Is Traditional Teaching really all that Bad? A Within-Student Between-Subject Approach

GUIDO SCHWERDT
AMELIE C. WUPPERMANN

CESIFO WORKING PAPER NO. 2634
CATEGORY 5: ECONOMICS OF EDUCATION
APRIL 2009

CESifo *Working Paper No. 2634*

Is Traditional Teaching really all that Bad? A Within-Student Between-Subject Approach

Abstract

Recent studies conclude that teachers are important for student learning but it remains uncertain what actually determines effective teaching. This study directly peers into the black box of educational production by investigating the relationship between lecture style teaching and student achievement. Based on matched student-teacher data for the US, the estimation strategy exploits between-subject variation to control for unobserved student traits. Results indicate that traditional lecture style teaching is associated with significantly higher student achievement. No support for detrimental effects of lecture style teaching can be found even when evaluating possible selection biases due to unobservable teacher characteristics.

JEL Code: I21, C23.

Keywords: teaching practices, educational production, TIMSS, between-subject variation.

Guido Schwerdt
Ifo Institute for Economic Research at the University of Munich
Poschingerstrasse 5
81679 Munich
Germany
schwerdt@ifo.de

Amelie C. Wuppermann
University of Munich
Kaulbachstrasse 45
80539 Munich
Germany
amelie.wuppermann@lrz.uni-muenchen.de

Preliminary Version
April 15, 2009
This research has profited from comments by seminar participants at the University of Munich, EDGE meeting in Copenhagen and the ifo Institute. We received insightful comments from Erik Hanushek, Ludger Wößmann, Joachim Winter, Florian Heiß and Jane Cooley to whom our gratitude goes.

1 Introduction

Recent studies stress the importance of teachers for student learning. However, the question, what actually determines teacher quality, i.e. what makes one teacher more successful in enhancing her students' performance than another, has not been settled so far (Aaronson et al., 2007). Different categories of teacher variables have been analyzed. Some studies focus on the impact of a teacher's gender and race on teacher quality (Dee, 2005, 2007). Others try to uncover the relationship between student outcomes and teacher qualifications such as teaching certificates, other paper qualifications or teaching experience (Kane et al., 2008). Such observable teacher characteristics are, however, generally found to have only little impact on student achievement and can only explain a relatively small part of overall teacher quality (Aaronson et al., 2007; Rivkin et al., 2005). Most of the variation in teacher quality can be attributed to unobserved factors.[1]

While most of these studies focus on characteristics of the teacher, this paper directly peers into the black box of educational production by focusing on the actual teaching process. More specifically, we contrast traditional lecture style teaching with teaching based on in-class problem solving and investigate the impact on student achievement. Lecture style teaching is often regarded as old-fashioned and connected with many disadvantages: Lectures fail to provide instructors with feedback about student learning and rest on the presumption that all students learn at the same pace. Moreover, students' attention wanes quickly during lectures and information tends to be forgotten quickly when students are passive. Finally, lectures emphasize learning by listening, which is a disadvantage for students who prefer other learning styles. Alternative instructional practices based on active and problem-oriented learning presumably do not suffer from these disadvantages. National standards (NCTM, 1991; National Research Council, 1996) consequently advocate engaging students more in hands-on learning activities and group work. Despite these recommendations traditional lecture and textbook methodologies continue to dominate science and mathematics instruction in US middle schools (Weiss, 1997). This raises the question whether the high share of total teaching time devoted to traditional lecture style presentations has a detrimental effect on overall student learning.

By addressing this question, this study adds to the literature analyzing the impact of teaching process variables such as teaching practices on student outcomes.[2] Despite

[1] This finding led researchers, concerned with providing recommendations for recruitment policies and designing optimal teacher pay schemes, to suggest to identify effective teachers by their actual performance on the job using "value added" measures of student achievement (Gordon et al., 2006).

[2] For an overview see Goe (2007).

the importance of teaching practices for student performance as recognized by educational researchers (Seidel and Shavelson, 2007) and their potential relatively low-cost implementation, economists have only recently begun to analyze the impact of teaching methods on student achievement.[3] Various dimensions of teaching practices have been shown to be able to explain a large share of the between-teacher variation in student achievement (Schacter and Thum, 2004). However, to our knowledge no rigorous empirical analysis of the impact of lecture style teaching on overall student achievement exists.

To study the effect of lecture style teaching, we construct the share of effective teaching time, that is time in class devoted to either lecture style presentation or in-class problem solving, using information on in-class time use provided by teachers in the 2003 wave of the Trends in International Math and Science Study (TIMSS) in US schools. Estimating a reduced form educational production function and exploiting between-subject variation to control for unobserved student traits, we find that the choice of teaching practices matters for student achievement. We find that a 10 percentage point shift from problem solving to lecture style presentation results in an increase in student achievement of about 1 percent of a standard deviation.

This result is highly robust. Consistent with other studies in this literature, we find no evidence for significant effects of commonly investigated observable teacher characteristics such as teaching certificates or teaching experience. While we are able to control for a huge array of observable teacher traits, selection of teachers based on unobservable characteristics into teaching methods remains an issue. The bias resulting from potential selection of teachers with different unobservable attributes into different teaching methods is assessed following the technique pioneered in Altonji et al. (2005). The results indicate that only relatively low selection on unobservables compared to the selection on observables is necessary to explain the entire estimated effect. We would thus not formulate policy conclusions that call for more lecture style teaching in general. However, a negative causal effect of lecture style presentation that is hidden in our results due to selection based on unobserved teacher traits can only exist if "good" teachers (teachers with favorable unobserved characteristics) predominately select themselves into an inferior teaching technique. This scenario, however, lacks any intuitive or theoretical support and thus appears extremely implausible. We therefore conclude that the high share of total teaching time devoted

[3]Brewer and Goldhaber (1997) and Aslam and Kingdon (2007) analyze the impact of many different teaching methods. Rouse and Krueger (2004), Banerjee et al. (2007), and Barrow et al. (2009) investigate the effectiveness of computer-aided instruction and Machin and McNally (2008) analyze an education policy that changed reading instruction.

to traditional lecture style teaching in science and mathematics instruction in US middle schools has no detrimental effect on overall student learning. This finding implies that attempts to reduce the amount of traditional lecture style teaching in US middle schools have little potential for raising overall achievement levels.

The remainder of the paper is structured as follows: the following section reviews the literature on teaching practices. Section 3 presents the data. Section 4 describes the estimation strategy. Headline results are presented in Section 5, while Section 6 provides a sensitivity analysis. Section 7 concludes.

2 Literature on Teaching Practices

A majority of studies considering teaching practices do not focus on a specific teaching practice. They rather analyze the relationship between a teacher's evaluation score on a standard-based teacher evaluation system and student achievement.[4] Most of these studies find that evaluation scores are correlated with student achievement. Similarly, Jacob and Lefgren (2008) analyze the relationship between the school principal's evaluation of a teacher and the part of actual achievement gain students have because they are taught by this teacher. The authors also find a relationship between the evaluation and teacher effectiveness. The different evaluation schemes certainly measure a part of teacher quality. Nevertheless, when analyzing the relationship between an evaluation score and student achievement it is unclear, which part of the evaluated practices is (most) important for the student outcome.

This problem also arises in some other studies that look at the impact of different categories of practices on student achievement. Smith et al. (2001) analyze if didactic or interactive teaching methods are more effective in teaching elementary school children. They find that interactive teaching is associated with higher gains in test scores. McGaffrey et al. (2001) and Cohen and Hill (2000) analyze whether students have higher test scores in math if their teacher uses methods in accordance with a teaching reform promoted by the National Science Foundation. Machin and McNally (2008) analyze the effect of the introduction of the "literacy hour" in English primary schools in the late 1990s by the British government. This policy intervention changed the practice of teaching primary students how to read. Using the fact that not all schools started the literacy hour at the same point in time in a difference-in-difference framework, the authors show that

[4]For example Borman and Kimball (2005), Gallagher (2004), Heneman et al. (2006), Holtzapple (2003), Kimball et al. (2004), Milanowski (2004), Matsumura et al. (2006) and Schacter and Thum (2004).

the literacy hour significantly increased reading skills for low ability student while high ability students were not affected. Again, didactic and interactive methods and traditional practices are measured at an aggregated level encompassing different teaching practices. The authors estimate an effect of a teaching style but not of a single teaching practice.

