had particular other unobserved characteristics that promote or hinder students' cognitive skills. For example, it might be the case that highly motivated teachers sort into modern teaching practices. If teacher motivation promotes student test scores via a channel other than teaching practices, this will lead to an upward bias in the estimated coefficient on the modern teaching index. This omitted-variable problem is a challenge faced by virtually all studies that try to identify the effect of a particular teacher trait on student outcomes, and there is usually no definite solution to this issue. In this paper, I partially address this concern by controlling for a rich set of teacher and class characteristics (shown in Table 1).

A final concern regarding the empirical strategy is that in contrast to the teacher characteristics usually studied in the literature, teachers' teaching practices are not a fixed characteristic. Indeed, as already noted above, it is plausible that teachers adjust their teaching practices according to the students they face. In order to discuss the implications of such an adjustment for the identification strategy, it is useful to think of teaching practices as being made up of two parts: a fixed part that varies across teachers but not across classes for a given teacher, and a variable part that depends on student characteristics. Clearly, if the variable part of teaching practices is a function of students' subject-invariant determinants of test scores only, the inclusion of student fixed effects in Equation 1 adequately accounts for any adjustment of teaching practices on the teacher's part. Only if there is subject-specific academic ability at the class level, and if the latter is an input into teaching practices, will the identifying assumption underlying the model be violated.¹⁶ In sum, bias through the channel of teachers' adjustment of

¹⁶Note that at the class level, individual subject-specific abilities of students will net out unless

teaching practices to their students is likely to be minimal due to the inclusion of student fixed effects.¹⁷

5. Results

5.1. Teaching Practices and Overall Student Achievement

Table 3 presents estimates of the model in Equation 1 in which the dependent variable is the overall test score in math and science. All regressions in this and in subsequent tables control for subject fixed effects and are weighted using the student sampling weights supplied with the TIMSS database. Furthermore, all of the estimates account for the uncertainty regarding each student’s true test score introduced by the incomplete-booklet design of the TIMSS tests (see Section 3.3). In particular, regressions are run separately for each possible combination of the five plausible test score values for math and for science, and tables report the mean coefficient estimates as well as the average R -squared from these 25 regressions. Standard errors are adjusted upward for the imputation variance using the formula provided in the TIMSS 2007 Technical Report (Foy et al., 2008) and additionally allow for clustering at the class level.

Table 3 shows a positive and highly significant effect of traditional teaching practices on overall test scores. The estimated coefficient on the traditional teaching index is virtually identical in all of the specifications, regardless of whether it is included in the regression as a single treatment variable (column 1) or together with the modern teaching index in a multiple-treatment model (columns 3-5). The coefficient in my preferred specification in column 5, which includes both teaching practice indices as well as the full set of teacher and class controls, implies that a one standard deviation (ten percent) increase in the traditional teaching index is associated with a 3.2% of a standard deviation increase in the overall test score.

In contrast, the impact of modern teaching practices on overall test scores is comparatively small in all of the specifications in Table 3 and is never statistically significant. The coefficient on the modern teaching index is reduced by one third when the traditional teaching index is included in the regression (column 3) as compared to when the modern teaching index is the only treatment variable in the model (column 2). The additional inclusion of teacher and class controls in columns 4 and 5 has very little impact on the coefficient estimate.¹⁸ Notably, in my preferred specification in column 5, the

students systematically sort into classes based on their subject-specific abilities.

¹⁷Another way to address the endogeneity of teaching practices arising through teachers’ adjustment would be to use a teacher’s teaching practices as observed with other classes as a proxy for her teaching practices in the current class. This is the strategy adopted in Kane et al. (2011). This approach is not feasible here because teachers in the TIMSS data are usually observed only with one class.

¹⁸The fact that the coefficients on both teaching practice indices hardly change with the inclusion of teacher and class controls suggests that there may not be strong relationship between these two sets of variables. I confirmed this in unreported regressions of the traditional and modern teaching indices on each of the observable teacher and class characteristics in Table 1. The estimated coefficients from

coefficient on the modern teaching index is significantly smaller than the coefficient on the traditional teaching index (the last row of Table 3 indicates that the null hypothesis of equal effects can be rejected with a p value of 0.087).

The estimates in Table 3 are in line with the results found in the previous literature. In particular, the positive effect of traditional teaching on overall test scores corroborates similar results by Lavy (2011) and by Schwerdt and Wuppermann (2011). In contrast, the close to zero and insignificant effect of modern teaching lies in between the negative estimates found by Goldhaber and Brewer (1997) and by Murnane and Phillips (1981) and the small positive estimate found by Lavy (2011). Taken at face value, the results in Table 3 therefore imply that an increase in modern teaching and an equally large decrease in traditional teaching will harm student achievement on standardized tests.

