Teaching Practices and Cognitive Skills*

Jan Bietenbeck

CEMFI†

May 1, 2013

Abstract

National Teaching Standards in the United States call for a shift from traditional teaching practices such as lecturing and rote memorization towards modern ones such as student discussion and group work in schools. Yet a small literature in economics has consistently found that teachers who emphasize traditional teaching practices raise test scores, while the evidence for modern teaching practices is less clear. In this paper, I show that traditional and modern teaching practices promote different cognitive skills in students. Exploiting a unique data set which contains test scores measuring performance on three distinct cognitive skill dimensions, I find that traditional teaching practices increase students' factual knowledge and their competency in solving routine exercises but have no significant effect on their reasoning skills. Exactly the reverse is true for modern teaching practices. I provide evidence that in many standardized tests, reasoning skills are not well measured, which explains the ambiguous findings regarding the effect of modern teaching practices in the literature.

JEL Code: *I21*

Keywords: *teaching practices; cognitive skills; TIMSS*

*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, Jesse Rothstein, and seminar participants at CEMFI and at the MIT labor lunch 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. All errors are my own.

†Address: CEMFI, Casado del Alisal 5, 28014 Madrid, Spain; Email: bietenbeck@cemfi.es

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 the use of drill worksheets in classrooms to teach students basic facts and procedures. Several reform movements during the 20th century attempted to introduce a more student-centered approach to teaching into schools, in which student 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 with the release of National Teaching Standards (e.g., NCTM 1989, 1991; NRC 1996). These prescribe a change of emphasis from traditional towards modern, student-centered teaching practices in order to promote students' reasoning skills over mere factual knowledge and routine problem-solving skills. This is motivated by the perception that reasoning skills are becoming increasingly important in the labor market. In practice, the implementation of the changes advocated by National Teaching Standards has often been met with resistance from parents and teachers, and has led to an intensification of the debate between advocates of traditional and modern teaching practices.[1]

In this paper, I study the effects of traditional and modern teaching practices (as currently implemented in schools) on students' cognitive skills in order to inform the debate about what works best in classrooms. In this way, I also hope to gain an insight into the potential consequences of the shift from traditional towards modern teaching practices advocated by National Teaching Standards. My starting hypothesis is that rather than being substitutes in educational production, traditional and modern teaching practices promote different cognitive skills in students. In particular, I hypothesize that just as National Teaching Standards imply, modern teaching practices promote reasoning skills, while traditional teaching practices promote the knowledge of basic facts and procedures. I test this claim using data from the 2007 wave of the Trends in International Mathematics and Science Study (TIMSS) for the United States. This data is unique in that it contains both information on teaching practices and test scores measuring eighth-grade students' performance on three distinct cognitive skill dimensions, one of which reflects reasoning skills, and two of which reflect factual knowledge and routine problem-solving skills.

I measure teaching practices using information from the TIMSS 2007 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 identify three activities reflecting traditional

[1] National Teaching Standards categorize teaching practices as "to be de-emphasized" (also: "to be given less attention") or "to be emphasized" (also: "to be given increased attention"). In line with the previous literature (e.g., Lavy 2011, Schwerdt and Wuppermann 2011), I adopt the terminology "traditional" and "modern" teaching practices here.

teaching practices and three activities reflecting modern teaching practices in the questionnaire. The traditional activities are listening to the teacher lecture, memorizing facts, formulas and procedures, and working routine problems. The modern activities are working in small groups, giving explanations, and relating what is learned to students' daily lives. I summarize the information on how frequently these activities are done in two class-level indices of traditional and modern teaching. These indices reflect the emphasis that a teacher places on traditional versus modern teaching practices in a particular class.

I begin my analysis of the effects of teaching practices on students' cognitive skills by relating the traditional and modern teaching indices to students' overall test scores in mathematics and science. The empirical model exploits the fact that each student is observed twice - once in math, and once in science - to include student fixed effects. This addresses the most obvious threats to identification and means that the effects of interest are identified using the variation of teaching practices between the two subjects for each student. I find that the traditional teaching index has a positive and sizable estimated effect on students' overall test scores, while the estimated effect of the modern teaching index is close to zero and not statistically significant. Taken at face value, this result suggests that a shift from traditional towards modern teaching practices as advocated by National Teaching Standards will harm student performance on standardized tests.

I then exploit the availability of subscores measuring performance on three distinct cognitive skill dimensions to investigate whether the effects of traditional and modern teaching practices are heterogeneous across these dimensions. There is a sizable estimated effect of the traditional teaching index on students' factual knowledge and on their routine problem-solving skills, but no significant effect on students' reasoning skills. Conversely, the estimated effect of the modern teaching index on students' factual knowledge and routine problem-solving skills is practically zero, while its estimated impact on reasoning skills is sizable. These results confirm my initial hypothesis. The positive effect of the modern teaching index 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, these results 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.

This paper is related to a small but growing literature in economics on the effects of teaching practices on student outcomes. This literature is characterized by stark differences in the definition and measurement of teaching practices. While this makes it difficult to draw general conclusions, the evidence points in the direction of a positive effect on test scores of teaching practices that would be considered traditional by National Teaching Standards, and a more ambiguous effect of teaching practices that would be considered modern.[2] The only study that defines traditional and modern teaching practices in a similar way as this paper does is Lavy (2011). The author uses data from a student survey in Israel in which students rated on a six-point scale which proportion of their teachers across all subjects engaged in specific teaching practices. From this information, Lavy (2011) constructs two composite indices of traditional and modern teaching at the class level. He then relates these indices to test scores in a student fixed effects regression, exploiting variation in teaching practices over time. It turns out that both the traditional and the modern teaching practice index are positively related to student achievement, but that the former has a larger estimated impact. These effects are heterogeneous across students with different socioeconomic backgrounds and across boys and girls.

This paper makes three important contributions to the literature discussed in the previous paragraph. First, I study the heterogeneity of the effects of traditional and modern teaching practices across different cognitive skill dimensions. This is important given the result of a sizable positive effect of the modern teaching index on students' reasoning scores which is masked in the overall test score regression.[3] Second, I provide the first comprehensive analysis of the effects of teaching practices on student test scores for the United States. In particular, this paper considers a larger number of teaching practices than Schwerdt and Wuppermann (2011) and Goldhaber and Brewer (1997), and is the only one to measure credibly both traditional and modern teaching practices as defined by National Teaching Standards.[4] Third, in an extension of my analysis, I exploit the international dimension of the TIMSS database to provide evidence on the effects of traditional and modern teaching practices from other countries. The finding that the estimates on these other samples are very similar to the ones obtained for the United States lends credibility to my results and is evidence of their external validity.