Most studies that analyze specific practices either focus on the impact of specific computer programs or analyze the effects of assignments and grading practices. The impact of computer-aided instruction on student achievement is analyzed by several studies that use random assignment to computer instruction as variation for identification. Rouse and Krueger (2004) analyze 3rd to 6th grade students in four schools in a US urban school district who were randomly assigned to learn with a computer program designed to increase language and reading skills. They do not find large significant effects of this program on literacy. Banerjee et al. (2007), on the other hand, find significant effects of a computer-assisted instruction program on achievement in math for primary school students in India. Barrow et al. (2009) also find significant effects of computer-aided instruction on student achievement in math. Classes in three districts in the US were randomly assigned to study math either in the computer lab with a special computer-aided instruction program for algebra and pre-algebra or in traditional classrooms with "chalk and talk" methods. The authors find that students in the computer lab classes did significantly better than students in the traditional classes in terms of achievement gains. Larger classes, classes with higher rates of absenteeism and more heterogeneous classes seem to benefit more from computer-aided instruction. This is interpreted as evidence that the individualization of instruction is the mechanism through which computer-aided instruction raises achievement.

Matsumura et al. (2002) look at the effect the quality of assignments has on student achievement. Using hierarchical linear modeling they find that a small part of student test score variance can be predicted by assignment quality. The relationship between assignments and student achievement is also analyzed by Newmann et al. (2001). The authors find that more intellectually challenging assignments are related to higher gains in test scores. Wenglinsky (2000, 2002) uses multilevel structural equation modeling to analyze the impact of different teaching practices on student test scores in math and science. He finds that the use of hands-on learning activities like solving real world problems and working with objects, an emphasis on thinking skills, and frequent testing of students are positively related to students' test scores taking into account student background and prior performance. Some evidence for the effectiveness of frequent student assessment is also found by Kannapel et al. (2005): High-performing high-poverty schools in Kentucky payed more attention to student assessment than other high-poverty schools. Bonesrønning (2004) looks at a different aspect of student assessment. He analyzes if grading practices affect student achievement in Norway and finds evidence that easy grading deteriorates student achievement.

More closely related to this paper in terms of the teaching practices analyzed and identification strategy are the analyses by Brewer and Goldhaber (1997) and Aslam and Kingdon (2007). Brewer and Goldhaber (1997) estimate different specifications of education production functions for tenth grade students in math with data from the National Educational Longitudinal Study of 1988. They conclude that teacher behavior is important in explaining student test scores. Controlling for student background, prior performance and school and teacher characteristics, they find that instruction in small groups and emphasis on problem solving lead to lower student test scores.

Aslam and Kingdon (2007) analyze the impact of different teaching process variables on student achievement in Pakistan. Their identification strategy rests on within pupil across subject (rather than across time) variation, which is similar to the identification strategy employed in this analysis. They find that students taught by teachers who spend more time on lesson planning and by teachers who ask more questions in class have higher test scores.

Similarly to Brewer and Goldhaber (1997) and Aslam and Kingdon (2007), this study focuses on the impact of a single teaching practice rather than a general teaching style. As in Brewer and Goldhaber (1997) problem solving is included in the analysis. But it is not taken as the mainly analyzed teaching practice. Instead we focus on the effect of spending time on lecture style presentation compared to time spent on problem solving. Since lecture style presentation and problem solving could be classified as belonging to different teaching styles this study also relates to other literature that compares the effects of different teaching styles.

3 Data

The data used in this study is the 2003 wave of the Trends in International Math and Science Study (TIMSS). In this study we focus on country information for the US. In TIMSS, students in 4th grade and in 8th grade were tested in math and science. We limit our analysis to 8th grade students, because 4th grade students are typically taught by one teacher in all subjects.

We standardize the test scores for each subject to be mean 0 and standard deviation 1.

In addition to test scores in the two tested subjects, the TIMSS data provide background information on student home and family. For the purpose of this analysis it is crucial that TIMSS allows linking students to teachers. Each student's teachers in math and science are surveyed on their characteristics, qualifications and teaching practices. Additionally, school principals provide information on school characteristics.

The key variable of interest in this paper is derived from question 20 in the teacher questionnaires in the 2003 wave of TIMSS. Unfortunately, this precise question was not asked in previous waves of TIMSS. We therefore limit our analysis to the 2003 wave. Teachers are asked in 2003 to report what percentage of time in a typical week of the specific subject's lessons students spend on certain in-class activities. These activities include reviewing homework, listening to lecture style presentation, working on problems with the teacher's guidance, working on problems without guidance, listening to the teacher reteach and clarify content, taking tests or quizzes, classroom management and other activities. Out of these 8 categories, we classify listening to lecture style presentation and working on problems with and without guidance as effective teaching time. Effective teaching time is meant to proxy for time in which students are taught new material. The percentage of time spent on effective teaching is included as a control in all specifications.

The primary interest of this study is to contrast time devoted to lecture style presentation with in-class teaching time used for problem solving. The latter includes both students solving problems on their own and solving problems with teachers' guidance. We therefore construct a variable measuring the share of effective teaching time devoted to lecture style presentation. An increase in this variable of 0.1 indicates that 10 percentage points of total effective teaching time are shifted from teaching based on problem solving to lecture style presentation.

The TIMSS 2003 US data set contains student-teacher observations on 8,912 students in 232 schools. 41 of those students have more than one teacher in science. These students are not included in the estimation sample. 8,871 students in 231 schools in 455 math classes taught by 375 different math teachers and in 1,085 science classes taught by 475 different science teachers remain in the sample. Not all of the students and teachers completed their questionnaires. In order not to loose a large amount of observations we impute missing values of all control variables and include indicators for imputed values in all estimations.[5]

[5]Experimenting with different imputation procedures revealed that our main results do not depend on the method of imputation. Main results remain also qualitatively unchanged when simply deleting observations with missing values. Results presented in the paper are based on a simple mean-imputation procedure.

2,561 students have, however, missing information on our teaching practice variable of interest. These observation are dropped from the analysis. 6,310 students in 205 schools with 639 teachers (303 math teachers and 355 science teachers, where 19 teachers teach both subjects) remain in the sample.

Furthermore due to the sampling design of TIMSS, students are not all selected with the same probability. A two stage sampling design makes it necessary to take probability weights into account when estimating summary statistics (Martin, 2005). All estimation results take the probability weights into account and allow for correlation between error terms within schools.[6]

Table 1 reports descriptive statistics on observable teacher characteristics separately for math and science teachers. Mean differences are reported in the last column of table 1. The share of teachers with math or science majors naturally differs between those two groups. Apart from mean differences in majors only a few other variables are significantly different between the two groups. The same is true for class characteristics. Mean differences between classes in science and math are reported in table A-1 in the appendix. To control for the potentially confounding impact of these differences in observable class and teacher characteristics, we include all variables presented in tables 1 and A-1 in our empirical analysis.

4 Estimation Strategy

To estimate the effect of effective teaching devoted to lecture style presentation we estimate a standard education production function:

$$Y_{ijk} = c_j + B'_{ijk}\beta_{1j} + S'_{ik}\beta_{2j} + T'_{ijk}\beta_{3j} + TP'_{ijk}\beta_{4j} + \epsilon_{ijk}. \tag{1}$$

The test score, Y_{ijk}, of student i in subject j in school k is explained by student background characteristics, B_{ijk}, school characteristics, S_{ik} and teacher characteristics, qualifications T_{ijk} and a vector TP_{ijk} including teaching process variables. This analysis limits its focus on two process variables: the share of effective teaching time devoted to lecture style presentation and the share of total time spent on effective teaching. The error term, ϵ_{ijk}, contains all unobservable influences on student test scores. In particular, it contains the effects of unobservable student, μ_i , teacher, ξ_j, and school characteristics, ν_k:

[6] In addition, the two step procedure of sampling could be incorporated in the estimation of standard errors. For simplicity, we ignore the latter in the following analysis which then gives us conservative estimates of the standard errors.

$$\epsilon_{ijk} = \mu_i + \xi_j + \nu_k + \psi_{ijk} \tag{2}$$

Estimating equation (1) by ordinary least squares produces biased estimates if unobserved school characteristics, ν_k, and TP_{ijk} are correlated. This could be the case if the choice of a particular teaching practice is partly determined by the school and if there exists sorting of high ability students and effective teachers into schools.