5.2. *Teaching Practices and Cognitive Skills*

Having established that the results for overall test scores are similar to those found in the previous literature, I now examine whether the effects of traditional and modern teaching practices are heterogeneous across the three cognitive skills assessed in the tests. Table 4 presents estimates of models in which the dependent variable is the knowing, applying, or reasoning score. All the regressions in this table control for the full set of teacher and class characteristics. In order to facilitate easy comparison, column 1 reproduces the results for the overall test score from column 5 in Table 3.

Columns 2 and 3 of Table 4 present estimates for the knowing and applying score, respectively. Recall that these scores measure the factual knowledge and routine problem-solving skills that have traditionally been emphasized in schools. There is a positive and highly significant effect of traditional teaching practices in both specifications, which is somewhat larger than the impact found for the overall test score in column 1. The coefficient in column 2, for example, implies that a one standard deviation increase in the traditional teaching index is associated with a 4.2% of a standard deviation increase in the knowing score. In contrast, the impact of modern teaching practices on the knowing and applying scores is almost exactly zero, and the hypothesis tests for equal effects of the modern and traditional teaching indices reject the null for both outcomes.

Column 4 of Table 4 presents the estimates for the reasoning score, which measures the skill that National Teaching Standards want to promote. The results in this column stand in stark contrast to those found in the previous two columns. In particular, the estimated coefficient on the traditional teaching index is close to zero and not statistically significant in this regression. Instead, there is a positive and significant effect of modern teaching on reasoning. The coefficient implies that a one standard deviation (eleven percent) increase in the modern teaching index is associated with a 2.4% of a standard

these regressions are generally small and almost never statistically significant. A notable exception is that teacher age is negatively associated with modern teaching, and that teacher experience is positively related to traditional teaching and negatively related to modern teaching.

deviation increase in the reasoning score. Note, however, that the null of equal impacts of traditional and modern teaching cannot be rejected at conventional levels of significance in this specification.

The results in Table 4 are in line with my initial hypothesis that traditional and modern teaching practices promote different cognitive skills in students. The finding that an increase in the traditional teaching index is associated with higher knowing and applying scores makes intuitive sense. After all, the index contains the practices of memorizing facts, formulas, and procedures and working routine problems that are specifically aimed at increasing the factual knowledge and routine problem-solving skills measured by these scores. Instead, the positive effect of modern teaching practices on reasoning skills may be more surprising, especially given their small and insignificant impact on overall test scores. The different coefficients on the modern teaching index in columns 1 and 4 are reconciled by the fact that questions assessing reasoning skills have the smallest weight in the computation of overall test scores in TIMSS. That is, the positive and significant effect of modern teaching on students' reasoning skills is masked by its close to zero impact on factual knowledge and on routine problem-solving skills in the overall test score regression.

What do these results imply for the recommendations of National Teaching Standards? For a start, the findings in Table 4 corroborate the idea that a greater emphasis on modern teaching practices in schools will increase students' reasoning skills, which National Teaching Standards perceive as increasingly important in the labor market. However, because questions aimed at measuring factual knowledge and routine problem-solving skills make up the bulk of today's standardized tests, an emphasis on modern teaching practices at the cost of traditional teaching practices will likely lead to a decrease in overall test scores. Again, this is not because modern teaching practices are generally ineffective at raising student learning; rather, they foster a very specific skill - reasoning - that is not measured well in standardized tests.

5.3. Robustness to Alternative Measures of Teaching Practices

The teaching practice indices used in the empirical analysis so far are not the only way in which teaching practices can be measured based on the information from the TIMSS student questionnaire. Therefore, I now consider several alternative definitions of traditional and modern teaching and show that my results are not sensitive to the choice of measurement. Table 5 presents estimates from the corresponding regressions. Results are reported for the overall test score in column 1, and for the knowing, applying, and reasoning scores in columns 2, 3, and 4, respectively. Descriptive statistics of the treatment variables used in this table can be found in panel A of Table A2.

In panel A of Table 5, I again define two indices of traditional and modern teaching at the class level. However, rather than assigning values to each answer category and computing the average of these values among all students in a given class, for each teaching practice I now compute the share of students that respond engaging in it "about

half the lessons” or “every or almost every lesson.”¹⁹ The traditional (modern) teaching index is then defined as the average of these class-level shares across all traditional (modern) teaching practices. Panel A shows that results from specifications that use these alternative indices as treatment variables are qualitatively and quantitatively similar to the ones obtained in Table 4.

In panel B of Table 5, I define the gap between traditional and modern teaching as the difference between the original traditional and modern teaching indices. The gap can loosely be interpreted as the share of lessons in which traditional but not modern teaching practices are used (if the gap is negative, it reflects the share of lessons in which modern but not traditional teaching practices are used). It is therefore a measure of the relative emphasis on traditional as compared to modern teaching practices in a given class. Results from specifications that use the gap between traditional and modern teaching as a treatment variable are again qualitatively similar to the ones obtained in Table 4. In particular, the coefficient estimates in panel B imply that a greater relative emphasis on traditional teaching increases knowing and applying scores, while a greater relative emphasis on modern teaching increases reasoning scores, though the latter estimate is not statistically significant.