The remainder of the paper is organized as follows. Section 2 describes the nature and contents of National Teaching Standards in more detail. Section 3 presents the data and the measurement of teaching practices and cognitive skills. Section 4 discusses the empirical strategy. Section 5 presents the headline

[2]Schwerdt and Wuppermann (2011), using data from a teacher time use survey in TIMSS 2003 for the United States, find that teachers who spend more time lecturing (a traditional teaching practice) are associated with higher test scores in math and science. Van Klaveren (2011), using data for the Netherlands from the same database and a very similar empirical setup, fails to find any significant effect of time spent lecturing, even though his confidence interval includes the point estimate found by Schwerdt and Wuppermann (2011). Goldhaber and Brewer (1997) include a variety of modern teaching practices in an education production function and find that they lower test scores in mathematics. Lavy (2011) reports that traditional teaching practices have a larger positive effect on test scores than modern teaching practices do (see main text for discussion). Kane, Taylor, Tyler, and Wooten (2011) find that classroom observations of teaching practices by trained evaluators are substantively related to student achievement growth, with some practices predicting achievement more than others. However, none of the practices they consider fits into the framework of traditional versus modern teaching.

[3]The finding that modern teaching practices promote skills that are not measured well by standardized tests is consistent with very recent work by Algan, Cahuc, and Shleifer (2012), who show that modern teaching practices promote the formation of social capital.

[4]Goldhaber and Brewer (1997) include only modern teaching practices in their education production function. Schwerdt and Wuppermann (2011) focus on lecturing and in-class problem solving. While they regard the latter as a modern teaching practice, this is not necessarily the case in National Teaching Standards (see the discussion in footnote 13).

results, robustness checks, and heterogeneity results. Section 6 extends the analysis to other countries. Section 7 concludes.

2 National Teaching Standards

In its 1983 report *A Nation At Risk*, 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" (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 a 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 this reform was supposed to be an improvement in the quality of teaching in schools.

The National Council of Teachers of Mathematics' *Curriculum and Evaluation Standards* (NCTM 1989) were a direct response to this call for reform. They set out a body of mathematical skills that students should master at different grade levels, placing a strong emphasis on reasoning skills relative to factual knowledge. This was supposed to reflect the demands of the modern labor market, in which "businesses no longer seek workers with strong backs, clever hands, and 'shopkeeper' arithmetic skills [...] [but] the ability to work with others on problems, the ability to see the applicability of mathematical ideas to common and complex problems, [and] preparation for open problem situations, since most real problems are not well formulated." The key reform needed in order to achieve this type of mathematical literacy according to the document was to change the way in which mathematics is taught in schools, reflecting the idea that not only what but also how students are taught matters. In particular, teachers should emphasize working in groups, discussion among students, and connecting mathematics to the world outside of the classroom, among other practices. In contrast, teachers should place less emphasis on lecturing, memorizing facts, formulas and procedures, and practicing routine problems.[5]

The *Curriculum and Evaluation Standards* received considerable attention by policy makers and the media, with the initial reactions being overwhelmingly positive. This led the U.S. Department of Education to commission other professional education bodies to develop similar standards for other school subjects (Zemelman, Daniels, and Hyde 2005). The most relevant of these standards for this paper are

[5]The proposed changes in teaching practices were later outlined in more detail in the *Professional Standards for Teaching Mathematics* (NCTM 1991).

the National Research Council's *National Science Education Standards* (NRC 1996) for science teaching in schools. Zemelman, Daniels, and Hyde (2005) survey the recommendations made by these subject standards and find that they are remarkably congruent. In particular, all of the standards emphasize the importance of students' reasoning skills over factual knowledge, and all of them view a shift from traditional towards modern teaching practices as a key ingredient for achieving this new type of literacy.[6] The entirety of these subject standards has become commonly referred to as National Teaching Standards.

Over the past two decades, National Teaching Standards have found their way into numerous state curricula as well as teacher education and training programs. However, the implementation of their recommendations has often been met with criticism. In particular, many parents fear that with the focus on student interactions rather than on rote memorization and practice their children will fail to learn basic facts and procedures.[7] A relevant question is therefore whether teaching practices have actually changed in American classrooms. To the best of my knowledge, the only study of long-term trends in teaching practices after the 1980s is Smith et al. (2002). The study compares the frequency of use of several teaching practices in 1993 and 2000 as reported by a nationally representative sample of mathematics and science teachers. Two of these teaching practices are of particular interest here: lecturing, which is a traditional practice according to National Teaching Standards, and working in groups, which is a modern practice. The authors find that there has been no significant change in the frequency that teachers report using each of this practices in math across the two waves. In contrast, there was a small reduction in the use of lecturing in science (but no significant change in the frequency of group work).

The existing empirical evidence therefore does not support the idea that teaching practices changed dramatically between 1993 and 2000. However, this period might simply be too short to measure a change in teaching practices. In particular, it might take some more years for curricula and textbooks to be updated to conform with National Teaching Standards and for teachers to undergo related training programs. In the following section, I therefore provide additional evidence on trends in teaching practices using data from TIMSS. This evidence covers the time period from 1995 to 2007 and supports the idea of a shift from traditional towards modern teaching practices in American schools.

[6]The *National Science Education Standards*, similarly to the *Curriculum and Evaluation Standards*, motivate the emphasis on reasoning skills by stating that scientific literacy based on reasoning reflects the "skills that people use every day, like solving problems creatively, thinking critically, [and] working cooperatively in teams" (p. ix).

[7]In mathematics, the implementation of curricula based on the NCTM (1989, 1991) documents has led to what has been called the "math wars" by some commentators. For an illustrative example of this conflict between advocators of traditional and modern teaching practices, see the article "The New Flexible Math Meets Parental Rebellion" published in the *New York Times* on April 27, 2000.

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) 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 for the three cognitive skill dimensions used in my analysis.[8]

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 here because it implies a greater variation in teaching practices across subjects (recall that this is the variation used in the empirical analysis in order to identify the effects of interest). 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.

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 performance on the tests.[9] Two studies by the National Center for Education Statistics (Smith Neidorf et al. 2006; Smith Neidorf, Binkley, and Stephens 2006) compare the standardized tests in TIMSS to the eighth-grade math and science tests used in its own National Assessment of Educational Progress (NAEP), the largest nationally

[8] My 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 necessary variation of teaching practices across subjects for each student does not exist in this sample.

[9] In contrast, the Programme for International Student Assessment, better known by its acronym PISA, tests students' overall (i.e. not grade-specific) skills in problem solving in math and science.

representative assessment of American students. The studies find that the tests in both assessments are very similar in terms of both the content covered and the cognitive skills measured. This confirms the validity of using TIMSS test scores for measuring eighth-grade curriculum knowledge in the United States.

4.1 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 not 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. Below, I describe how I measure teaching practices at the class level using information from the TIMSS student questionnaire. I guarantee a minimum of precision in this measurement by requiring that at least five students that answered the questions on teaching practices be observed per class.[10] This means that another 372 students are dropped from the sample. The final sample consists of 6,057 students in 425 math classes and 462 science classes in 221 schools.