To eliminate the effects of between-school sorting, we use school fixed effects, s_k, to exclude any systematic between-school variation in performance or teaching practice, whatever its source:

$$Y_{ijk} = c_j + B'_{ik}\beta_{1j} + s_k + T'_{ijk}\beta_{3j} + TP'_{ijk}\beta_{4j} + \mu_i + \xi_j + \psi_{ijk}. \tag{3}$$

The estimates produced by equation (3) could still be biased by within-school sorting wherever schools have more than one class per subject per grade. We therefore eliminate the influence on constant student traits by differencing between subjects:

$$\begin{aligned} \Delta Y_i \;=\; & c_m - c_s + B'_i(\beta_{1m} - \beta_{1s}) + S'_i(\beta_{2m} - \beta_{2s}) \\ & + T'_{im}\beta_{3m} - T'_{is}\beta_{3s} + TP'_{im}\beta_{4m} - TP'_{is}\beta_{4s} + \eta_i \end{aligned} \tag{4}$$

where $\Delta Y_i = Y_{im} - Y_{is}$ and $\eta_i = \xi_m - \xi_s + \psi_{im} - \psi_{is}$.

In our headline specification we follow Dee (2005, 2007) by assuming that coefficients for each variable are equal across the two subjects:[7]

$$\Delta Y_i \;= \Delta T'_i\beta_3 + \Delta TP'_i\beta_4 + \eta_i. \tag{5}$$

The estimate of the effect of teaching practice on student achievement produced by equation (5) is not biased due to between or within school sorting of students based on unobservable student traits. We do, however, have to make the identifying assumption that unobservable teacher characteristics that directly influence student achievement are not related to the choice of the teaching method. In other words, η_i is uncorrelated with all other right-hand side variables. This is a strong assumption and we, therefore refrain from interpreting β_4 as causal effect. We rather interpret β_4 as a measure for the link between a teaching practice and student achievement that is not driven by between or within school

[7] We do, however, estimate equation (4) as a robustness check.

sorting of students. It might, however, be partly driven by sorting of teachers into a special teaching method based on unobservable teacher traits.

We evaluate the concern of selection on unobservables by borrowing a procedure from Altonji et al. (2005) which allows to evaluate the bias of the estimate under the assumption that selection on unobservables occurs to the same degree as selection on observables. For the following, let *Lecture* denote our variable of interest: the percent of effective teaching time spent on giving lecture style presentations and β_4 its coefficient. As developed in the appendix in our application the asymptotic bias of $\widehat{\beta_4}$ is

$$Bias(\widehat{\beta_4}) = \frac{Cov(\Delta \widetilde{Lecture}, \eta)}{Var(\Delta \widetilde{Lecture})} = \frac{Cov(\Delta Lecture, \eta)}{Var(\Delta \widetilde{Lecture})} \tag{6}$$

where $\Delta \widetilde{Lecture}$ is the residual of a linear projection of $\Delta Lecture$ on all other control variables, now represented by T. The second equality holds if the other controls (T) are orthogonal to η. The condition that selection on unobservables is equal to selection on observables can be stated as

$$\frac{Cov(\Delta T'\beta_3, \Delta Lecture)}{Var(\Delta T'\beta_3)} = \frac{Cov(\Delta Lecture, \eta)}{Var(\eta)} \tag{7}$$

Equation (7) can be used to estimate the numerator of the bias of $\widehat{\beta_4}$, once we have consitent estimates for β_3. Under the assumptions that the true effect of lecture style teaching is zero and again that T is orthogonal to η, β_3 can be consistently estimated (see appendix).

The estimated bias displays the effect we would estimate even if the true effect was zero when selection on unobservables is as strong as selection on observables. In addition, we report the ratio of the estimated β_4 from equation (5) and the estimated bias giving a hint of how large selection on unobservables would have to be compared to selection on observables to explain the entire estimated effect. A value higher than one indicates that selection of unobservables needs to be stronger than selection on observables to explain the entire estimate, in case of a ratio lower than one already weaker selection on unobservables than on observables suffices to explain the entire estimate.

5 Results

Estimates of the effect of teaching practices based on the different methods advanced in Section 4 are presented in Tables 2 and 3. Each regression is performed at the level of the individual student and each of the estimations also takes into account the complex data structure produced by the survey design and the multi-level nature of the explanatory variables.

Table 2 reports results from estimating equation (1) and equation (3). We estimate both equations separately for math and science. Columns 1 and 3 present regressions results for math and science based on equation (1). These regressions include a complete set of student- and family-background variables, controls for teacher and class characteristics as well as characteristics of the school. Given the purpose of this study, only estimated coefficients for the teaching practice variable of interest and selected teacher characteristics are reported.

Our key variable of interest, the share of effective teaching time devoted to lecture style presentation, is estimated to have a positive impact on test scores in both subjects. In math the estimate is highly significant, while the estimate in science falls short of achieving statistical significance at any common significance level.

As discussed in the previous section, these results might be confounded by between school sorting based on unobservable characteristics of students. Column 2 and 4 therefore report estimation results based on equation (3), which includes school fixed effects. Lecture style presentation is now highly significant in science and the point estimate significantly increased compared to column 3. The estimate in math, however, did not change but lost its statistical significance due to increased standard errors.

To gain statistical power we pool both estimation samples and estimate equations (1) and (3) with the joint sample. This approach assumes that the effects of all right-hand side variables are identical in both subjects. Based on this estimation sample the relationship between more lecture style presentation and test scores is positive and significantly different from zero in both specifications.

The evidence presented in table 2 suggests a positive relation between more lecture style presentation and student achievement. However, within school selection of students based on unobservable student characteristics might drive this relationship. For instance, it is reasonable to assume that teachers adjust their teaching style according to class composition and average student ability. We therefore difference out unobserved constant student traits by taking between-subject differences of test scores and all right-hand side variables as presented in equation (5).

Table 3 presents estimation results of the between-subject differences approach. We start out with a very basic specification without further controls in column 1 and successively add more controls that vary between subjects to account for subject-specific differences. In particular column 6 presents our headline results. In this specification we additionally control for all observable teacher and class characteristics presented in tables 1 and A-1. It is quite astonishing to see that adding more control variables in the between-subject specification leaves our estimate for the effect of more lecture style teaching almost unchanged. Moreover, in contrast to all other control variables the share of lecture style teaching is estimated to be statistically significant throughout all specifications presented in table 3. However, the magnitude of the estimated coefficient decreased in comparison to the regression results presented in table 2. This indicates that within school sorting matters for the estimation of teachers' choice variables such as the degree of lecture style teaching. Estimates for the effect of more lecture style teaching on student learning in table 3 range from 0.13 to 0.1. Our headline estimate is reported in column 6 with an estimated size of 0.1. This parameter suggests that a 10 p.p. increase in the share of effective teaching time devoted to lecture style teaching is associated with an increase in student test scores of 1 percent of a standard deviation.

In turn, our results imply a negative correlation between more in-class problem solving and student achievement. This finding is consistent with evidence on instruction based on problem solving and student learning presented in Brewer and Goldhaber (1997). Moreover, time devoted to in-class problem solving is positively correlated with a categorial variable included in the TIMSS data that indicates the amount of group work done in class. This suggests a link between group work and student achievement, which would also be consistent with the results of Brewer and Goldhaber (1997) regarding instruction in small groups.

Furthermore, results presented in tables 2 and 3 indicate that for none of the other commonly investigated teacher characteristics a significant and robust impact on student achievement can be found. This is in line with previous findings in this literature and emphasizes the importance of the statistical significant relationship between more lecture style teaching and student achievement.