Finally, I address the concern that some of the individual teaching practices might be wrongly classified as traditional or modern. For example, as discussed in Section 3.2, working problems could potentially be considered a modern teaching practice in the context of National Teaching Standards. In order to ensure that no single, potentially misclassified teaching practice drives the results, I exclude individual teaching practices from the traditional and modern teaching indices one at a time and re-run the regressions from Table 4.²⁰ Table A3 shows that the results from these specifications are again qualitatively and quantitatively similar to those from the headline regressions above.

5.4. Complementarities Between Traditional and Modern Teaching Practices

Traditional and modern teaching practices are typically substitutes in teaching; for example, spending more time lecturing necessarily reduces the amount of time available for small group work. Nevertheless, it is quite possible that these two sets of practices are complements in the production of cognitive skills. For instance, a teacher might introduce a new topic to her class by teaching basic facts in a lecture, only to have students develop their reasoning skills in small group discussions based on these facts afterwards. The policy implications of the findings in the previous subsections depend

¹⁹In computing these class-level shares, I again exclude each student’s own answer, although results are not sensitive to this step. Using the share of answers falling into the “every or almost every lesson” category gives qualitatively similar results.

²⁰I also experimented with including each of the six teaching practices separately in regressions. The corresponding coefficient estimates confirm that modern teaching practices are associated with higher reasoning scores while traditional teaching practices mostly raise knowing and applying scores. However, most of these effects are imprecisely estimated.

in important ways on whether these kind of complementarities exist. If they do, extreme combinations of teaching practices, such as using only modern and no traditional teaching, are unlikely to yield the desired results, in this case higher reasoning skills.

In order to shed some light on this issue, I now examine the impacts of different combinations of frequencies of use of traditional and of modern teaching practices on cognitive skills. More specifically, I compare the effect of having a teacher who uses a balanced mix of traditional and modern teaching practices with that of having a teacher who uses mainly traditional or mainly modern teaching practices. To this end, I construct an indicator for balanced teaching that takes value 1 if both the traditional teaching index and the modern teaching index in a given class are above the 25th percentile of their respective distributions, and 0 otherwise.²¹ I then use this indicator as a treatment variable in regressions like those in Table 4.

Panel A of Table 6 shows the results from these regressions. Being exposed to a balanced mix of traditional and modern teaching practices has a significant positive impact on all three cognitive skills. The effect size ranges from 5.7% of a standard deviation for the applying score to 7.1% of a standard deviation for the knowing score. In panel B of Table 6, I further divide classes in which the mix of traditional and modern teaching practices is not balanced into two groups: classes with traditional-intensive teaching (traditional teaching index > modern teaching index) and classes with modern-intensive teaching (traditional teaching index < modern teaching index). I then estimate specifications like those in panel A that include an indicator for modern-intensive teaching in addition to the indicator for balanced teaching. The results from these regressions confirm that exposure to a balanced mix of traditional and modern teaching practices raises students' cognitive skills. The estimates also show that being in a modern-intensive class rather than a traditional-intensive class has no significant effect on knowing and applying scores, but a modest positive effect on reasoning scores.

The results in Table 6 point to the existence of complementarities between traditional and modern teaching practices in the production of cognitive skills. In combination with the results in Table 4, this suggests that while increasing the amount of modern teaching can be expected to raise students' reasoning skills, the optimal combination of teaching practices for improving these skills will also include some traditional teaching.

6. International Evidence

The debate whether traditional or modern teaching practices are better for student learning in schools is not exclusive to the United States. In Israel, for example, a recent reform calls for a reduction in the use of traditional teaching practices at the post-

²¹Sixty percent of students are in classes where the combination of teaching practices is balanced according to this definition. Using the 20th percentile or the 33rd percentile as cutoff points gives qualitatively similar results.

primary level (Lavy, 2011). In contrast, the English Education Secretary Michael Gove has been calling for a return to traditional teaching in schools (Walker, 2012). In this section, I exploit the international dimension of TIMSS in order to analyze the effects of teaching practices on cognitive skills in other countries. This is an interesting topic in its own right, and it will provide an insight into the external validity of the results obtained for the United States.