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 a female-teacher dummy, dummies for being 30-39, 40-49, and 50 or more years of age, dummies for having a teaching certificate, having a postgraduate (Master's or PhD) degree, and having majored in the subject taught, and dummies for having 1-2, 3-5, and 6 or more years of teaching experience.[11] The class variables are class size and teaching time in minutes per week. Table A1 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 variable are included in the regressions. The results in the later parts of this paper are 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.

[10] The response rate to the teaching practice questions is 93%.

[11] I choose this functional form for teaching experience because prior research (Rockoff 2004, Clotfelter, Ladd, and Vigdor 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.

4.2 Measuring Teaching Practices

I use information on classroom activities 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 each subject. I assign a value of 1 to the answer "never", 2 to "some lessons", 3 to "about half of the lessons", and 4 to "every or almost every lesson." Note that in contrast to Lavy (2011), where students indicate which proportion of their teachers across all subjects employs a particular teaching practice, students respond separately for math and science here. This allows me to identify a more direct effect of teaching practices on students' cognitive skills in the same subject.

I refer to National Teaching Standards (NCTM 1989, 1991; NRC 1996; Zemelman, Daniels, and Hyde 2005) in order to select those classroom activities from the questionnaire that can be categorized as reflecting either a traditional or a modern teaching practice. I should note that the list of classroom activities that students are asked about differs slightly between the two subjects. I concentrate here on the activities that are available for both math and science and identify three traditional and three modern activities in the questionnaire.[12] The three traditional practices are listening to the teacher lecture, memorizing facts, formulas and procedures, and working problems. The three modern practices are working in small groups, giving explanations, and relating what is learned to students' daily lives.[13] Table 1 presents the six activities with their exact wording from the questionnaire separately for math and science.

In order to gain precision in the measurement of teaching practices, I aggregate students' answers to the class level as follows. First, I calculate the mean of each student's answers across traditional and modern teaching practices separately for math and science. In a second step, I then aggregate the resulting composite measures of traditional and modern teaching to the class level by taking the simple average across all students in the class while excluding each student's own answer.[14] In this way, I hope to overcome the potential measurement error that might confound each student's individual answers. The resulting class-level indices of traditional and modern teaching measure the emphasis that a teacher

[12]By concentrating on activities available for both subjects, I ensure that my results are not driven by mechanical differences in the construction of the treatment variables between subjects. The results are however very similar when one further modern math activity - using a variety of solution paths to solve complex problems - and one further traditional science activity - reading textbooks - that are available in the questionnaire are included in the analysis. These results are available upon request.

[13]It 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 33% compared to 22% for modern teaching practices). In a robustness check, I show that my results do not depend on the inclusion of this practice in either index. In the rest of the paper, I will refer to this practice as "working routine problems" in order to emphasize that it is considered traditional.

[14]The results are very similar when each student's own answer is included in the index. These results are available upon request.

places on traditional versus modern teaching practices in a particular class. Table 2 shows the means, standard deviations, and the 20th and 80th percentile for the traditional and modern teaching indices as well as for the individual teaching practices included in them. The traditional teaching index has a mean of 3.09 points, while the mean of the modern teaching index is lower at 2.79 points. The means of the individual teaching practice variables range from 2.54 to 3.18.

It is important to note that the categorical nature of students' answers implies that the two indices do not stand in a mechanical trade-off to each other: scoring one point higher on the traditional teaching index does not necessarily imply that the modern teaching index decreases by one point. For example, a teacher that frequently mixes traditional teaching practices with modern ones will score high on both of the indices. Indeed, it turns out that the two indices are weakly positively correlated with a correlation coefficient of 24%.[15] Importantly, this does not prevent me from answering my question of interest, namely what the effects of traditional and modern teaching practices are on students' cognitive skills.

In the empirical analysis below, I include the traditional and modern teaching indices in the same regressions. This means that the estimated coefficients can be interpreted as the effect of the traditional (modern) teaching index on test scores, *holding the modern (traditional) teaching index constant.* Moreover, as a robustness check I construct a treatment variable which forces there to be a trade-off between traditional and modern teaching practices and show that the results using this treatment are qualitatively similar to the ones obtained using the two teaching practice indices. I discuss this issue further in Section 5.

4.3 Measuring Cognitive Skills

The math and science tests in TIMSS are organized around three so-called cognitive domains reflecting 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 standard in classroom exercises. The *reasoning* domain assesses students' capacity for logical, systematic thinking by confronting them with complex problems set in unfamiliar contexts. Each test question belongs to one of these three domains and gives a certain number of score points if answered correctly. The distribution of questions and score points over the three domains is determined by education experts based on what they deem appropriate for eighth-grade students. In particular, the shares of score points in the knowing, applying and reasoning domains are 35%, 41%, and 24%,

[15]Lavy (2011) reports a correlation coefficient of 81% for his modern and traditional teaching practice indices.

respectively, in math and 37%, 41%, and 22%, respectively, in science (Ruddock et al. 2007).

From the description in the previous paragraph, it is clear that the knowing and applying domains reflect the skills that schools traditionally promoted, whereas the reasoning domain reflects the skills emphasized by National Teaching Standards.[16] It is important to note that these latter skills are not reflected well in the standardized tests, with their share of score points being only a fifth to a quarter of the overall test. Importantly, this is similar to the share of items measuring reasoning skills in the eighth-grade NAEP math and science tests (Smith Neidorf et al. 2006; Smith Neidorf, Binkley, and Stephens 2006). This means that the low emphasis given to reasoning skills is not an artefact of the TIMSS assessment, but is likely a common feature of many standardized tests.

The TIMSS assessment uses an incomplete-booklet design which means that each individual student 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 test scores are made available in the data in the form of five random draws from this distribution (five so-called *plausible values*). Regressions in the later parts of this paper account for the uncertainty regarding a student's true test score introduced by this design feature.[17] In addition to the overall math and science scores, the TIMSS data contains separate test scores for achievement on the three cognitive domains for each of the two subjects. I standardize the overall and cognitive-domain specific test scores to have a mean of zero and a standard deviation of one in the full sample.

4.4 Evidence on Trends in Teaching Practices Between 1995 and 2007

I now present evidence on trends in teaching practices in American math and science classrooms using information from the student questionnaires of the first four waves of TIMSS. The student questionnaires changed considerably between the waves, and not all of them contain comparable information on teaching practices, with the 2011 questionnaire containing no information on teaching practices at all. It is however possible to track the prevalence of one modern teaching practice, working in groups, across the waves between 1995 and 2007. In particular, in 1995, 22% of students reported working in groups in almost every lesson, and this number rose to 23% in 1999, 26% in 2003, and 28% in 2007.[18] There is therefore

[16]The three cognitive domains in TIMSS also map directly into the concepts of fluid and crystallized intelligence that are widely used in psychology. Originally developed by Cattell (1971), crystallized intelligence is defined as the ability to use previously acquired knowledge to solve problems. This corresponds to the knowing and applying domains here. In contrast, fluid intelligence measures the ability to reason logically and to solve complex problems in unfamiliar situations. This is exactly what the reasoning domain measures.