As pointed out in the previous section, estimates might still be biased due to selection of teachers into more (or less) lecture style teaching based on unobservable teacher characteristics. This concern is fostered by previous findings in the literature that emphasize the importance of unobservable teacher traits for student achievement. This raises the question: How can these results be interpreted?

The bias and ratio at the end of table 3 allow us to shed some light on the question of the influence of unobservables. The underlying assumption for the estimation of each bias is that selection on unobservables occurs to the same degree as selection on observables. As from left to right in table 3 the number of included controls increases, the potential for selection on unobservables increases and thus the estimated bias gets larger. In all columns the estimated bias is larger than the point estimate of the impact of lecture style teaching on student test scores. This is reflected in the ratios at the end of each column that are always smaller than one, indicating that selection on unobservables that is weaker than selection on observables suffices to explain the entire estimated coefficient. In our headline specification in column 6 selection on unobservables that is only 0.04 times as strong as selection on observables would explain the entire estimated coefficient given that the true effect is zero. On the one hand, we have included a great amount of control variables so that we believe that selection on unobservables is likely weaker than selection on observables. On the other hand only very little selection on unobservables compared to the selection on observables suffices to explain the entire effect. Given this uncertainty, we refrain from interpreting the results as evidence for a causal effect as the positive coefficient could also reflect selection of teachers with desirable unobserved characteristics into lecture style teaching.

This raises another question: Why would teachers with different desirable unobserved characteristics select different degrees of lecture style teaching? While a reduced form approach of educational production cannot mirror the full complexity of the choices involved in the teaching process, we are, nevertheless, able to pin down the relationship between potential selection based on unobserved teacher traits and the causal effect of lecture style teaching as our estimation approach eliminated all other likely biases. If no selection based on unobservable teacher traits exists, our estimates speak for a positive effect of lecture style teaching. Our estimates might, however, be biased upwards if teachers with desirable unobserved characteristics more frequently base their instruction on lectures. Theoretically, this selection bias could be large enough to hide a true negative effect of lecture style teaching, which would imply that teachers with desirable unobserved characteristics predominately select themselves into an inferior teaching practice. This scenario, however, lacks any intuitional or theoretical support. We thus argue that this scenario is highly implausible and can be excluded, which allows a conservative interpretation of our results:

We find no evidence for any detrimental effect of lecture style teaching on overall student learning.

6 Robustness Checks

This section tests the sensitivity of the results presented in section 5 with respect to other definitions of the lecture style variable and with respect to specifications allowing for heterogenous effects. The results of these robustness checks are presented in table 4.

As our grouping of the response categories available to the teacher in question 20 of the 2003 teacher questionnaires in TIMSS could be criticized, we provide evidence on the effect of interest based on different approaches to construct the lecture style variable. We test four alternative definitions of the lecture style variable with corresponding estimation results presented in each of the four columns of the upper panel of table 4.

In column 1 time spent re-teaching and clarifying content/procedures is included in effective teaching and in lecture style teaching. In column 2 effective teaching time includes taking tests or quizzes in addition to giving lecture style presentation and problem solving. Lecture style teaching in column 2 is defined in relation to the latter three. In column 3 we decompose the effective teaching time into its elements and separately control for each category. In column 4 lecture style is defined as the percent of overall time in class spent on giving lecture style presentation.

The coefficients in the upper panel of table 4 reveal that redefining our key variable of interest does not change the estimated impact of more lecture style teaching on student achievement. Only the estimate in the fourth column just falls short of achieving statistical significance (p-value .112). Note however, that in column four we directly compare lecture style teaching with all other possibilities of in-class time use. Naturally taking tests, reviewing homework and classroom management are very different aspects of the teaching process. A successful teaching strategy requires an optimal mix of all these categories. While the main purpose of this study is to analyze the teaching of new material by giving lecture style presentations rather than by letting pupils solve problems, it is reassuring to see that increasing the total amount of time in class devoted to lecture style presentations (without defining effective teaching time) is also associated with higher student achievement.

Additionally, we present evidence for various sub-samples in the middle panel of table 4. In column 1 and 2 we estimate equation (5) for pupils with the same peers in math and science and pupils with different peers, respectively. This distinction is motivated by the concern that the main effect might be driven by differences in the classroom composition. In the sub-sample with identical peers in both subjects our within-student between-subject identification strategy takes care of any potential peer effects. Note that, while both estimates lack significance, the estimated coefficient in column 1 even exceeds our headline estimate, while the estimate in the sub-sample with different peers decreases to .08. Hence, our headline estimate does not hide significantly different effects of lecture style teaching in these two sub-groups.

Column 3 and 4 of the middle panel of table 4 report estimates separately for students in schools where either no or both subjects are tracked by ability and for students in schools where tracking on ability exists in only one of the two subjects. This distinction is motivated by the consideration that tracking on ability might induce teachers to chose different degrees of lecture style teaching. The results indicate that the positive association between more lecture style teaching and student achievement is indeed less pronounced when looking at schools with identical tracking policies in both subjects. The point estimate is, however, again positive, but insignificant.

In the lower panel of table 4 we specifically investigate subject-specific effects. Column 1 and 2 presents estimation results from estimating versions of equation (4), where we abandon the assumption that coefficients for each right-hand side variable are equal across subjects. As all science variables enter negatively on both sides of equation (4), a negative coefficient for any variable in science masks a positive relationship between the variable and science test scores. All estimates thus have the expected signs. They are not statistically significant for science, though. We thus find evidence for a stronger effect of lecture style teaching in math.

In sum, we find positive relationships between more lecture style teaching and student achievement in all robustness analyses. The magnitude of the estimated effects varies between specifications and between sub-samples. The latter finding indicates that there exists indeed a substantial variation in the positive association between more lecture style teaching and student achievement. We also find no evidence for a detrimental effect of lecture style teaching.

7 Conclusion

Existing research on teacher quality allows two conclusions: First, there exists a large variation in teachers' ability to improve student achievement. Second, this variation cannot be explained by common, observable teacher characteristics. The results presented in this study confirm that these observable teacher characteristics have little potential for explaining the variation in student achievement. We provide, however, new evidence on a significant link between teaching practice and student achievement.

The specific teaching practice variable analyzed in this paper is the share of effective teaching time devoted to lecture style presentation (in contrast to in-class problem solving). We construct this variable based on information on in-class time use provided by teachers in the 2003 wave of the Trends in International Math and Science Study (TIMSS) in US schools. Exploiting between-subject variation to control for unobserved student traits and estimating a reduced form educational production function, we find that a 10 percentage point shift from problem solving to lecture style presentation results in an increase in student achievement of about one percent of a standard deviation. We further show that this result is extremely robust to definitional changes in the construction of the main variable of interest as well as to specifications allowing for heterogenous effects.

This finding suggests that students taught by teachers, who devote more effective teaching time to lecture style presentation rather than letting students solve problems on their own or with the teacher's guidance, learn more (in terms of competencies tested in the TIMSS student achievement test). This result stands in contrast to constructivist theories of learning. It is, however, in line with previous findings in the literature (Brewer and Goldhaber, 1997) showing that instruction in small groups and emphasis on problem solving lead to lower student test scores.

We emphasize, however, that our results demand a careful interpretation and need to be taken for what they are: Evidence for a positive association between more time devoted to lecture style teaching and student achievement that is neither driven by selection of particular students into schools or classes nor by selection of teachers based on various observable characteristics into a particular teaching method. However, selection based on unobservable teacher characteristics remains a worry. Following the method developed in Altonji et al. (2005), we show that only a relatively small selection based on unobservables suffices to explain the entire estimated coefficient. We thus refrain from formulating any policy conclusions that call for more lecture style teaching in general.

We are nevertheless able to draw an important conclusion about the nature of the causal effect of lecture style teaching on student achievement as we eliminated any potential biases arising from sorting of students, differences in schools and observable differences in teacher traits in our empirical approach. The existence of a sizeable negative causal effect of lecture style teaching would only be consistent with our results if teachers with favorable unobserved characteristics predominantly select themselves into an inferior teaching practice. Such a scenario, however, lacks any intuitional and theoretical support. We can thus exclude the possibility of a sizeable detrimental effect of lecture style teaching in math and science instruction on overall student achievement in US middle schools.