The analysis focuses on nine advanced economies (as defined by the International Monetary Fund) that participated in TIMSS 2007: three Anglo-Saxon economies (Australia, England, and Scotland), Israel, and five East and Southeast Asian economies (Hong Kong, Japan, Singapore, South Korea, and Taiwan). For each of these countries and territories, I draw a sample of students using the same sampling criteria as for the United States above. I then pool the samples of the Anglo-Saxon countries and Israel on the one hand, and of the five East and Southeast Asian countries on the other hand. This classification reflects the idea that educational production, including the effects of teaching practices, might work differently in the more rarely studied Asian education systems. Panel B of Table A2 shows means and standard deviations of the traditional and modern teaching indices for these two groups of countries.²²

Table 7 presents estimates of Equation 1 for the three Anglo-Saxon countries and Israel and for the five East and Southeast Asian economies. All regressions in this table control for the full set of teacher and class characteristics as well as for country-by-subject dummies in order to allow for systematic differences in achievement in a given subject across countries. Panel A shows estimates from the pooled sample of the three Anglo-Saxon countries and Israel which are qualitatively similar to those obtained for the United States. The heterogeneity of the effects of traditional and modern teaching practices across cognitive skills is even starker here than it is in Table 4. This is reflected by the fact that the null hypothesis of equal impacts of the two treatment variables can be rejected in all of the specifications, with traditional teaching having a significantly more positive effect on overall, knowing, and applying scores, and modern teaching having a significantly more positive impact on reasoning scores.

Estimates from the pooled sample of the five East and Southeast Asian economies in panel B of Table 7 are also qualitatively and quantitatively similar to those obtained for the United States, although the different impacts of traditional and modern teaching are somewhat less pronounced than for the countries considered in panel A. Finally, panel C presents estimates from the pooled sample of all countries including the United States. The results are again very similar to the ones obtained in Table 4.²³ Due to the large

²²The Data Appendix provides additional details regarding the choice of countries and the construction of the country samples.

²³Results from country-by-country regressions are also qualitatively similar to those presented in Table 4. In particular, in specifications where the dependent variable is the knowing (applying) score, the estimated coefficient on traditional teaching is greater than the estimated coefficient on modern teaching in seven out of ten (nine out of ten) countries. In regressions where the dependent variable

sample size, the effects are very precisely estimated, and the null hypothesis of equal impacts of traditional and modern teaching can easily be rejected in all specifications. In conclusion, the evidence presented here suggests that traditional and modern teaching practices have very distinct impacts on students' cognitive skills beyond the specific setting of the United States.

7. Conclusion

There is an ongoing debate about the relative merits of traditional versus modern teaching for raising student learning in schools. Recent economic evidence suggests that traditional teaching practices, such as lecturing and rote memorization, are associated with higher student test scores, while modern teaching practices, such as working in small groups and having student discussions, are at best associated with smaller test score gains. These results cast doubt on the usefulness of National Teaching Standards in the United States, which call for a shift in emphasis from traditional towards modern teaching in classrooms.

In this paper, I provide the first evidence that traditional and modern teaching practices promote different cognitive skills in students. In particular, traditional teaching practices foster factual knowledge and competency in solving routine problems, skills that have traditionally been emphasized in schools. In contrast, modern teaching practices promote reasoning, which is precisely the skill that National Teaching Standards want to foster. I document that only a small fraction of the questions in standardized tests, both in TIMSS and elsewhere, measure students' reasoning skills. The modest or negative impacts of modern teaching practices on test scores that were documented in the previous literature are therefore not due to general ineffectiveness of these practices, but due to the design of standardized tests, which give little weight to the reasoning skills that they promote.

The identification strategy based on student fixed effects controls for most factors which could potentially confound my results. Nevertheless, not all potential biases (e.g., due to unobserved correlated teacher traits) can be eliminated, which implies that caution should be exercised when it comes to formulating policy recommendations. Keeping this caveat in mind, the results in this paper suggest two novel implications for education policy. First, the fact that teaching practices are a potentially important determinant of student learning implies that instructing teachers to emphasize modern teaching might be a useful way to increase reasoning skills among students. Second, if policy makers and educators are serious about promoting reasoning skills over factual knowledge and routine problem-solving skills, standardized tests need to be adapted to

is the reasoning score, the coefficient on traditional teaching is smaller than the coefficient on modern teaching in eight out of ten countries. However, many of these individual-country results are less precisely estimated than those for the United States due to smaller sample sizes.

give more weight to questions that measure these skills. Otherwise, teachers and schools have little incentive to employ modern teaching practices and to thus promote reasoning if their goal is to raise test scores.

Finally, an important issue that this paper has not been able to address is that of the potential effects of traditional and modern teaching practices on non-cognitive skills. It seems plausible that modern teaching in particular, with its emphasis on student interactions, promotes non-cognitive skills such as the ability to work in teams. Recent evidence by [Algan et al. \(2013\)](#), which shows that modern teaching practices foster trust between students, already points in this direction. However, more research on this topic is clearly desirable.

