TIER

How are Teachers Teaching?
A Nonparametric Approach

C. Van Klaveren and K. de Witte

TIER WORKING PAPER SERIES
TIER WP 10/23

Universiteit van Amsterdam   Maastricht University Leading in Learning!   rijksuniversiteit groningen

How are Teachers Teaching? A Nonparametric Approach

Kristof De Witte*

Chris Van Klaveren†

July 12, 2010

Abstract

This paper examines which configuration of teaching activities (expressed in, e.g., problem solving, homework, lecturing) maximizes student performance. To do so, it formulates a non-parametric efficiency model that is rooted in the Data Envelopment Analysis literature. In the model, we account for (1) self selection of students and teachers in better schools, and (2) complementary teaching activities. The analysis distinguishes both individual teaching (i.e., a personal teaching style adapted to the individual needs of the student) and collective teaching (i.e., a similar style for all students in a class). Exploiting the data set, we compare the actual teaching style as revealed by the teacher in the data to the model estimations. As such, we analyse which students in the class the teacher is targeting with his/her teaching style. The main results show that high test scores are associated with teaching styles that emphasise problem solving and homework. In addition, teachers seem to adapt their optimal teaching style on the 70 percent least performing students.

JEL Codes: C14, C61, C23, I21

Keywords: Data Envelopment Analysis; Teacher Quality; Student Performance; Nonparametric estimation; Revealed teaching style

1 Introduction

For outsiders, it is often unclear what makes succesful teachers (in terms of higher educational attainments of their students) and less succesful teachers. The economic literature on teacher quality

*The corresponding author is affiliated with Maastricht University, Top Institute for Evidence Based Education Research (TIER), Faculty of Humanities and Science, P.O. Box 616, 6200 MD Maastricht, The Netherlands. The author is also affiliated with the University of Leuven, Naamsestraat 69, 3000 Leuven, Belgium. Email: kristof.dewitte@econ.kuleuven.be

†The author is affiliated with Maastricht University, Top Institute for Evidence Based Education Research (TIER), Faculty of Humanities and Science, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Email: cp.vanklaveren@maastrichtuniversity.nl.

and student performance roughly distinguishes between two perspectives. The first perspective starts with the recognition that better teachers produce better students. The focal point in this literature is pinpointing the teacher characteristics that improve student performance significantly (see, among others, Hanushek, Kain and Rivkin, 2005; Hanushek and Rivkin, 2006; Clotfelter, Ladd and Vigdor, 2006; Clotfelter, Ladd and Vigdor, 2007; Aaronson, Barrow and Sander, 2007). The second perspective acknowledges that teacher practices matter for student performance. This literature focuses less on the characteristics that teachers possess, but the more on the activities that teachers undertake to improve student performance (see, among others, Brewer and Goldhaber, 1997; Smith, Lee and Newmann, 2001; Aslam and Kingdon, 2007; Machin and McNally, 2008; Barrow, Markman and Rouse, 2009; Schwerdt and Wuppermann, 2009; Van Klaveren, 2009).

Both strands generally use parametric models to examine the effect of teacher inputs (either teacher characteristics or practices) on student performance. In these parametric models, the researcher formulates *a priori* a parametric education production function, brings this reduced form model to the data and estimates the model parameters. The parameters associated with teacher inputs are then interpreted as the influence of teacher inputs on student performance, under the assumption that the model is correctly specified. The parametric results thus depend crucially on the *a priori* specified functional form of the educational production function, while the researcher usually does not have any *a priori* information on the underlying production technology (see Yatchew (1998) in general, and Rothstein (2008) for educational settings). As a consequence the model is often wrongly specified, which results in biased estimation results (Hjalmarsson et al., 1996).

Contributions

This study contributes to the literature in three different dimensions. First of all, it examines how teaching time relates to student performance using a fully nonparametric efficiency model. By nonparametrically estimating the relationship between teaching practices, teacher characteristics and educational attainments, we recognize the critique of Rothstein (2008), and, more generally, the inherent deficiencies of a parametric analysis. The nonparametric efficiency model is rooted in the Data Envelopment Analysis (DEA; Charnes et al., 1978) literature. DEA is a linear programming technique that allows us to estimate the most efficient set of inputs (in casu, teaching activities) to obtain the outputs (i.e., test scores). From a methodological point of view, the suggested nonparametric framework is attractive as, by definition, it does not require an *a priori* specified educational production function and, therefore, the model is less prone to misspecification bias.