We believe that this result is relevant for the debate on optimizing the teaching process. Various dimensions of teaching practices have been shown to matter for student achievement. Moreover, the low-cost implementation of changes in the teaching process compared to other policy measures designed to foster student learning makes improvements in the teaching process particularly appealing. There exists, however, little consensus about what measures could improve the teaching process. Reducing the amount of traditional instruction based on lecture style teaching is typically a key candidate. Lectures are potentially connected with many disadvantages and might therefore be an inferior teaching method. National standards (NCTM, 1991; National Research Council, 1996) also advocate engaging students more in hands-on learning activities and group work but traditional lecture and textbook methodologies remain dominant in science and mathematics instruction in US middle schools. This raises the concern that the high share of total teaching time devoted to traditional lecture presentations has a detrimental effect on overall student learning in US middle schools. Our results, however, show that there exists no empirical support for this concern. Moreover, while reducing traditional lecture style teaching might have beneficial effects on non-cognitive outcomes or cognitive outcomes not measured by TIMSS test scores, our findings imply that policies designed to reduce the amount of traditional lecture style teaching in US middle schools contain little potential for raising overall achievement levels in math and science.

References

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

Altonji, J., Elder, T., and Taber, C. (2005). Selection on observed and unobserved variables: assessing the effectiveness of catholic schools. *Journal of Political Economy*, 113(1):151–184.

Aslam, M. and Kingdon, G. (2007). What can Teachers do to Raise Pupil Achievement? The centre for the study of african economies working paper series.

Banerjee, A. V., Cole, S., Duflo, E., and Linden, L. (2007). Remedying education: Evidence from two randomized experiments in india. *The Quarterly Journal of Economics*, 122(3):1235–1264.

Barrow, L., Markman, L., and Rouse, C. E. (2009). Technology's edge: The educational benefits of computer-aided instruction. *American Economic Journal: Economic Policy*, 1(1):52–74.

Bonesrønning, H. (2004). Do the teachers' grading practices affect student achievement? *Education Economics*, 12(2):151–167.

Borman, G. D. and Kimball, S. M. (2005). Teacher quality and educational quality: Do teachers with higher standards-based evaluation ratings close student achievement gaps? *The Elementary School Journal*, 106(1):3–20.

Brewer, D. J. and Goldhaber, D. (1997). Why don't schools and teachers seem to matter? *The Journal of Human Resources*, 32(3):505–523.

Cohen, D. K. and Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in california. *Teachers College Record*, 102(2):294–343.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, 95(2):158–165.

Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *The Journal of Human Resources*, 42(3):528–554.

Gallagher, H. A. (2004). Vaughn elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education*, 74(4):79–107.

Goe, L. (2007). The link between teacher quality and student outcomes: A research synthesis. Report, National Comprehensive Center for Teacher Quality, Washington DC.

Gordon, R., Kane, T. J., and Staiger, D. O. (2006). The hamilton project: Identifying effective teachers using performance on the job. Report Washington, DC, The Brookings Institution.

Heneman, H. G., Milanowski, A., Kimball, S. M., and Odden, A. (2006). Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay. CPRE Policy Brief RB-45, Consortium for Policy Research in Education, Philadelphia.

Holtzapple, E. (2003). Criterion-related validity evidence for standard-based teacher evaluation systems. *Journal of Personnel Evaluation in Education*, 17(3):207–219.

Jacob, B. A. and Lefgren, L. (2008). Can principals identify effective teachers? evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26:101–136.

Kane, T. J., Rockoff, J. E., and Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? evidence from new york city. *Economics of Education Review*, 27(6):615–631.

Kannapel, P. J., Clements, S. K., Taylor, D., and Hibpshman, T. (2005). Inside the black box of high-performing high-poverty schools. Report, Prichard Committee for Academic Excellence.

Kimball, S. M., White, B., Milanowski, A. T., and Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in washoe county. *Peabody Journal of Education*, 79(4):54–78.

Machin, S. and McNally, S. (2008). The literacy hour. *Journal of Public Economics*, 92(5-6):1441–1462.

Martin, M. E. (2005). *TIMSS 2003 User Guide for the International Data Base*. International Association for the Evaluation of Educational Achievement (IEA), Boston.

Matsumura, L. C., Garnier, H. E., Pascal, J., and Valdes, R. (2002). Measuring instructional quality in accountability systems: Classroom assignments and student achievement. *Educational Assessment*, 8(3):207–229.

Matsumura, L. C., Slater, S. C., Junker, B., Peterson, M., Boston, M., Steele, M., and Resnick, L. (2006). Measuring reading comprehension and mathematics instruction in urban middle schools: A pilot study of the instructional quality assessment. CSE Technical Report 681, National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles.

McGaffrey, D. F., Hamilton, L. S., Stecher, B. M., Klein, S. P., Bugliari, D., and Robyn, A. (2001). Interactions among instructional practices, curriculum and student achievement: The case of standards-based high school mathematics. *Journal for Research in Mathematics Education*, 35(5):493–517.

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from cincinnati. *Peabody Journal of Education*, 79(4):33–53.

National Research Council (1996). National science education standards. Report, Washington, DC.

NCTM (1991). Professional standards for teaching mathematics. Report, National Council of Teachers of Mathematics, Reston, VA.

Newmann, F. M., Bryk, A. S., and Nagaoka, J. (2001). Authentic intellectual work and standardized tests: Conflict or coexistence? Report, Consortium on Chicago School Research, Chicago.

Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2):417–458.

Rouse, C. E. and Krueger, A. B. (2004). Putting computerized instruction to the test: a randomized evaluation of a "scientifically based" reading program. *Economics of Education Review*, 23(4):323–338.

Schacter, J. and Thum, Y. M. (2004). Paying for high- and low-quality teaching. *Economics of Education Review*, 23(4):411–430.

Seidel, T. and Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4):454.

Smith, J. B., Lee, V. E., and Newmann, F. M. (2001). Instruction and achievement in chicago elementary schools. Report, Consortium on Chicago School Research, Chicago.

Weiss, I. R. (1997). The status of science and mathematics teaching in the united states: Comparing teacher views and classroom practice to national standards. NISE Brief 3, National Institute for Science Education, Madison: University of WisconsinMadison.

Wenglinsky, H. (2000). How teaching matters: Bringing the classroom back into discussion of teacher quality. Report, Policy Information Center, Princeton.

Wenglinsky, H. (2002). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives*, 10(12).

Appendix

Selection on unobservables following Altonji et al. (2005)

Formally, in our application the assumption that selection on unobservables occurs to the same degree as selection on observables as imposed by Altonji, Elder and Taber (2005) can be stated as:

$$\begin{aligned} Proj(\Delta Lecture|\Delta T\beta_3, \eta) &= \phi_0 + \phi_{\Delta T'\beta_3}\Delta T'\beta_3 + \phi_\eta \eta && (8) \\ \text{with} \qquad \phi_{\Delta T'\beta_3} &= \phi_\eta && (9) \end{aligned}$$

Where $\Delta Lecture$ captures the between subject differences in the percent of effective teaching time devoted to lecture style presentation and β_4 indicates its coefficient while T includes all other k control variables (Teacher characteristics and effective teaching time as well as class characteristics) and β_3 is a $kx1$ vector of coefficients. When $\Delta T'\beta_3$ is orthogonal to η assumption (9) is equal to

$$\frac{Cov(\Delta T'\beta_3, \Delta Lecture)}{Var(\Delta T'\beta_3)} = \frac{Cov(\eta, \Delta Lecture)}{Var(\eta)} \tag{10}$$

We proceed now to answer the question how large selection on unobservables relative to selection on observabels would have to be in order to explain the entire estimate of β_4 under the assumption that the true β_4 is 0. Following Altonji et al. (2005) we regress $\Delta Lecture$ on ΔT, to get

$$\Delta Lecture = \Delta T'\delta + \Delta\widetilde{Lecture}.$$

Plugging this into our equation (5) yields:

$$\Delta Y \quad = c + \Delta T'(\beta_3 + \delta * \beta 4) + \Delta\widetilde{Lecture}'\beta_4 + \eta \tag{11}$$

As $\Delta\widetilde{Lecture}$ is by construction orthogonal to ΔT the probability limit of $\widehat{\beta_4}$ can be written as

$$plim \ \widehat{\beta_4} = \beta_4 + \frac{Cov(\Delta\widetilde{Lecture}, \eta)}{Var(\Delta\widetilde{Lecture})}$$

where

$$\frac{Cov(\Delta\widetilde{Lecture}, \eta)}{Var(\Delta\widetilde{Lecture})} = \frac{Cov(\Delta Lecture, \eta)}{Var(\Delta\widetilde{Lecture})}$$

as ΔT is orthogonal to η.