Data Appendix

Forty-nine countries and territories participated in the assessment of eighth-grade students in TIMSS 2007. From these, I select the sixteen advanced economies (as defined by the International Monetary Fund) other than the United States for my analysis in Section 6. In five of these economies, science is taught as a series of separate subjects (biology, chemistry, physics, etc.) rather than as an integrated subject. Unfortunately, separate test scores by cognitive domain are available in the data only for integrated science. I therefore have to exclude these five economies from the study. Moreover, all students in Italy and 62% of students in Norway have the same teacher in math and in science. These countries therefore lack the necessary within-student between-subject variation in teaching practices and have to be excluded from the analysis. This leaves me with the following nine economies: Australia, England, Scotland, Israel, Hong Kong, Japan, Singapore, South Korea, and Taiwan.

For each of these countries and territories, I select the sample in the same way as I did for the United States (see Section 3.1). This reduces the sample size by more than 40% in England, Scotland, and Israel (the average decrease in sample size in the other six economies is 17%). The principal reason for this reduction is that many students in these countries have several science teachers, and that students with more than one teacher in either subject are excluded from the sample. The rationale for this step is that these students' responses to the classroom activity questions reflect the teaching practices of all of their teachers in a given subject rather than those of a specific teacher (recall that students were asked about classroom activities in a given subject rather than with a specific teacher). In unreported regressions, I find that the results both for the United States and for the nine countries and territories considered in Section 6 are robust to keeping students who have several teachers per subject in the sample.

References

- Algan, Yann, Pierre Cahuc, and Andrei Shleifer (2013) ‘Teaching Practices and Social Capital.’ *American Economic Journal: Applied Economics* 5(3), 189–210
- Chira, Susan (1992) ‘Testing Rote, Not Reasoning.’ *The New York Times*. October 18, 1992
- Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor (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
- Cuban, Larry (1993) *How Teachers Taught: Constancy and Change in American Classrooms 1890-1990*, 2nd ed. (New York, NY: Teachers College Press)
- Dee, Thomas S. (2007) ‘Teachers and the Gender Gaps in Student Achievement.’ *Journal of Human Resources* 42, 528–554
- Foy, Pierre, Joseph Galia, and Isaac Li (2008) ‘Developing the TIMSS 2007 Mathematics and Science Assessments and Scoring Guides.’ In *TIMSS 2007 Technical Report*, ed. John F. Olson, Michael O. Martin, and Ina V.S. Mullis (Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College)
- Goldhaber, Dan D., and Dominic J. Brewer (1997) ‘Why Don’t Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity.’ *Journal of Human Resources* 32(3), 505–523
- Hamilton, Laura S., Daniel F. McCaffrey, Brian M. Stecher, Stephen P. Klein, Abby Robyn, and Delia Bugliari (2003) ‘Studying Large-Scale Reforms of Instructional Practice: An Example from Mathematics and Science.’ *Educational Evaluation and Policy Analysis* 25(1), 1–29
- Hartocollis, Anemona (2000) ‘The New, Flexible Math Meets Parental Rebellion.’ *The New York Times*. April 27, 2000
- Kane, Thomas J., Eric S. Taylor, John H. Tyler, and Amy L. Wooten (2011) ‘Identifying Effective Classroom Practices Using Student Achievement Data.’ *Journal of Human Resources* 46, 587–613
- Lavy, Victor (2011) ‘What Makes an Effective Teacher? Quasi-Experimental Evidence.’ NBER Working Paper 16885
- Murnane, Richard J., and Barbara R. Phillips (1981) ‘What Do Effective Teachers of Inner-City Children Have in Common?’ *Social Science Research* 10(1), 83–100
- NCEE (1983) ‘A Nation at Risk: The Imperative for Educational Reform.’ Report, National Commission on Excellence in Education, Washington, D.C.
- NCTM (1989) ‘Curriculum and Evaluation Standards for School Mathematics.’ Report, National Council of Teachers of Mathematics, Reston, VA
- (1991) ‘Professional Standards for Teaching Mathematics.’ Report, National Council of Teachers of Mathematics, Reston, VA