[17]Test score regressions are run separately for each of the five plausible values. Tables report the mean estimate from these regressions for each coefficient as well as the average R-squared. Standard errors are adjusted for the imputation variance using the formula provided in the TIMSS 2007 Technical Report (Foy, Galia, and Li 2008).

[18]The labelling 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"

a clear upward trend in the frequency that this modern teaching practice is being used. Unfortunately, there is no traditional teaching practice that appears in student questionnaires in all of these four waves. However, the fraction of students reporting listening to the teacher lecture every or almost every lesson declined from 44% to 43% between 2003 and 2007. While this evidence is not conclusive, together with the evidence presented in the previous section it is consistent with the view that a slow but steady shift from traditional towards modern teaching practices is underway.

4 Empirical Strategy

The ideal experiment to estimate the effect of teaching practices on test scores would randomly vary teaching practices across students. There are two reasons why in practice the pairing of students and teaching practices will not be random. First, on the demand side, 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, on the supply side, it is plausible that teachers partially adjust their teaching practices to the students they face. If teaching practices are (partly) a function of student-level determinants of test scores that are not controlled for in a regression (e.g., students' unobserved academic ability), this will again lead to a bias in the estimated effects of teaching practices.

Previous studies have addressed these issues by including student fixed effects in the empirical model. This accounts for student sorting into teaching practices across schools and classrooms based on fixed student characteristics such as academic ability. Moreover, under some assumptions, which I discuss 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 effects 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

in 2003 and 2007. The figures are averages across math and science, and the trends in both subjects were similar (separate figures by subject are available upon request).

subjects for each student.[19]

Below, I present estimates of the following empirical model:

$$A_{ijs} = \alpha + TradTI_{ijs}\beta + ModnTI_{ijs}\gamma + X_{js}\delta + \lambda_i + \varepsilon_{ijs}, \tag{1}$$

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, X_{js}. λ_i is the student fixed effect, which controls for any subject-invariant unobservable 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 unobservable school characteristics. In the following section, I will first show results where A_{ijs} is the overall math or science test score. I then replace the A_{ijs} by the cognitive-domain specific test scores in order to estimate separately the effect of traditional and modern teaching practices on students' performance on the three distinct cognitive skill dimensions.

The parameters of interest in (1) are β and γ. 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 are correlated with the teaching practice indices. That is, the student fixed effects in my model do not account for sorting of students into teaching practices across schools and classrooms based on subject-specific academic ability. I cannot address this issue definitely with the data at hand. However, the fact that math and science are closely related subjects somewhat mitigates this concern here. Moreover, in a related study that also relies on between-subject variation for identification, Clotfelter, Ladd, and Vigdor (2010) provide suggestive evidence based on tracking patterns that academic ability is indeed highly correlated across subjects. Finally, I should emphasize again that any sorting based on students' overall academic ability is accounted for by the estimation strategy.

Another way in which the identifying assumption in (1) could be violated is if teachers who emphasize certain teaching practices have 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 practice index. This omitted-variable problem is a challenge which virtually all studies that try to identify the

[19] Identification based on within-student between-subject variation has been used by Dee (2007), Clotfelter, Ladd, and Vigdor (2010), and Schwerdt and Wuppermann (2011), among others.

effect of a particular teacher trait on student outcomes face, and there is usually no definite solution to this problem. In this paper, I partially address this concern by controlling for a rich set of teacher and classroom characteristics (shown in Table A1). I can however not completely exclude the possibility that my coefficient estimates actually pick up the effect of some other unobserved teacher trait.

A final concern regarding the empirical strategy is that in contrast to the teacher characteristics usually studied in the literature such as experience and certification status, teachers' teaching practices are not a fixed characteristic. Indeed, as already noted above, it is quite plausible that teachers partially adjust their teaching practices to the students they face. In order to discuss the implications of this adjustment for the identification strategy, it is useful to think of teaching practices as being made up of a fixed part that varies across teachers but not across classes for a given teacher, and a variable part that depends on student and class characteristics. Clearly, if the variable part of teaching practices only depends on students' subject-invariant determinants of test scores, the inclusion of student fixed effects in (1) adequately accounts for this adjustment. If instead the subject-specific academic abilities of students are an input into teaching practices, and if there is subject-specific ability at the class level, this will violate the identifying assumption in (1).[20] Therefore, while I cannot completely exclude that teachers' adjustment of teaching practices partially drives the results presented below, bias through this channel is most likely minimal due to the inclusion of student fixed effects.[21]

5 Results

5.1 The Relationship Between Teaching Practices and Students' Cognitive Skills

Table 3 presents estimates of the relationship between teaching practices and students' cognitive skills from variations of equation (1). In columns (1)-(3), the dependent variable is the overall test score, while columns (4), (5), and (6) report results from regressions with the knowing, applying, and reasoning test score as the dependent variable, respectively. Column (1) includes as regressors only the two teaching practice indices and a subject dummy, which is included in all of the specifications. Teacher and class controls as shown in Table A1 are then added consecutively in columns (2) and (3). Regressions are weighted using the student sampling weight supplied with the TIMSS database and are run separately

[20]Similarly, if teaching practices depend on other class-level characteristics, I have to assume that the class controls included in (1) adequately capture these.

[21]A way to address the issue of endogeneity 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 strategy is not feasible here because as mentioned in the previous section, teachers in the TIMSS data are usually observed only with one class. In particular, in the final estimation sample, only 30% of students have teachers in both subjects that are observed also with other classes.

for each of the five plausible values. Table 3 reports the mean coefficient estimates as well as the average R-squared from these five regressions. Standard errors in parentheses are adjusted for the imputation variance and allow for clustering at the class level.

Column (1) shows a positive and highly significant estimated effect of the traditional teaching index on students' overall test scores. The coefficient of 0.118 implies that moving a student from the 20th percentile of the traditional teaching index to the 80th percentile is associated with a 5.4% of a standard deviation increase in her overall test score. In contrast, the estimated coefficient on the modern teaching index is less than a quarter of that in size and not statistically significant.[22] While the hypothesis that the two coefficients are equal cannot be rejected in this specification (last row of Table 3; p value = 0.142), the results point in an interesting direction: taken at face value, they suggest that the shift of emphasis from traditional towards modern teaching practices advocated by National Teaching Standards will harm student achievement on standardized tests.

Columns (2) and (3) show that the inclusion of teacher and class controls has very little impact on the coefficient estimates of the two teaching practice indices. This robustness is reassuring because it can be interpreted as evidence of the validity of the identification strategy. In particular, if the selection on unobservables mirrors the selection on observables in Table 3, the stability of the coefficients across columns (1)-(3) implies that the estimated effects are unlikely to be confounded by unobserved teacher and class characteristics. Note also that the results in columns (1)-(3) are in line with those found in the previous literature. In particular, the sizable positive effect of the traditional teaching index on overall test scores corroborates similar result by Lavy (2011) and Schwerdt and Wuppermann (2011). In contrast, the small estimated coefficient on the modern teaching index is in between the negative effect found by Goldhaber and Brewer (1997) and the positive effect found by Lavy (2011).