To estimate the numerator of the bias we can use equality (10):

$$\frac{Cov(\Delta T'\beta_3, \Delta Lecture)}{Var(\Delta T'\beta_3)} * Var(\eta).$$

For this however, we need a consistent estimate of β_3 which we obtain by estimating equation (11) under the assumption that $\beta_4 = 0$.

Table 1: Descriptive Statistics- Teacher variables

	Math		Science		
	303 teachers		355 teachers		
Variable	Mean	SD	Mean	SD	Difference
Teaching practices					
Lecture style teaching (share of effective)	0.32	0.187	0.374	0.202	-0.054***
Effective teaching (share of overall)	0.572	0.119	0.554	0.161	0.018
Teacher variables					
Teacher is female	0.649	0.473	0.540	0.496	0.109**
Full teaching certificate	0.970	0.163	0.957	0.188	0.013
Major in math	0.473	0.492	0.099	0.294	0.374***
Major in science	0.146	0.348	0.584	0.486	-0.438***
Major in education	0.598	0.483	0.456	0.491	0.142***
Teacher younger than 30	0.119	0.322	0.143	0.349	-0.024
Teacher aged 40-49	0.293	0.452	0.335	0.469	-0.042
Teacher at least 50	0.315	0.462	0.299	0.455	0.017
Teaching experience $<$ 1 year	0.043	0.201	0.042	0.199	0.001
Teaching experience 1-5 years	0.178	0.370	0.224	0.404	-0.046
Teacher training 0 years	0.102	0.301	0.154	0.359	-0.051**
Teacher training 1 year	0.578	0.491	0.523	0.497	0.055
Teacher training 2 years	0.209	0.404	0.193	0.392	0.016
Teacher training 3 years	0.039	0.192	0.048	0.213	-0.009
Teacher training 4 years	0.056	0.228	0.035	0.184	0.021
Teacher training 5 years	0.008	0.090	0.039	0.193	-0.031**
Motivation					
Pedagogy classes in last 2 years	0.748	0.431	0.648	0.472	0.100***
Subject content classes in last 2 years	0.840	0.364	0.827	0.374	0.014
Subject curriculum classes in last 2 years	0.830	0.372	0.853	0.349	-0.023
Subject related IT classes in last 2 years	0.729	0.441	0.803	0.393	-0.074**
Subject assessment classes in last 2 years	0.756	0.426	0.649	0.471	0.107***
Classes on improving student's skills last 2 years	0.759	0.424	0.766	0.418	-0.007
Working hours scheduled per week	21.119	8.276	20.159	7.291	0.960
Weekly hours spent on lesson planning	3.704	2.708	4.680	3.276	-0.976***
Weekly hours spent on grading	5.252	3.930	6.083	4.407	-0.830**
Teaching requirements					
Requirement probationary period	0.502	0.493	0.496	0.479	0.007
Requirement licensing exam	0.526	0.493	0.558	0.479	-0.032
Requirement finished Isced5a	0.891	0.307	0.824	0.371	0.067**
Requirement minimum education classes	0.833	0.368	0.777	0.399	0.056
Requirement minimum subject specific classes	0.799	0.395	0.744	0.420	0.056

Note: Probability weights and within school correlation are taken into account when estimating means and standard deviations. Teacher variables are weighted by the number of students taught by each teacher.

Table 2: Estimation Results OLS

	Math1	Math2	Science1	Science2	Pooled1	Pooled2
Lecture teaching	.488**	.429	.193	.651**	.361***	.291***
	(.19)	(.33)	(.13)	(.28)	(.11)	(.09)
Effective teaching time	.220	.781	−.0387	.465	.0101	.318**
	(.29)	(.49)	(.16)	(.28)	(.14)	(.13)
Female teacher	−.148**	−.175	−.0751	−.0131	−.0942**	−.0196
	(.06)	(.11)	(.06)	(.09)	(.05)	(.04)
Teacher younger than 30	.0986	.00829	.100	.245*	.102	.0551
	(.12)	(.23)	(.09)	(.13)	(.08)	(.08)
Teacher's age 40-49	.0656	.107	.0360	.124	.0415	.00161
	(.08)	(.18)	(.07)	(.10)	(.05)	(.05)
Teacher at least 50	.0667	.0943	.0680	−.00910	.0352	.0160
	(.08)	(.17)	(.07)	(.12)	(.06)	(.06)
Teaching experience	−.452***	−.734***	−.0919	.00678	−.235**	−.160*
<1 years	(.15)	(.25)	(.16)	(.17)	(.10)	(.09)
Teaching experience	−.155	−.204	−.0210	−.0253	−.104	−.0774
$1-5$ years	(.10)	(.19)	(.07)	(.10)	(.07)	(.07)
Teaching certificate	−.411**	−1.521***	.00225	−.305	−.0840	−.134
	(.16)	(.26)	(.17)	(.38)	(.13)	(.16)
Constant	1.524***	2.492***	.281	1.957***	.761	.0403
	(.53)	(.87)	(.58)	(.67)	(.48)	(.41)
Student Background	Yes	Yes	Yes	Yes	Yes	Yes
School variables	Yes	No	Yes	No	Yes	No
School fixed effects	No	Yes	No	Yes	No	Yes
Teacher variables	Yes	Yes	Yes	Yes	Yes	Yes
Class variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6310	6310	6310	6310	12620	12620
R^2	.333	.527	.339	.503	.316	.476

* p<0.10, ** p<0.05, *** p<0.01

Note: a) Bias estimated as in equation (6) using condition (7) b) Ratio of the coefficient of lecture style presentation and the bias. Clustered standard errors in parentheses. Dependent variable is the standardized student test score. Effective teaching time is measured as the share of overall time in class spent on either problem solving or giving lecture style presentation. Lecture style teaching is measured as the share of effective teaching time spent on giving lecture style presentations. Student background includes students' gender, age in years, a dummy if born in first 6 months of the year, number of books at home, English spoken at home, migration background, household size and parental education. School variables include dummies capturing different levels of parental involvement in school activities and dummies for community size. Not reported teacher variables are teacher's major in math, science and education and years of teacher training. Class variables are class size, spread of test scores and a tracking indicator. Imputation indicators are included in all regressions.

Table 3: Estimation Results First Difference

	1	2	3	4	5	6
Lecture style teaching	.112*	.112*	.125**	.106**	.0997**	.0977*
	(.06)	(.06)	(.05)	(.05)	(.05)	(.05)
Effective teaching time		.00952	–.0258	–.0355	–.0417	–.0388
		(.07)	(.06)	(.06)	(.06)	(.06)
Teacher female			–.00766	–.0252	–.0336	–.0341
			(.02)	(.02)	(.02)	(.02)
Teaching certificate			.00503	.0335	.0554	.0504
			(.06)	(.11)	(.12)	(.13)
Teacher younger than 30			–.0266	–.0609*	–.0507	–.0430
			(.04)	(.04)	(.04)	(.04)
Teacher's age 40-49			.0143	.0431	.0509*	.0415
			(.03)	(.03)	(.03)	(.03)
Teacher at least 50			.0212	.0431	.0405	.0352
			(.03)	(.03)	(.03)	(.03)
Teaching experience			–.00595	–.0134	–.0138	–.0290
<1 years			(.04)	(.05)	(.05)	(.05)
Teaching experience			.0116	.0278	.0257	.0140
$1-5$ years			(.03)	(.03)	(.03)	(.03)
Constant	–.0103	–.0105	.0188	.00341	–.000330	–.00976
	(.01)	(.01)	(.03)	(.03)	(.03)	(.03)
Class var	No	No	Yes	Yes	Yes	Yes
Teacher var	No	No	Yes	Yes	Yes	Yes
Limit to teach	No	No	No	Yes	Yes	Yes
Motivation	No	No	No	No	Yes	Yes
Teaching requirements	No	No	No	No	No	Yes
Observations	6310	6310	6310	6310	6310	6310
R^2	.002	.002	.012	.031	.034	.037
Bias^a)			.958	2.041	2.253	2.399
Ratio^b)			.131	.052	.044	.041

* p<0.10, ** p<0.05, *** p<0.01

Note: Clustered standard errors in parentheses. Dependent variable is the within student difference of standardized math and science test scores. All teacher variables are included as within student between subject differences. Effective teaching time is the share of time in class spent on problem solving and giving lecture style presentation. Lecture style presentation is the share of effective teaching time spent on giving lecture style presentation. Variables included in Motivation, Teacher variables and Add teacher variables are shown in table 1. Variables included in Class variables and teaching limits are shown in table A-1. Imputation indicators included in all but the first two columns.