- NRC (1996) ‘National Science Education Standards.’ Report, National Research Council, Washington, D.C.
- Rockoff, Jonah E. (2004) ‘The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data.’ *American Economic Review* 94(2), 247–252
- Ruddock, Graham J., Christine Y. O’Sullivan, Alka Arora, and Ebru Erberber (2008) ‘Developing the TIMSS 2007 Mathematics and Science Assessments and Scoring Guides.’ In *TIMSS 2007 Technical Report*, ed. John F. Olson, Michael O. Martin, and Ina V.S. Mullis (Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College)
- Schwerdt, Guido, and Amelie C. Wuppermann (2011) ‘Is Traditional Teaching Really All that Bad? A Within-Student Between-Subject Approach.’ *Economics of Education Review* 30(2), 365–379
- Smith Neidorf, Teresa, Marilyn Binkley, and Maria Stephens (2006a) ‘Comparing Science Content in the National Assessment of Educational Progress (NAEP) 2000 and Trends in International Mathematics and Science Study (TIMSS) 2003 Assessments.’ Technical Report, National Center for Education Statistics, Washington, D.C.
- Smith Neidorf, Teresa, Marilyn Binkley, Kim Gattis, David Nohara, and Mark Schneider (2006b) ‘Comparing Mathematics Content in the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 Assessments.’ Technical Report, National Center for Education Statistics, Washington, D.C.
- Smith, P. Sean, Eric R. Banilower, Kelly C. McMaho, and Iris R. Weiss (2002) ‘The National Survey of Science and Mathematics Education. Trends from 1977 to 2000.’ Report, Horizon Research Inc., Chapel Hill, NC
- Walker, Peter (2012) ‘Tough exams and learning by rote are the keys to success, says Michael Gove.’ *The Guardian*. November 13, 2012
- Zemelman, Steven, Harvey Daniels, and Arthur Hyde (2005) *Best Practice. Today’s Standards for Teaching and Learning in America’s Schools*, 3rd ed. (Portsmouth, NH: Heinemann)

Table 1
Descriptive Statistics: Teacher and Class Characteristics

	Mean	SD
<i>Teacher characteristics</i>		
Female	0.63	0.48
Age (base: <30 years)		
30-39 years	0.29	0.45
40-49 years	0.25	0.43
>50 years	0.28	0.45
College major in subject taught	0.71	0.45
Postgraduate degree	0.59	0.49
Teaching certificate	0.97	0.17
Teaching experience (base: 1 year) ^a		
2 years	0.04	0.20
3-5 years	0.17	0.38
>5 years	0.72	0.45
<i>Class characteristics</i>		
Class size	24.50	6.74
Teaching time per week (minutes)	239.21	72.45

^a Because the TIMSS teacher questionnaire was administered at the end of the school year, all teachers in the sample have at least one year of teaching experience.

Table 2
Descriptive Statistics: Teaching Practices

	Mean	SD	Distribution of students' answers (%)			
			never	some lessons	about half the lessons	(almost) every lesson
Traditional teaching index	0.63	0.10				
We listen to the teacher lecture	0.62	0.14	8.35	22.15	26.07	43.43
We memorize facts, formulas, and procedures	0.62	0.15	5.63	23.95	29.56	40.85
We work (routine) problems	0.66	0.13	3.37	19.06	33.02	44.55
Modern teaching index	0.53	0.11				
We work in small groups	0.49	0.21	12.46	32.58	27.98	26.98
We give explanations	0.65	0.14	5.39	21.09	28.48	45.04
We relate what is learned to our daily lives	0.44	0.13	16.75	38.84	26.35	22.06

Notes: Class-level means of individual teaching practices exclude each student's own answer and are computed by assigning a value of 0 to the answer "never", 0.25 to "some lessons", 0.5 to "about half the lessons", and 1 to "every or almost every lesson." The traditional (modern) teaching index is constructed as the average of these class-level means across the three traditional (across the three modern) teaching practices.

Table 3
Teaching Practices and Overall Student Achievement

	(1)	(2)	(3)	(4)	(5)
Traditional teaching index	0.328** (0.094)		0.317** (0.097)	0.307** (0.094)	0.317** (0.095)
Modern teaching index		0.105 (0.097)	0.065 (0.099)	0.064 (0.099)	0.058 (0.100)
Teacher controls				Y	Y
Class controls					Y
Number of students	6,057	6,057	6,057	6,057	6,057
Average <i>R</i> -squared	0.912	0.912	0.912	0.912	0.913
H_0 : TradTI = ModnTI (<i>p</i> value)			0.103	0.107	0.087

Notes: Results from student fixed-effects regressions with the standardized overall test score as the dependent variable. All regressions control for subject dummies and are run separately for each possible combination of the five plausible test score values for each subject. The table reports the mean coefficient estimates as well as the the average *R*-squared from these 25 regressions. Teacher and class controls are listed in Table 1. Standard errors in parentheses are adjusted for the imputation variance and are robust to clustering at the class level. ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

Table 4
Teaching Practices and Cognitive Skills

	Overall	Knowing	Applying	Reasoning
	(1)	(2)	(3)	(4)
Traditional teaching index	0.317** (0.095)	0.418** (0.130)	0.359** (0.129)	0.036 (0.132)
Modern teaching index	0.058 (0.100)	0.007 (0.102)	-0.014 (0.108)	0.221~ (0.114)
Teacher controls	Y	Y	Y	Y
Class controls	Y	Y	Y	Y
Number of students	6,057	6,057	6,057	6,057
Average <i>R</i> -squared	0.913	0.856	0.853	0.824
H ₀ : TradTI = ModnTI (<i>p</i> value)	0.087	0.020	0.036	0.327