I now investigate whether the results discussed in the previous paragraphs mask some heterogeneity of the effects of traditional and modern teaching practices across different cognitive skills. In columns (4) and (5) of Table 3, I present estimates of models in which the dependent variable is the knowing and the applying score, respectively. Recall that these scores measure the factual knowledge and the routine problem-solving skills that have traditionally been emphasized in schools. There is a positive and highly significant estimated effect of the traditional teaching index in both specifications, which is slightly larger than the effect found in column (3). In contrast, the estimated coefficient on the modern teaching index is practically zero. The two coefficients are significantly different from each other in both specifications

[22]Results from specifications that include each of the teaching practice indices individually are quantitatively similar to those presented here and are available upon request.

(p value = 0.018 in column (4), and p value = 0.044 in column (5)). Note that it makes intuitive sense that the traditional teaching index has such a large effect here. After all, the practice of memorizing facts, formulas, and procedures, which is included in the index, specifically aims at raising students' factual knowledge. Likewise, the practice of working routine problems, which is also included in the traditional teaching index, is meant to improve the skills measured by the applying score.

In column (6) of Table 3, I present the results for the reasoning domain, which measures the skills 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 less than a sixth in size of that found in columns (4) and (5) and is not statistically significant. In contrast, there is a positive and significant estimated effect of the modern teaching index. The coefficient of 0.085 implies that moving a student from the 20th to the 80th percentile of the modern teaching index is associated with a 4.8% of a standard deviation increase in her reasoning score. This effect is comparable in size to the one of the traditional teaching index in columns (4) and (5). Finally, note that I cannot reject the hypothesis that the coefficients on the two teaching practice indices are equal in this specification (p value = 0.363).

What do these results imply for the recommendations made by National Teaching Standards? They suggest that a higher emphasis on modern teaching practices will not be associated with an increase in test scores, and that a lower emphasis on traditional teaching practices will decrease test scores. This is so because traditional and modern teaching practices promote different cognitive skills in students, and in particular because the reasoning skills which are promoted by modern teaching practices are not reflected well in standardized tests.[23] That a lower emphasis on traditional teaching practices reduces students' factual knowledge and routine problem-solving skills does not invalidate the recommendations of National Teaching Standards given that one of their presumptions is that factual knowledge is becoming increasingly less important. Most importantly, however, and in line with my initial hypothesis and with the assumption underlying National Teaching Standards, placing more emphasis on modern teaching practices in classrooms will indeed promote students' reasoning skills.[24]

[23]The low share of questions measuring reasoning skills in TIMSS also explains why the sizable estimated effect of modern teaching practices on these skills is not visible in the first three columns of Table 3. Recall that this is not an artefact of the TIMSS tests, but that the NAEP (and presumably other tests) are similar in their measurement of skills.

[24]Recent work by Carlsson, Dahl, and Roth (2012) shows that schooling raises crystallized but not fluid intelligence in Sweden. My finding that traditional teaching practices do not significantly affect fluid intelligence as measured by reasoning scores can be seen as corroborating this result if one assumes that like in the United States, traditional teaching practices dominate in Swedish classrooms.

5.2 Robustness of Results to Alternative Definitions of the Teaching Practice Indices

The way in which I constructed the traditional and modern teaching indices above can be criticized on several grounds. I address these criticisms here by considering several alternative definitions of the treatment variables and showing that the results are not sensitive to which exact definition is used. Table 4 presents estimates from variations of equation (1) that include these alternative definitions of the teaching practice indices. Results are reported for the overall test score in column (1), and for the knowing, applying, and reasoning test scores in columns (2), (3), and (4), respectively. All regressions in the table include the full set of teacher and class controls. Descriptive statistics of the treatment variables used here can be found in Table A2.

In panel (A) of Table 4, the traditional (modern) teaching index is defined as the proportion of answers among all the traditional (modern) teaching practices that fall into the "about half the lessons" or "every or almost every lesson" categories.[25] This redefinition addresses the concern that the results might be driven by the above indices' linearization of the ordinal scale on which teaching practices are measured. The results from these specifications are quantitatively similar to the ones found in Table 3. The estimated coefficient of 0.271 on the traditional teaching index in column (1) implies that moving a student from the 20th to the 80th percentile of the traditional teaching index is associated with a 5.7% of a standard deviation increase in her overall test score. Similarly, the estimated coefficient of 0.231 on the modern teaching index in column (4) implies that moving a student from 20th to 80th percentile of the modern teaching index is associated with a 6.0% of a standard deviation increase in her reasoning score.

In panel (B) of Table 4, the treatment variable is the normalized "share" of traditional teaching, which is defined as the traditional teaching index divided by the sum of the traditional and the modern teaching indices. The attractive feature of this treatment variable is that it forces a trade-off between traditional and modern teaching practices. Its major drawback is that it is based on the division of variables measured on an ordinal scale. The results from these specifications are qualitatively in line with those found in Table 3. In particular, the "share" of traditional teaching is estimated to have a positive effect on the overall, knowing, and applying scores, and a negative effect on the reasoning score. The estimated size of the effect is such that a ten percent increase in the "share" of traditional teaching is associated with a 5.6% (5.3%) of a standard increase in students' knowing (applying) score, and with a 5.0% of a standard deviation decrease in students' reasoning score. However, these effects are imprecisely

[25] Defining the indices as the proportion of answers falling into the "every or almost every lesson" category, or as the proportion of answers falling into any category other than "never", gives qualitatively similar results which are available upon request.

estimated and none of the coefficients is statistically significant.

Finally, the categorization of some of the individual teaching practices as traditional or modern may be criticized. For example, as discussed above, working problems could potentially be considered a modern teaching practice. I address this concern by excluding individual teaching practices from the teaching practice indices one at a time and re-running the regressions. This ensures that no single teaching practice (which might be miscategorized) drives the results. Table A3 shows that the results from these specifications are quantitatively similar to those found in Table 3. Results are reported for the overall test score in column (1), and for the knowing, applying, and reasoning test scores in columns (2), (3), and (4), respectively. All regressions in the table include the full set of teacher and class controls.

In conclusion, the results from specifications using alternatively defined teaching practice indices are consistent with the evidence provided in Table 3. That is, the results do not depend on the exact way that the indices are defined. I will therefore continue using the initial indices in the rest of the paper.

5.3 Heterogeneous Effects by Subject

The results reported so far assume that the effects of teaching practices on cognitive skills are identical in math and science. I now relax this assumption by estimating variations of the model in equation (1) that include as regressors interaction terms between the subject dummy and the two teaching practice indices. Table 5 presents the results from these regressions.