Table 4: Estimation Results: Robustness Checks

Other Definitions				
	Def 1	Def 2	Def 3	Def 4
Lecture style teaching	.108*	.129**	.104**	.119
	(.06)	(.06)	(.05)	(.07)
Observations	6310	6310	6310	6310
R^2	.037	.038	.038	.037

Subsamples				
	Same Peers	Diff Peers	No Track	Track
Lecture style teaching	.166	.0806	.0576	.139*
	(.13)	(.06)	(.06)	(.08)
Observations	2205	4105	3529	2292
R^2	.103	.049	.073	.085

Heterogenous effects		
	Diff	Background
Lecture style teaching	.133*	.162**
(math)	(.07)	(.07)
Lecture style teaching	–.0113	–.0560
(science)	(.07)	(.07)
Observations	6310	6310
R^2	.067	.088

* p<0.10, ** p<0.05, *** p<0.01

Note: Clustered standard errors in parentheses. Dependent variables in all panels and columns are the within student between subject differences in standardized test scores. All teacher variables, class variables, motivation and teaching limits are included as controls. Upper panel: In Def 1 time spent re-teaching and clarifying content/procedures is included in effective teaching and in lecture style teaching. In Def 2 effective teaching time includes taking tests or quizzes in addition to giving lecture style presentation and problem solving. Lecture style teaching in column 2 is defined in relation to the latter three. In Def 3 effective teaching is further divided into different categories. Def 4 takes time spent on giving lecture style presenation in relation to all other time-use categories. Middle panel: Separate estimation for different sub-samples: Column 1 only students with same classmates in both subjects, column 2 students with different classmates. Column 3 students who are tracked according to ability in either both or none of the two subjects, column 4 students who are tracked in at least one of the two subjects. Lower panel: Column 1 and 2 allow different coefficients in the two subjects, column 2 includes student background as additional controls. Imputation indicators are included in all estimations.

Table A-1: Descriptive Statistics- Class Characteristics

	Math		Science		
	359 classes		734 classes		
Variable	Mean	SD	Mean	SD	Difference
Class variables					
Class size	23.458	6.543	24.571	7.514 1	-1.113**
Student's tracked according to ability	0.550	0.480	0.171	0.363	0.379***
Spread of test scores	0.614	0.124	0.678	0.187	-0.064***
Teaching limits (reference not at all/not applicable)					
Differing academic ability - a little	0.339	0.469	0.340	0.473	-0.002
Differing academic ability - some	0.330	0.466	0.321	0.466	0.008
Differing academic ability - a lot	0.204	0.399	0.174	0.378	0.031
Wide range of backgrounds - a little	0.308	0.456	0.277	0.447	0.031
Wide range of backgrounds - some	0.205	0.399	0.238	0.425	-0.032
Wide range of backgrounds - a lot	0.060	0.234	0.0778	0.267	-0.018
Special need students - a little	0.309	0.457	0.328	0.469	-0.019
Special need students - some	0.147	0.350	0.184	0.387	-0.037
Special need students - a lot	0.0642	0.243	0.081	0.272	-0.016
Uninterested students - a little	0.376	0.480	0.357	0.477	0.019
Uninterested students - some	0.276	0.443	0.298	0.455	-0.022
Uninterested students - a lot	0.172	0.374	0.161	0.365	0.011
Low morale students - a little	0.420	0.489	0.324	0.466	0.095**
Low morale students - some	0.199	0.395	0.262	0.438	-0.063
Low morale students - a lot	0.094	0.290	0.082	0.273	0.013
Disruptive students - a little	0.440	0.492	0.393	0.487	0.047
Disruptive students - some	0.228	0.416	0.291	0.453	-0.062
Disruptive students - a lot	0.109	0.309	0.142	0.349	-0.033
Shortage computer hardware - a little	0.140	0.343	0.236	0.423	-0.097***
Shortage computer hardware - some	0.197	0.396	0.207	0.404	-0.011
Shortage computer hardware - a lot	0.110	0.310	0.189	0.390	-0.079***
Shortage computer software - a little	0.168	0.371	0.294	0.454	-0.125***
Shortage computer software - some	0.146	0.350	0.198	0.397	-0.052
Shortage computer software - a lot	0.145	0.349	0.174	0.378	-0.029
Shortage support pc use - a little	0.181	0.380	0.217	0.411	-0.036
Shortage support pc use - some	0.148	0.351	0.185	0.387	-0.037
Shortage support pc use - a lot	0.089	0.282	0.137	0.343	-0.048*
Shortage of textbooks - a little	0.055	0.225	0.088	0.283	-0.034
Shortage of textbooks - some	0.045	0.205	0.044	0.205	0.001
Shortage of textbooks - a lot	0.011	0.103	0.083	0.275	-0.072***

Table A-2: Descriptive Statistics- Class Characteristics (cont.)

	Math		**Science**		
	359 classes		734 classes		
Variable	Mean	SD	Mean	SD	Difference
Shortage instructional equipment - a little	0.180	0.380	0.314	0.463	-0.134***
Shortage instructional equipment - a some	0.123	0.326	0.193	0.394	-0.070**
Shortage instructional equipment - a lot	0.038	0.190	0.141	0.347	-0.103 ***
Shortage demonstrative equipment - a little	0.253	0.431	0.318	0.465	-0.065
Shortage demonstrative equipment - some	0.117	0.318	0.196	0.396	-0.080**
Shortage demonstrative equipment - a lot	0.044	0.203	0.189	0.391	-0.146***
Inadequate physical facilities - a little	0.148	0.352	0.219	0.413	-0.071*
Inadequate physical facilities - some	0.051	0.219	0.158	0.364	-0.107***
Inadequate physical facilities - a lot	0.030	0.169	0.131	0.337	-0.101***
High student teacher ratio - a little	0.230	0.417	0.292	0.454	-0.062
High student teacher ratio - some	0.132	0.335	0.204	0.402	-0.071**
High student teacher ratio - a lot	0.091	0.285	0.129	0.334	-0.038

Note: Probability weights and within school correlation are taken into account when estimating means and standard deviations. Class variables are weighted by the number of students in each class.