Secondly, we adapt the nonparametric model to account for selection bias. The literature argues that students with more favorable background characteristics tend to self-select in better schools (see, e.g., Hanushek, Kain and Rivkin, 2005, Clotfelter, Ladd and Vigdor, 2006 and Clotfelter, Ladd and Vigdor, 2007). Therefore we apply a conditional nonparametric efficiency model (Daraio and Simar, 2005, 2007) to control for characteristics that determine the selection of high ability students to better schools, such as ethnic background and the education level of the parents. Furthermore, we account for self-selection of presumably better teachers to better schools and control for the teachers' experience and education level. We emphasize that in a DEA setting controlling for covariates means that we simulate a control group for each student in the sample who is evaluated on the basis of these covariates. In essence, the non-parametric Data Envelopment Analysis is combined with a kernel matching approach, because for each student we simulate an appropriate control group to determine the optimal configuration of teaching activities.

Thirdly, and from an empirical point of view, the attractiveness of the nonparametric DEA model arises from its endogenous weight specification for each of the teaching activities (i.e., the inputs in the model). As such, and in contrast to previous literature (e.g., Rouce and Krueger, 2004; Banerjee et al. 2007; Barrow et al., 2009; Schwerdt and Wuppermann, 2009; Van Klaveren, 2009), we allow different teaching activities to be complements rather than substitutes. Complementing lecturing activities correspond better to the real life activities of the teachers. For example, if a teacher lets students solve problems on their own, it may only be effective if the teacher gives individual guidance during the process.

The model

To set the scene, we briefly present the idea of the model. We estimate the optimal teaching style for each student (and later for each teacher) using the conditional DEA model. To do so, we include the complementary teaching activities (e.g., time allocated to homework, lecturing in front of the class, problems with guidance, problems without guidance, or reteaching) as inputs in the DEA model. The DEA model assigns endogenous weights to each of the teaching activities, such that the total sum of activities (1) equals the total teaching time and (2) maximizes the educational attainments of the students.

The model consists of three independent steps. In the first step, the proposed method estimates a 'first best optimum' where teachers adopt the optimal teaching style for each student *individually*. The first best optimum relates to the idea that lecturing style teaching is old fashioned and that a more personal teaching style allows teachers to better adjust their teaching method to the needs of the individual student, which on its turn increases student performance [see Schwerdt and Wuppermann (2009) and Van Klaveren (2009)]. In efficiency terms, however, a more personal approach is time intensive and eliminates the complementary effects of lecturing style teaching.

The second step corresponds to a 'second best optimum' where teachers are assumed to adopt an optimal teaching style for the average student in the class. This is later relaxed to a large part of the students in the class, such that we assume that teachers target their optimal teaching style on a percentage of the least advantageous student in the class. The second best optimum is more in line with reality, since teachers are not able to give each student a private lecture, but at the same time they do observe the abilities of their students and can use this information to choose an optimal teaching style. This study does not favor a 'first best' or 'second best' teaching style, it merely aims at comparing the teaching styles in both optima. As a promising side effect of the second best optimum, it enables us to detect the students that qualify for additional resources for student counseling. In particular, we can detect students who have difficulties in taking the lecturing style which is optimal for other students in the class.

The third step exploits the actual weights that teachers give to the different teaching activities we distinguish. By comparing these revealed weights (as indicated by the teachers and included in the data set) with the 'second best' model outcomes, we can discover on which students in the class the teacher is targeting his or her teaching style.

We illustrate the model by using the Trends in International Mathematics and Science Study (TIMSS) 2003 data. In particular we use data on math performance for Dutch students who are in their second year of secondary education. The data contain information on math performance as well as information on student, teacher, school and class characteristics. Furthermore there is detailed information on how teachers fill in the students' day with respect to their math lectures.