Notes: Results from student fixed-effects regressions with the overall/known/apply/reasoning score as the dependent variable (column 1/2/3/4, respectively). All regressions control for subject dummies and are run separately for each possible combination of the five plausible test score values for each subject. The table reports the mean coefficient estimates as well as the the average *R*-squared from these 25 regressions. Teacher and class controls are listed in Table 1. Standard errors in parentheses are adjusted for the imputation variance and are robust to clustering at the class level. ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

Table 5
Robustness to Alternative Measures of Teaching Practices

	Overall	Knowing	Applying	Reasoning
	(1)	(2)	(3)	(4)
<u>(A) Share of answers \geq “about half the lessons”</u>				
Traditional teaching index	0.270** (0.085)	0.315** (0.114)	0.283** (0.101)	0.087 (0.113)
Modern teaching index	0.044 (0.075)	0.024 (0.084)	-0.020 (0.088)	0.228* (0.091)
Average <i>R</i> -squared	0.913	0.856	0.853	0.825
H_0 : TradTI = ModnTI (<i>p</i> value)	0.073	0.054	0.034	0.400
<u>(B) Gap between traditional and modern teaching</u>				
TradTI – ModnTI	0.103 (0.077)	0.176* (0.087)	0.162~ (0.088)	-0.110 (0.094)
Average <i>R</i> -squared	0.912	0.856	0.853	0.824

Notes: Results from variations of the specifications used in Table 4. In panel A, the traditional (modern) teaching index measures the share of students that report engaging in traditional (modern) classroom activities “about half the lessons” or “every or almost every lesson.” In panel B, the treatment variable is the difference between the traditional and the modern teaching index. The number of students is 6,057 in all regressions. See the notes to Table 4 for additional controls included in all of the specifications. Standard errors in parentheses are adjusted for the imputation variance and are robust to clustering at the class level. ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

Table 6
Complementarities Between Traditional and Modern Teaching Practices

	<u>Overall</u>	<u>Knowing</u>	<u>Applying</u>	<u>Reasoning</u>
	(1)	(2)	(3)	(4)
<u>(A) Balanced vs. unbalanced teaching</u>				
Balanced teaching	0.056** (0.019)	0.071** (0.025)	0.057* (0.026)	0.062* (0.026)
Average <i>R</i> -squared	0.912	0.856	0.853	0.825
<u>(B) Balanced vs. unbalanced modern-intensive vs. unbalanced traditional-intensive teaching</u>				
Balanced teaching	0.061** (0.020)	0.073** (0.026)	0.059* (0.027)	0.081** (0.029)
Modern-intensive teaching	0.019 (0.028)	0.007 (0.033)	0.008 (0.031)	0.068~ (0.039)
Average <i>R</i> -squared	0.912	0.856	0.853	0.825

Notes: Results from variations of the specifications used in Table 4. In panel A, the indicator for balanced teaching takes value 1 if both the traditional teaching index and the modern teaching index in a given class are above the 25th percentile of their respective distributions, and 0 otherwise. In panel B, classes for which this indicator is 0 are further divided into traditional-intensive and modern-intensive classes based on whether the traditional or the modern teaching index is greater. The number of students is 6,057 in all regressions. See the notes to Table 4 for additional controls included in all of the specifications. Standard errors in parentheses are adjusted for the imputation variance and are robust to clustering at the class level. ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

Table 7
Teaching Practices and Cognitive Skills: International Evidence

	Overall	Knowing	Applying	Reasoning
	(1)	(2)	(3)	(4)
(A) Other Anglo-Saxon countries + Israel				
Traditional teaching index	0.392** (0.081)	0.651** (0.129)	0.411** (0.116)	-0.105 (0.125)
Modern teaching index	-0.019 (0.078)	-0.130 (0.104)	-0.122 (0.104)	0.324** (0.106)
Average <i>R</i> -squared	0.913	0.851	0.852	0.822
H ₀ : TradTI = ModnTI (<i>p</i> value)	0.001	0.000	0.002	0.018
(B) East and Southeast Asian countries				
Traditional teaching index	0.340** (0.088)	0.437** (0.129)	0.372** (0.118)	0.103 (0.143)
Modern teaching index	0.144~ (0.084)	0.025 (0.116)	0.137 (0.100)	0.367** (0.105)
Average <i>R</i> -squared	0.913	0.848	0.861	0.829
H ₀ : TradTI = ModnTI (<i>p</i> value)	0.170	0.048	0.207	0.201
(C) All countries pooled				
Traditional teaching index	0.362** (0.051)	0.530** (0.073)	0.392** (0.066)	0.011 (0.076)
Modern teaching index	0.048 (0.048)	-0.047 (0.061)	-0.007 (0.063)	0.313** (0.067)
Average <i>R</i> -squared	0.913	0.850	0.857	0.826
H ₀ : TradTI = ModnTI (<i>p</i> value)	0.000	0.000	0.000	0.006