CESifo Working Paper Series

for full list see www.cesifo-group.org/wp
(address: Poschingerstr. 5, 81679 Munich, Germany, office@cesifo.de)

2573 Jo Thori Lind, Karl Ove Moene and Fredrik Willumsen, Opium for the Masses? Conflict-Induced Narcotics Production in Afghanistan, March 2009

2574 Silvia Marchesi, Laura Sabani and Axel Dreher, Agency and Communication in IMF Conditional Lending: Theory and Empirical Evidence, March 2009

2575 Carlo Altavilla and Matteo Ciccarelli, The Effects of Monetary Policy on Unemployment Dynamics under Model Uncertainty - Evidence from the US and the Euro Area, March 2009

2576 Falko Fecht, Kjell G. Nyborg and Jörg Rocholl, The Price of Liquidity: Bank Characteristics and Market Conditions, March 2009

2577 Giorgio Bellettini and Filippo Taddei, Real Estate Prices and the Importance of Bequest Taxation, March 2009

2578 Annette Bergemann and Regina T. Riphahn, Female Labor Supply and Parental Leave Benefits – The Causal Effect of Paying Higher Transfers for a Shorter Period of Time, March 2009

2579 Thomas Eichner and Rüdiger Pethig, EU-Type Carbon Emissions Trade and the Distributional Impact of Overlapping Emissions Taxes, March 2009

2580 Antonios Antypas, Guglielmo Maria Caporale, Nikolaos Kourogenis and Nikitas Pittis, Selectivity, Market Timing and the Morningstar Star-Rating System, March 2009

2581 António Afonso and Christophe Rault, Bootstrap Panel Granger-Causality between Government Budget and External Deficits for the EU, March 2009

2582 Bernd Süssmuth, Malte Heyne and Wolfgang Maennig, Induced Civic Pride and Integration, March 2009

2583 Martin Peitz and Markus Reisinger, Indirect Taxation in Vertical Oligopoly, March 2009

2584 Petra M. Geraats, Trends in Monetary Policy Transparency, March 2009

2585 Johannes Abeler, Armin Falk, Lorenz Götte and David Huffman, Reference Points and Effort Provision, March 2009

2586 Wolfram F. Richter, Taxing Education in Ramsey's Tradition, March 2009

2587 Yin-Wong Cheung, Menzie D. Chinn and Eiji Fujii, China's Current Account and Exchange Rate, March 2009

2588 Alexander Haupt and Silke Uebelmesser, Voting on Labour-Market Integration and Education Policy when Citizens Differ in Mobility and Ability, March 2009

2589 Hans Jarle Kind, Marko Koethenbuerger and Guttorm Schjelderup, Should Utility-Reducing Media Advertising be Taxed?, March 2009

2590 Alessandro Cigno, How to Avoid a Pension Crisis: A Question of Intelligent System Design, March 2009

2591 Helmut Lütkepohl and Fang Xu, The Role of the Log Transformation in Forecasting Economic Variables, March 2009

2592 Rainald Borck, Hyun-Ju Koh and Michael Pflüger, Inefficient Lock-in and Subsidy Competition, March 2009

2593 Paolo M. Panteghini, On the Equivalence between Labor and Consumption Taxation, March 2009

2594 Bruno S. Frey, Economists in the PITS?, March 2009

2595 Natalie Chen and Dennis Novy, International Trade Integration: A Disaggregated Approach, March 2009

2596 Frédérique Bec and Christian Gollier, Term Structure and Cyclicity of Value-at-Risk: Consequences for the Solvency Capital Requirement, March 2009

2597 Carsten Eckel, International Trade and Retailing, March 2009

2598 Gianni De Nicolò and Iryna Ivaschenko, Global Liquidity, Risk Premiums and Growth Opportunities, March 2009

2599 Jay Pil Choi and Heiko Gerlach, International Antitrust Enforcement and Multi-Market Contact, March 2009

2600 Massimo Bordignon and Guido Tabellini, Moderating Political Extremism: Single Round vs Runoff Elections under Plurality Rule, April 2009

2601 Ana B. Ania and Andreas Wagener, The Open Method of Coordination (OMC) as an Evolutionary Learning Process, April 2009

2602 Simon Gächter, Daniele Nosenzo, Elke Renner and Martin Sefton, Sequential versus Simultaneous Contributions to Public Goods: Experimental Evidence, April 2009

2603 Philippe Jehiel and Andrew Lilico, Smoking Today and Stopping Tomorrow: A Limited Foresight Perspective, April 2009

2604 Andreas Knabe, Steffen Rätzel, Ronnie Schöb and Joachim Weimann, Dissatisfied with Life, but Having a Good Day: Time-Use and Well-Being of the Unemployed, April 2009

2605 David Bartolini and Raffaella Santolini, Fiscal Rules and the Opportunistic Behaviour of the Incumbent Politician: Evidence from Italian Municipalities, April 2009

2606 Erkki Koskela and Jan König, Can Profit Sharing Lower Flexible Outsourcing? A Note, April 2009

2607 Michel Beine, Frédéric Docquier and Çağlar Özden, Diasporas, April 2009

2608 Gerd Ronning and Hans Schneeweiss, Panel Regression with Random Noise, April 2009

2609 Adam S. Booij, Bernard M.S. van Praag and Gijs van de Kuilen, A Parametric Analysis of Prospect Theory's Functionals for the General Population, April 2009

2610 Jeffrey R. Brown, Julia Lynn Coronado and Don Fullerton, Is Social Security Part of the Social Safety Net?, April 2009

2611 Ali Bayar and Bram Smeets, Economic, Political and Institutional Determinants of Budget Deficits in the European Union, April 2009

2612 Balázs Égert, The Impact of Monetary and Commodity Fundamentals, Macro News and Central Bank Communication on the Exchange Rate: Evidence from South Africa, April 2009

2613 Michael Melvin, Christian Saborowski, Michael Sager and Mark P. Taylor, Bank of England Interest Rate Announcements and the Foreign Exchange Market, April 2009

2614 Marie-Louise Leroux, Pierre Pestieau and Gregory Ponthiere, Should we Subsidize Longevity?, April 2009

2615 Ronald MacDonald, Lukas Menkhoff and Rafael R. Rebitzky, Exchange Rate Forecasters' Performance: Evidence of Skill?, April 2009

2616 Frederick van der Ploeg and Steven Poelhekke, The Volatility Curse: Revisiting the Paradox of Plenty, April 2009

2617 Axel Dreher, Peter Nunnenkamp, Hannes Öhler and Johannes Weisser, Acting Autonomously or Mimicking the State and Peers? A Panel Tobit Analysis of Financial Dependence and Aid Allocation by Swiss NGOs, April 2009

2618 Guglielmo Maria Caporale, Roman Matousek and Chris Stewart, Rating Assignments: Lessons from International Banks, April 2009

2619 Paul Belleflamme and Martin Peitz, Asymmetric Information and Overinvestment in Quality, April 2009

2620 Thomas Dohmen, Armin Falk, David Huffman and Uwe Sunde, Are Risk Aversion and Impatience Related to Cognitive Ability?, April 2009

2621 Yin-Wong Cheung and Xingwang Qian, The Empirics of China's Outward Direct Investment, April 2009

2622 Frédérique Bec and Christian Gollier, Assets Returns Volatility and Investment Horizon: The French Case, April 2009

2623 Ronnie Schöb and Marcel Thum, Asymmetric Information Renders Minimum Wages Less Harmful, April 2009

2624 Martin Ruf and Alfons J. Weichenrieder, The Taxation of Passive Foreign Investment – Lessons from German Experience, April 2009

2625 Yao Li, Borders and Distance in Knowledge Spillovers: Dying over Time or Dying with Age? – Evidence from Patent Citations, April 2009

2626 Jim Malley and Ulrich Woitek, Technology Shocks and Aggregate Fluctuations in an Estimated Hybrid RBC Model, April 2009

2627 Jin Cao and Gerhard Illing, Endogenous Systemic Liquidity Risk, April 2009

2628 Thiess Buettner and Bjoern Kauder, Revenue Forecasting Practices: Differences across Countries and Consequences for Forecasting Performance, April 2009

2629 Håkan Selin, The Rise in Female Employment and the Role of Tax Incentives – An Empirical Analysis of the Swedish Individual Tax Reform of 1971, April 2009

2630 Nick Johnstone and Ivan Hascic, Environmental Policy Design and the Fragmentation of International Markets for Innovation, April 2009

2631 Spiros Bougheas, Richard Kneller and Raymond Riezman, Optimal Education Policies and Comparative Advantage, April 2009

2632 Jay Pil Choi and Heiko Gerlach, Multi-Market Collusion with Demand Linkages and Antitrust Enforcement, April 2009

2633 Thor O. Thoresen, Income Mobility of Owners of Small Businesses when Boundaries between Occupations are Vague, April 2009

2634 Guido Schwerdt and Amelie C. Wuppermann, Is Traditional Teaching really all that Bad? A Within-Student Between-Subject Approach, April 2009