We proceed as follows. In Section 2, we discuss the literature on how teacher practices and teacher characteristics influence student performance and place this literature into one, comprehensive non-parametric model. In Section 3, we describe the probabilistic formulation of the production process and the conditional efficiency approach. Section 4 describes the data and discusses the empirical results. Finally, Section 5 concludes.

2 Combining teacher practices and teacher characteristics

The literature on how teacher inputs affect student performance can, roughly, be divided in two streams. A first stream considers the impact of teachers practices, while a second stream considers the influence of teacher characteristics. The paper at hand attempts to include both streams in one model. In this section, we explore the key points of both streams and indicate how they fit into our non-parametric model.

First consider the studies that focus on the impact of teacher practices and in particular on how student performance is influenced by the teachers allocation of time to different teaching activities (e.g., Aslam and Kingdon, 2007; Schwerdt and Wuppermann, 2009; Van Klaveren, 2009). The studies of Schwerdt and Wuppermann (2009) and Van Klaveren (2009) are inspired by the idea that lecturing style teaching would be old fashioned. A more personal teaching style would allow teachers to better adjust their teaching method to the needs of the individual student, which in turn increases student performance. Schwerdt and Wuppermann (2009) examine whether there is empirical support for this view for the U.S., but do not find that a more personal teaching style can be associated with significant higher student achievement. In fact, their findings show that traditional lecturing style teaching is associated with somewhat higher student achievement. Van Klaveren (2009) performs a comparable study for the Netherlands but he also finds no relation between student achievement and the time teachers spend on lecturing style teaching. Aslam and Kingdon (2007) focus on a broader spectrum of teacher activities and find that lesson planning, involving students by asking questions during class and quizzing them on past material, all substantially benefit pupil learning.

Many studies do not focus on the time that teachers spend on different teaching activities, but instead focus on teaching practices themselves (see, among others, Rouce and Krueger, 2004; Banerjee, Cole, Duflo and Linden, 2007; Barrow et al., 2009; Brewer and Goldhaber, 1997; Smith et al., 2001; Wenglinsky, 2002). These studies are informative with respect to this study, because some teaching practices can better characterize an optimal configuration of teaching activities than others. Brewer and Goldhaber (1997) find that instruction in small groups and emphasis on problem solving lead to lower student test scores. The first results show that it is important to control for class-size differences while the latter result shows that a configuration where a relatively large amount of the teaching time is spent on problem solving may have a negative impact on the performance of students. Smith et al. (2001) find that children on elementary schools benefit from more didactic or interactive teaching methods. If this result applies for children on secondary schools as well, i.e. the children in our sample, it shows that a distinction should be made between problem solving with guidance and without guidance. In the latter case the absence of student-teacher interaction can have an negative impact on student performance, which may be the result of Brewer and Goldhaber (1997). In the former case, problem solving in the presence of student-teacher interaction may have a positive impact on student performance. Wenglinsky (2002) examines the impact of different teaching practices on student test scores in math and science using parametric multi-level structural equation models. He finds that the use of hands-on learning activities, such as solving real world problems, working with objects, an emphasis on thinking skills and frequent testing of students are all positively related to student performance. Another study in that respect is Machin and McNally (2008) who analyze the effect of a literacy hour in English primary schools in the late 1990s. They find that the literacy hour significantly increased the reading skills for low ability students while high ability students were not affected. If we could generalize their result beyond the British school system, we would argue that students who are below the desired skill level benefit from extra teaching time.

Besides teaching practices and how teachers allocate time to different teaching activities, the effectiveness of a particular teaching style depends on the quality of the teacher, and, therefore, a second stream of literature focuses more on teacher characteristics. Most studies focus on how student performance is influenced by teacher characteristics that can directly be linked to the productivity of the teacher. The characteristics usually considered are years of teaching experience, education level, and whether the teacher has a teacher license. Intuitively, one would expect that these characteristics affect the performance of students, however, there is little empirical evidence that supports this intuition (Murnane and Phillips, 1981, Hanushek, 1997, Wayne and Youngs, 2003, Hanushek, Kain and Rivkin, 2005, Hanushek and Rivkin, 2006, Clotfelter, Ladd and Vigdor, 2006, Clotfelter, Ladd and Vigdor, 2007, Aaronson, Barrow and Sander, 2007).