Focusing first on the results for math, the estimates in columns (1)-(4) are qualitatively similar to those found in the headline regressions. The estimated coefficient on the traditional teaching index is somewhat larger than the one found in Table 3 in columns (1)-(3), and is negative (though not significant) in column (4). The estimated coefficient on the modern teaching index is four times (about three times) as large in column (4) as it is in column (2) (in column (3)). Turning to the results for science, the estimates are again qualitatively similar to those found in Table 3. In particular, the largest estimate of the coefficient on the traditional teaching index is found in column (3), while the largest estimate of the coefficient on the modern teaching index is found in column (4). Note that due to the more demanding specifications used here, standard errors are generally larger than in Table 3 and not all of the discussed coefficients are statistically significant. This also means that the hypothesis of equal effects of the traditional and modern teaching indices in a particular subject cannot be rejected in most cases. To summarize, the main takeaway from Table 5 is that the effects of teaching practices on cognitive skills are qualitatively similar for both subjects, with the effects for math being quantitatively larger and driving the headline results.[26]

6 International Evidence

The debate whether traditional or modern teaching practices are better for student learning is not exclusive to the United States. In England, for example, Education Secretary Michael Gove has been calling for a greater emphasis on traditional teaching practices in schools.[27] In contrast, a recent reform in Israel calls for a reduction in traditional teaching practices at the post-primary level (Lavy 2011). In this section, I therefore extend my analysis of the effects of teaching practices on cognitive skills to education systems in other countries by exploiting the international dimension of TIMSS. As comparison countries, I select nine advanced economies (as defined by the International Monetary Fund) from the TIMSS database: three Anglo-Saxon countries (Australia, England, and Scotland), Israel, and five East and Southeast Asian countries (Hong Kong, Japan, Singapore, South Korea, and Taiwan). For each of these countries, I define the sample in the same way as I did for the United States above. The choice of countries as well as further issues regarding sample selection are discussed in more detail in the online appendix to this paper.

I estimate variations of the model in equation (1) for each country individually and for two groups of countries. The first group is made up of countries with a "Western"-style education system (the three Anglo-Saxon countries and Israel), while the second group comprises the five East and Southeast Asian countries. This categorization reflects the idea that educational production, including the effects of teaching practices, might work differently in the rarely studied Asian education systems. Table 6 presents the results from the regressions for these two groups of countries.[28] The dependent variable is the overall test score in column (1), and the knowing, applying, and reasoning test score in columns (2), (3), and (4), respectively. All regressions include the full set of teacher and class controls as well as country-by-subject dummies in order to allow for systematic differences in achievement levels in a given subject across countries.

Panel (A) of Table 6 shows that the results from the sample of three Anglo-Saxon countries and Israel are qualitatively similar to those obtained for the United States. The traditional teaching index is estimated to have a large and highly significant effect on both knowing and applying scores, and

[26]I also investigated whether the effects of teaching practices differ by students' gender or socioeconomic status, but found no evidence of this kind of heterogeneity.

[27]See, for example, the article "Tough exams and learning by rote are the keys to success, says Michael Gove" published in the *The Guardian* on November 13, 2012.

[28]For the sake of brevity, results from regressions for individual countries are not presented in the paper. These results are available upon request.

a small negative effect on reasoning scores, which is however not statistically significant. In contrast, the estimated coefficient on the modern teaching index is negative and not statistically significant in columns (2) and (3), and positive and highly significant in column (4). Importantly, the hypothesis that the coefficients on the two treatment variables are equal can now be rejected in all of the specifications. Panel (B) shows that the results from the sample of East and Southeast Asian countries are quantitatively similar to those obtained for the United States. For the sake of brevity, I do not discuss these results in further detail here.

Panel (C) of Table 6 presents estimation results from the pooled samples of all countries including the United States. The estimates are quantitatively similar to the ones obtained in Table 3.[29] In particular, the coefficient of 0.129 on the traditional teaching index in column (1) implies that moving a student from the 20th to the 80th percentile of the traditional teaching index in this sample is associated with a 6.5% of a standard deviation increase in her overall test score (Table 3: 5.4%). Similarly, the coefficient of 0.108 on the modern teaching index in column (4) implies that moving a student from the 20th to the 80th percentile of the modern teaching index in this sample is associated with a 7.5% of a standard deviation increase in her reasoning score (Table 3: 4.8%). Due to the large sample size, the effects are very precisely estimated such that even relatively modest impacts of the modern teaching index in columns (2) and (3), and of the traditional teaching index in column (4) can be excluded.

The finding that the effects of traditional and modern teaching practices are very similar across education systems is highly interesting and reassuring because it proves the external validity of the results obtained for the United States.

7 Conclusion

The question whether traditional or modern teaching practices are better for student learning in schools has long been debated by educational researchers, teachers, and parents. Very recently, economists have become interested in the issue. Their empirical studies show that teachers who emphasize traditional teaching practices are associated with higher test scores, while the evidence for modern teaching practices is less clear. These results cast doubt on the recommendations of National Teaching Standards in the United States, which call for a shift in emphasis from traditional towards modern teaching practices in

[29]Results from individual country regressions are also qualitatively similar. In particular, in seven out of ten (ten out of ten) countries is the estimated coefficient on the traditional teaching index greater than the estimated coefficient on the modern teaching index in regressions where the dependent variable is the knowing (applying) score. Moreover, in eight out of ten countries is the estimated coefficient on the traditional teaching index smaller than the estimated coefficient on the modern teaching index when the dependent variable is the reasoning score.

classrooms.

In this paper, I show that traditional and modern teaching practices promote different cognitive skills in students. In particular, traditional teaching practices promote the knowledge of basic facts and procedures, skills that have traditionally been emphasized in schools. In contrast, modern teaching practices promote reasoning, which is the skill that National Teaching Standards want to foster because it is perceived to be increasingly important in the labor market. I provide evidence that only a small fraction of the questions in standardized tests, both in TIMSS and elsewhere, measure students' reasoning skills. This explains the smaller estimates of the effects of modern teaching practices than for traditional teaching practices found in the literature.

The results in this paper have important implications for education policy. First, they show that teaching practices are a sizable determinant of student learning. This implies that instructing teachers to teach in a certain way is a potentially very cost-effective policy to increase a certain set of skills among students. Second, and relatedly, the results in this paper imply that if policymakers and educators are serious about promoting reasoning over rote in schools, standardized tests need to be adapted to measure reasoning skills. Otherwise, teachers, whose salary nowadays is often related to their students' performance on these kind of tests, have no incentive to employ modern teaching practices and to thus instill reasoning skills in their students.

References

Algan, Yann, Cahuc, Pierre, and Andrei Shleifer (2012). "Teaching Practices and Social Capital." *American Economic Journal: Applied Economics* forthcoming.

Carlsson, Magnus, Dahl, Gordon B., and Dan-Olof Rooth (2012). "The Effect of Schooling on Cognitive Skills." IZA Discussion Paper No. 6913.