Notes: Results from specifications as reported in Table 4 run on the samples of eighth-grade students from other countries participating in TIMSS 2007. Countries in panel A: Australia, England, Scotland, and Israel (9,126 students). Countries in panel B: Hong Kong, Japan, Singapore, South Korea, and Taiwan (17,110 students). Countries in panel C: all countries in panels A and B and the United States (32,293 students). All regressions include country-by-subject fixed effects in addition to the controls mentioned in the notes to Table 4. Standard errors in parentheses are adjusted for the imputation variance and are robust to clustering at the class level. ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

Table A1

Traditional and Modern Teaching Practices in Math and Science in the TIMSS Student Questionnaire

Traditional Teaching Practices in Mathematics

We listen to the teacher give a lecture-style presentation.

We memorize formulas and procedures.

We work problems on our own.

Traditional Teaching Practices in Science

We listen to the teacher give a lecture-style presentation.

We memorize science facts and principles.

We work problems on our own.

Modern Teaching Practices in Mathematics

We work together in small groups.

We explain our answers.

We relate what we are learning in mathematics to our daily lives.

Modern Teaching Practices in Science

We work in small groups on an experiment or investigation.

We give explanations about what we are studying.

We relate what we are learning in science to our daily lives.

Notes: Students responded to the question, "How often do you do these things in your mathematics lesson (in your science lesson)?" Possible answers are "never", "some lessons", "about half the lessons", and "every or almost every lesson."

Table A2
Additional Descriptive Statistics

	N	Mean	SD
<u>(A) Treatments used in Table 5</u>			
<i>Share of answers \geq "about half the lessons"</i>			
Traditional teaching index	6,057	0.71	0.12
Modern teaching index	6,057	0.58	0.15
<i>Gap between traditional and modern teaching</i>			
TradTI – ModnTI	6,057	0.11	0.13
<u>(B) Treatments used in Table 7</u>			
<i>Other Anglo-Saxon countries + Israel</i>			
Traditional teaching index	9,126	0.56	0.12
Modern teaching index	9,126	0.47	0.10
<i>East and Southeast Asian countries</i>			
Traditional teaching index	17,110	0.54	0.10
Modern teaching index	17,110	0.41	0.12
<i>All countries pooled</i>			
Traditional teaching index	32,293	0.56	0.11
Modern teaching index	32,293	0.44	0.12

Table A3*Robustness of Results to the Exclusion of Individual Teaching Practices*

	Overall	Knowing	Applying	Reasoning
	(1)	(2)	(3)	(4)
Traditional teaching index excluding listening to lectures	0.332** (0.080)	0.429** (0.105)	0.381** (0.110)	-0.047 (0.129)
Modern teaching index	0.084 (0.097)	0.042 (0.101)	0.016 (0.107)	0.228* (0.113)
Traditional teaching index excluding memorizing	0.166~ (0.091)	0.214~ (0.112)	0.169 (0.117)	0.060 (0.115)
Modern teaching index	0.080 (0.099)	0.036 (0.103)	0.013 (0.109)	0.218~ (0.112)
Traditional teaching index excluding working problems	0.232** (0.080)	0.318* (0.118)	0.273* (0.107)	0.072 (0.111)
Modern teaching index	0.048 (0.102)	-0.008 (0.104)	-0.026 (0.109)	0.209~ (0.115)
Traditional teaching index	0.350** (0.099)	0.494** (0.138)	0.390** (0.135)	0.039 (0.133)
Modern teaching index excluding working in groups	-0.065 (0.096)	-0.217~ (0.115)	-0.097 (0.113)	0.102 (0.111)
Traditional teaching index	0.324** (0.093)	0.415** (0.129)	0.356** (0.128)	0.063 (0.130)
Modern teaching index excluding giving explanations	0.063 (0.076)	0.071 (0.091)	0.012 (0.097)	0.197* (0.096)
Traditional teaching index	0.316** (0.094)	0.409** (0.130)	0.354** (0.129)	0.052 (0.131)
Modern teaching index excluding relating to daily lives	0.095 (0.080)	0.080 (0.080)	0.021 (0.084)	0.185~ (0.096)

Notes: Results from variations of the specifications in Table 4 in which one teaching practice is excluded from the treatment variables at a time. The results presented here are drawn from a total of 24 different specifications (six teaching practices to exclude times four dependent variables). The number of students is 6,057 in all regressions. See the notes to Table 4 for additional controls included in all of the regressions. Standard errors in parentheses are adjusted for the imputation variance and are robust to clustering at the class level. ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.