The empirical results related to teacher productivity are often based on the so-called value added model. Value added models use panel data to relate the knowledge gain of students during a particular school year to various teacher characteristics, while controlling for school, student and class effects. Rothstein (2008) mentions that thanks to panel data one can better control for individual heterogeneity as compared to cross-sectional data. Nevertheless, he also shows that the parametric assumptions underlying value added models are often substantially incorrect. Due to a wrongly specified model, one obtains inconsistent and contradictory results with respect to the significance of teacher productivity characteristics. The only robust finding across the different studies performed and methodologies chosen, is that teaching experience influences student performance and that these experience effects are concentrated in the first few years of teaching. Specifically, teachers in their first, and to a somewhat lesser extent, their second year tend to perform significantly worse in the classroom (Hanushek and Rivkin, 2006).

In sum, the model, which is proposed in the next section, meets the critique of Yatchew (1998) and Rothstein (2008) in that it is fully non-parametric and, as such, avoids a specification bias. In addition, we bridge the two strands in the literature by linking teaching activities with teacher characteristics, and control in the analysis for student, parent, class and school characteristics.

3 A non-parametric model

3.1 Teaching in first versus second best

In every school, the school management decides to some extent discretionary upon the time that is allocated for each subject. Given this time constraint, we assume that teachers can freely allocate their time x to different teaching activities $(x_1, ..., x_p)$. As teachers are supposed to educate students their subject, we assume that teachers allocate the time such that it maximizes the test scores of the students. To a division of teaching time over different teaching activities we refer to as a teaching style. Given these benevolent and easy to defend assumptions, we develop a methodology that deduces the optimal teaching styles of teachers.

We distinguish two situations. In the first situation, each student is assumed to receive a tailor made teacher instruction and this scenario is labeled as 'first best optimum'. In this first best optimum teachers adopt an optimal teaching style on a *student*-level, given the time constraint they face and conditional on the characteristics of the teachers and students. In practice, however, teachers are not able to adopt an optimal teaching style at the student-level, since they are teaching a class of students. Therefore, the second situation assumes that teachers adopt an optimal teaching style at a class level, i.e teachers maximize the average attainments scores for the class, and to this situation we refer to as a 'second best optimum'. Later, we will relax the second best optimum and assume that teachers adopt a teaching style that is targeted to different student groups within the class (i.e., they target their optimal teaching style on a subgroup of students).

3.2 The data

The data arise from the Trends in International Mathematics and Science Study 2003 (TIMSS 2003). These data include information on student achievement in mathematics and science courses (such as biology, physics and chemistry) as well as information on students, teachers, schools and classes. The TIMSS data is a rich and unique database as teachers are extensively surveyed and teacher and student information can be linked at the student level. To limit the scope of the study, we use data for Dutch students who are in their second year of secondary school and focus on the math performance of these students. By focusing on one particular country, we remove cultural differences across countries.[1]

From the data set at hand, we observe pupil specific test scores on math (denote by y). In addition, we observe from teacher questionnaires (1) how many minutes teachers give math instruction to their class (denoted by t) and (2) how these minutes are divided over various teaching activities (x). The lecturing style of a teacher corresponds to the time allocated to the teaching activities (i.e., the vector of x).

We can distinguish 8 teaching activities ($x_1, ..., x_8$). In particular, the time allocated to (1) homework[2], (2) lecturing in front of the class, (3) problems with guidance, (4) problems without guidance, (5) reteaching, (6) tests and quizzes, (7) classroom management, and (8) other activities. Naturally, the time that teachers spend on these eight activities add up to the total teaching minutes ($\sum x = t$). The lecturing style of each teacher is now represented by how the teacher divides his or her time over these eight activities.

Although we observe the inputs (i.e., allocated teaching time) and the outputs (i.e., math test scores), we do not observe the functional form by which these inputs are transformed into the outputs. If we would assume some parametric functional relationship (e.g., translog or Cobb-Douglas) then all model predictions would depend crucially on this chosen functional form, and this functional form cannot be empirically validated (Yatchew, 1998 and Rothstein, 2008). Consequently, the obtained estimates may be biased due to a wrongly specified model. Therefore, a nonparametric procedure seems more appropriate as it avoids the specification bias (see before).