Cattell, Raymond B. (1971). *Abilities: Their Structure, Growth and Action.* Boston (MA): Houghton-Mifflin.

Clotfelter, Charles T., Ladd, Helen F., 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):655-81.

Cuban, Larry (1993). *How Teachers Taught: Constancy and Change in American Classrooms 1890-1990.* New York (NY): Teachers College Press (2nd ed.).

Dee, Thomas S. (2007). "Teachers and the Gender Gaps in Student Achievement." *Journal of Human Resources* 42(3):528-54.

Foy, Pierre, Galia, Joseph, and Isaac Li (2008). "Scaling the Data from the TIMSS 2007 Mathematics and Science Assessments", in: Olson, John F., Martin, Michael O., and Ina V.S. Mullis (eds.):

TIMSS 2007 Technical Report. Published by the TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Goldhaber, Dan D., and Dominic J. Brewer (1997). “Why Dont Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity.” *Journal of Human Resources* 32(3):50523.

Kane, Thomas J., Taylor, Eric S., Tyler, John H., and Amy L. Wooten (2011). “Identifying Effective Classroom Practices Using Student Achievement Data.” *Journal of Human Resources* 46(3):587-613.

Lavy, Victor (2011). “What Makes an Effective Teacher? Quasi-Experimental Evidence.” NBER Working Paper No. 16885.

National Commission on Excellence in Education (1983). *A Nation At Risk: The Imperative for Educational Reform.* Report, Washington, D.C.

NCTM (1989). *Curriculum and Evaluation Standards for School Mathematics.* Report, National Council of Teachers of Mathematics, Reston (VA).

NCTM (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.: The National Academies Press.

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., O’Sullivan, Christine Y., Arora, Alka, and Erberber, Ebru (2007). “Developing the TIMSS 2007 Mathematics and Science Assessments and Scoring Guides”, in: Olson, John F., Martin, Michael O., and Ina V.S. Mullis (eds.): *TIMSS 2007 Technical Report.* Published by the TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Smith, P. Sean, Banilower, Eric R., McMaho, Kelly C., 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).

Smith Neidorf, Teresa, Binkley, Marilyn, and Maria Stephens (2006). *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, Binkley, Marilyn, Gattis, Kim, Nohara, David, and Mark Schneider (2006). *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.

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):365-79.

Van Klaveren, Chris (2011). “Lecturing style teaching and student performance.” *Economics of Education Review* 30(4):729-39.

Zemelman, Steven, Daniels, Harvey, and Arthur Hyde (2005). *Best Practice. Today’s Standards for Teaching and Learning in America’s Schools.* Portsmouth (NH): Heinemann (3rd ed.).

Table 1
Traditional and Modern Teaching Practices in the TIMSS Student Questionnaire

Traditional Math Teaching Practices	Traditional Science Teaching Practices
We listen to the teacher give a lecture-style presentation.	We listen to the teacher give a lecture-style presentation.
We memorize formulas and procedures.	We memorize science facts and principles.
We work problems on our own.	We work problems on our own.
Modern Math Teaching Practices	**Modern Science Teaching Practices**
We work together in small groups.	We work in small groups on an experiment or investigation.
We explain our answers.	We give explanations about what we are studying.
We relate what we are learning in mathematics to our daily lives.	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)?" Answers are coded on a four-point scale, with 1 corresponding to "never", 2 to "some lessons", 3 to "about half the lessons", and 4 to "every or almost every lesson."

Table 2
Descriptive Statistics: Teaching Practices

	Mean	SD	p20	p80
Traditional teaching index	3.09	0.28	2.87	3.33
Listening to the teacher lecture	3.05	0.40	2.74	3.39
Memorizing facts, formulas, and procedures	3.05	0.40	2.73	3.40
Working routine problems	3.18	0.35	2.90	3.47
Modern teaching index	2.79	0.33	2.50	3.07
Working in small groups	2.69	0.62	2.15	3.25
Giving explanations	3.13	0.39	2.82	3.46
Relating what is learned to students' daily lives	2.54	0.40	2.20	2.88

Table 3
Estimates of the Relationship Between Teaching Practices and Students' Cognitive Skills

	Overall Test Score			Knowing	Applying	Reasoning
	(1)	(2)	(3)	(4)	(5)	(6)
Traditional teaching index	0.118**	0.114**	0.118**	0.152**	0.133**	0.022
	(0.040)	(0.040)	(0.040)	(0.047)	(0.045)	(0.049)
Modern teaching index	0.026	0.027	0.024	0.007	0.001	0.085*
	(0.040)	(0.039)	(0.040)	(0.037)	(0.042)	(0.038)
Teacher controls		Y	Y	Y	Y	Y
Class controls			Y	Y	Y	Y
Number of students	6,057	6,057	6,057	6,057	6,057	6,057
Average *R*-squared	0.924	0.924	0.924	0.857	0.853	0.824
H_0: Traditional = modern (*p* value)	0.142	0.152	0.130	0.018	0.044	0.363

Notes: The dependent variable is the standardized overall test score in columns (1) to (3) and the standardized cognitive-domain specific test score in columns (4) to (6). Teacher and class controls as reported in Table A1. All regressions include subject dummies and are run five times (once for each plausible value). The average *R*-squared from the five regressions is reported in the second to last row. Standard errors which adjust for the imputation variance and which are robust to clustering at the class level are reported in parentheses. ~/*/** denote statistical significance at the 10/5/1 percent level.

Table 4

Robustness of Results to Alternative Measurements of Teaching Practices

	Overall	Knowing	Applying	Reasoning
	(1)	(2)	(3)	(4)
	(A) Proportion of answers 3 and above			
Traditional teaching index	0.271**	0.316**	0.284**	0.087
	(0.100)	(0.112)	(0.103)	(0.109)
Modern teaching index	0.042	0.021	-0.023	0.231*
	(0.085)	(0.085)	(0.092)	(0.090)
Average *R*-squared	0.924	0.857	0.853	0.824
H_0: Traditional = modern (*p* value)	0.107	0.044	0.034	0.334
	(B) Normalized "share" of traditional teaching			
Traditional / (traditional + modern)	0.292	0.558	0.531	-0.504
	(0.358)	(0.343)	(0.383)	(0.394)
Average *R*-squared	0.924	0.856	0.853	0.824

Notes: The regressions in columns (1) to (4) are variations of the specifications in columns (3) to (6) in Table 3. In panel (A), the treatment variables are the proportions of answers among all the teaching practices included in each index that fall into the "about half the lessons" or "every or almost every lesson" categories. In panel (B), the treatment variable is the traditional teaching index divided by the sum of the traditional and modern teaching indices. The number of students is 6,057 in all regressions. See the notes to Table 3 for additional controls included in all of the specifications. Standard errors which adjust for the imputation variance and which are robust to clustering at the class level are reported in parentheses. ~/*/** denote statistical significance at the 10/5/1 percent level.