3.3 First best optimum

Consider the following setting. At the end of a math course, students are assessed and obtain test scores $y \in \Re^q$ on the subject. Each teacher has his or her individual lecturing style and divides the

[1] As a major disadvantage is that the data do not allow for initial test scores such that information on added value is hidden.

[2] With respect to the homework category, we note that this represents time where students make homework in the class.

total teaching time over p activities, i.e. $x_1, ..., x_p$, and where $p = 8$ in the setting at hand. The total teaching is thus the sum of time allocated to each of the p activities, $\sum x = t$.

In a first best optimum, each teacher chooses the set of teaching activities (i.e., the teaching style) that maximizes the test score of each individual student. In other words, the first best optimum estimates the maximal feasible test score for students if each student would receive the optimal configuration of teaching activities. A natural strategy to estimate the student specific optimal teaching style would consist of comparing the student to similar students in the sample. Students can be considered as similar if they obtained similar (or less) teaching hours, and if they have similar characteristics. The former is discussed in the first subsection (i.e., a deterministic model), while the latter is discussed in the subsection that follows (i.e., a conditional model).

A. Deterministic model

Intuitively, the optimal configuration of the set x can be deduced from comparing each student (denote the evaluated student by 'o') to the other students ($j = 1, ..., o, ..., n$) in the sample Ψ. The optimal lecturing style for student o is the set $\nu_i x_{io}$. The setting is analogous to a Data Envelopment Analysis (DEA) setting (Charnes et al., 1978). DEA is a nonparametric technique that estimates the performance of entities without making *a priori* assumptions on the shape of the production technology. In particular, it allows us to examine which configuration of x maximizes the output y (i.e., output-oriented model; for alternative settings, see Fried et al., 2008). To do so, DEA determines by linear programming virtual weights ν_i on the inputs x. In its (dual) multiplier formulation, the deterministic output-oriented DEA model is formulated as:[3]

$$\min_{\nu_i} \theta = \sum_{i=1}^{p} \nu_i x_{io} \tag{1}$$

such that:

$$\begin{array}{lr} (1)\ \sum_{i=1}^{p} \nu_i x_{ij} - \mu_r y_j \geq 0, & \text{(feasibility constraint)} \\ (2)\ \mu_r y_o = 1, & \text{(normalization constraint)} \\ (3)\ \mu_r, \nu_i \geq \epsilon > 0, & \text{(positive weights)} \end{array}$$

where θ represents the efficiency score and ϵ a small positive number. If $\theta > 1$, output y_o can be increased by $(\theta - 1)\%$ if the teacher would teach according to the student-specific optimal teaching style $\nu_i x_{io}$. If $\theta = 1$, teachers employ the optimal teaching style, and, thus, are relatively efficient. Multiplying the virtual weights on the inputs by the inputs, $\nu_i x_{io}$, results in the first best optimal teaching style, or the optimal configuration of x for a given output score y_o. However, the latter

[3]In the current setting, constant returns to scale are a straightforward assumption as the set of lecturing activities sums for all students to t.

is the observed output score and not the optimal output score. If the teacher was teaching in a non-optimal way (i.e., if $\theta > 1$), the output score could be further increased by teaching optimally. In addition, as can be observed from Model 1, the weights sum to the inefficiency score (and thus do not necessarily equal 1). In the current setting, we are less interested in the inefficiency, but the more in the optimal use of the inputs (cfr. the benefit of the doubt model of Cherchye et al., 2007). Inefficiency is ignored by projecting the individual test scores of the students to their highest feasible output level (i.e., $y^* = y \cdot \theta$) and re-estimating the DEA model. In this way, the resulting the weights add up to one (no inefficiency) and represent the optimal teaching style.

B. Robust and conditional model

The DEA model estimates the maximal feasible output for each student by comparing the evaluated student with all students in the sample. As a consequence, the estimate (1) is vulnerable to extreme best practices and (2) ignores heterogeneity among students and teachers.