Table 5
Teaching Practices and Students' Cognitive Skills by Subject

	Overall	Knowing	Applying	Reasoning
	(1)	(2)	(3)	(4)
Traditional teaching index: math	0.151**	0.222**	0.136*	-0.039
	(0.045)	(0.067)	(0.056)	(0.057)
Modern teaching index: math	0.037	0.016	0.024	0.064
	(0.040)	(0.056)	(0.055)	(0.056)
Traditional teaching index: science	0.088~	0.073	0.148*	0.081
	(0.052)	(0.060)	(0.069)	(0.065)
Modern teaching index: science	0.028	0.039	-0.023	0.073
	(0.046)	(0.051)	(0.059)	(0.048)
Average *R*-squared	0.924	0.857	0.854	0.824
H_0: Traditional = modern: math (*p* value)	0.091	0.002	0.117	0.201
H_0: Traditional = modern: science (*p* value)	0.454	0.726	0.133	0.929

Notes: The regressions in columns (1) to (4) are variations of the specifications in columns (3) to (6) in Table 3 in which the treatment variables are interacted with a subject dummy. The number of students is 6,057 in all regressions. See the notes to Table 3 for additional controls included in all of the regressions. Standard errors which adjust for the imputation variance and which are robust to clustering at the class level are reported in parentheses. ~/*/** denote statistical significance at the 10/5/1 percent level.

Table 6
Teaching Practices and Students' Cognitive Skills: International Evidence

	Overall	Knowing	Applying	Reasoning
	(1)	(2)	(3)	(4)
	(A) Other Anglo-Saxon countries + Israel			
Traditional teaching index	0.142**	0.230**	0.153**	-0.033
	(0.030)	(0.047)	(0.046)	(0.048)
Modern teaching index	-0.012	-0.050	-0.047	0.109**
	(0.026)	(0.032)	(0.036)	(0.036)
Average *R*-squared	0.925	0.850	0.852	0.822
H_0: Traditional = modern (*p* value)	0.000	0.000	0.002	0.041
	(B) East and Southeast Asian countries			
Traditional teaching index	0.117**	0.153**	0.129**	0.031
	(0.030)	(0.044)	(0.046)	(0.060)
Modern teaching index	0.052~	0.020	0.050	0.126**
	(0.030)	(0.040)	(0.035)	(0.039)
Average *R*-squared	0.923	0.849	0.861	0.829
H_0: Traditional = modern (*p* value)	0.207	0.059	0.272	0.259
	(C) All countries pooled			
Traditional teaching index	0.129**	0.186**	0.141**	0.004
	(0.018)	(0.025)	(0.024)	(0.030)
Modern teaching index	0.016	-0.014	-0.002	0.108**
	(0.017)	(0.021)	(0.022)	(0.024)
Average *R*-squared	0.924	0.850	0.857	0.826
H_0: Traditional = modern (*p* value)	0.000	0.000	0.000	0.014

Notes: The regressions in columns (1) to (4) are variations of the specifications in columns (3) to (6) in Table 3 run on 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 3. Standard errors which adjust for the imputation variance and which are robust to clustering at the class level are reported in parentheses. ~/*/** denote statistical significance at the 10/5/1 percent level.

Table A1

Descriptive Statistics: Teacher and Class Characteristics

	Mean	SD
Teacher characteristics		
Teacher is female	0.63	0.48
Teacher is aged 30-39	0.29	0.45
Teacher is aged 40-49	0.25	0.43
Teacher is older than 50	0.28	0.45
Teacher has teaching certificate	0.97	0.17
Teacher holds postgraduate degree	0.59	0.49
Teacher majored in subject taught	0.71	0.45
Teacher has 1-2 years of experience	0.11	0.31
Teacher has 3-5 years of experience	0.17	0.38
Teacher has > 5 years of experience	0.72	0.45
Class characteristics		
Class size	24.50	6.74
Teaching time per week (minutes)	239.21	72.45

Table A2
Additional Descriptive Statistics

	Mean	SD	p20	p80
	(A) Treatments used in Table 4			
Proportion of answers 3 and above				
Traditional teaching index	0.71	0.12	0.61	0.82
Modern teaching index	0.58	0.15	0.45	0.71
Modern / (modern + traditional)	0.47	0.03	0.45	0.50
	(B) Treatments used in Table 6			
Other Anglo-Saxon countries + Israel				
Traditional teaching index	2.91	0.34	2.63	3.20
Modern teaching index	2.62	0.31	2.36	2.88
East and Southeast Asian countries				
Traditional teaching index	2.86	0.28	2.64	3.08
Modern teaching index	2.46	0.39	2.11	2.83
All countries pooled				
Traditional teaching index	2.90	0.31	2.66	3.16
Modern teaching index	2.55	0.38	2.20	2.89

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	0.128 **	0.163 **	0.147 **	-0.009
	(0.034)	(0.040)	(0.035)	(0.047)
Modern teaching index	0.032	0.018	0.011	0.090 *
	(0.039)	(0.037)	(0.042)	(0.037)
Traditional teaching index excluding memorizing	0.107 ~	0.127 *	0.115 *	-0.015
	(0.060)	(0.055)	(0.056)	(0.062)
Modern teaching index	0.041	0.029	0.020	0.089 *
	(0.039)	(0.037)	(0.042)	(0.036)
Traditional teaching index excluding working problems	0.142 **	0.195 **	0.169 **	-0.008
	(0.037)	(0.060)	(0.051)	(0.058)
Modern teaching index	0.027	0.010	0.004	0.090 *
	(0.040)	(0.037)	(0.042)	(0.037)
Traditional teaching index	0.132 **	0.184 **	0.148 **	0.025
	(0.041)	(0.051)	(0.046)	(0.050)
Modern teaching index excluding working in groups	-0.023	-0.075 ~	-0.037	0.039
	(0.038)	(0.044)	(0.040)	(0.038)
Traditional teaching index	0.120 **	0.151 **	0.132 **	0.034
	(0.039)	(0.048)	(0.045)	(0.048)
Modern teaching index excluding explaining answers	0.027	0.029	0.013	0.076 *
	(0.029)	(0.031)	(0.036)	(0.031)
Traditional teaching index	0.117 **	0.149 **	0.131 **	0.029
	(0.039)	(0.048)	(0.045)	(0.048)
Modern teaching index excluding relating to students' lives	0.036	0.030	0.012	0.072 *
	(0.033)	(0.029)	(0.032)	(0.032)

Notes: The regressions in columns (1) to (4) are variations of the specifications in columns (3) to (6) in Table 3 in which one teaching practice is excluded from the treatment variables at a time. The total number of regressions underlying these results is 24 (six teaching practices to exclude times four dependent variables). The number of students is 6,057 in all regressions. See the notes to Table 3 for additional controls included in all of the regressions. Standard errors which adjust for the imputation variance and which are robust to clustering at the class level are reported in parentheses. ~/*/** denote statistical significance at the 10/5/1 percent level.