With respect to the first issue, extreme best practices could arise from measurement errors and anomalies in the data. They are problematic, in the sense that they heavily influence the estimated optimal configuration of teaching activities for other students in the sample by setting unrealistic optimal teaching styles. Therefore, following Cazals et al. (2002) and Daraio and Simar (2007), we limit the influence of these outlying observations by drawing repeatedly and with replacement a reference set, Ψ^m, consisting of m observations which are using less or equal teaching time than the evaluated observation (so-called order-m model). Relative to this draw of m observations, Model 1 is estimated. As the reference set consists only of m drawn observations, the evaluated entity will not always be included in its reference set Ψ^m. As such, super-efficient observations could arise (i.e., $\theta < 1$). The latter denotes that the observation is doing better than the m reference observations in the sample. The impact of outlying observations is mitigated by repeating this procedure B times after which the virtual weights and the efficiency estimates are averaged. Besides mitigating the impact of outlying observations, the order-m model of Cazals et al. (2002) is attractive in the setting at hand thanks to its statistical features. In particular, Jeong et al. (2010) showed the consistency of the model and its fast rate of convergence (for a discussion see De Witte and Kortelainen, 2009).

With respect to the second issue, background characteristics of students and teachers differ. For example, some of the students are raised in advantageous environments or have higher intellectual capacities, while others do not. As teachers generally know the characteristics of their students, they account for it in their teaching style. The procedure at hand accounts for student heterogeneity by adapting the DEA model to the conditional efficiency estimations of Daraio and Simar (2005, 2007). The conditional efficiency model only compares like with likes by smoothly adapting the order-m procedure such that the m observations are drawn by a probability. In particular, a multidimensional Kernel is estimated around the background characteristics of the evaluated entity. Using the Kernel, we obtain probability weights by which the m observations are drawn. As the characteristics include both continuous and discrete variables, we use the conditional model adaption of De Witte and Kortelainen (2009) who suggested a mixed Kernel for estimating the conditional efficiency. Relative to this new reference set $\Psi^{m,c}$ the optimal configuration of $\nu_i x_{io}$, and, thus, the optimal teaching style is estimated.

Besides allowing for different background characteristics of the students, comparing like with likes is important to allow for potential selection effects. Selection effects represent the fact that students with more favorable background characteristics have a higher probability to self-select themselves in better schools. Ignoring selection effects, could result in biased results. For example, we could find that teachers choose non-optimal teaching styles, while in reality these teachers give instruction to low ability students, while other teachers give instruction to high ability students. Although teaching styles affect the math performance of students, at the same time, the chosen teaching styles and math performance is affected by the ability of the students. In the analysis, we should therefore condition on characteristics that determine the selection of high ability students to better schools, such as ethnic background and the education level of the parents. In other words, we obtain accurate estimates if we compare like with likes (see also Van Klaveren, 2009).

3.4 Second best optimum

In real life, teachers cannot teach every student individually, but are more likely to choose one teaching style for all students in the class. Therefore, we arrive at a second best optimum in which an optimal teaching style is deduced for all students in the class.

A. Uniform teaching style

To compute a uniform optimal teaching style, the math test score of each student is replaced by the average test score of the class (we will relax this assumption later on), after which the conditional DEA model is estimated (see Section 3.3). Performing the second best analysis (with the average test scores as output) creates an attractive and policy relevant feature. First consider a homogeneous class in which all students have a similar background. Estimating the optimal teaching style by the outlined conditional efficiency model will result in a similar teaching style for all students in the class. Indeed, in the conditional efficiency model, all students will be drawn with a similar probability. Next, consider a class in which students have heterogeneous backgrounds. When some students in the class are significantly different from the other students (i.e., have different background characteristics), the optimal teaching style will be different from the other students in the class. Indeed, in the conditional efficiency model, students of the class will be drawn with a different probability (in other words, their 'likes' are in different classes). In the Netherlands, similar to other other countries, the teaching style difference from the other students in the class corresponds to students who (should) obtain additional student counseling (e.g., additional training on maths, logopeadics or pedagogical training) because of their disadvantageous background (so-called care-students). As such, the second best model at hand allows us to detect students who should need additional help.[4]

B. Revealed teaching style

In a second phase, within a class, teaching styles are allowed to be differentiated according to the intellectual capacities of the student. We assume that the teacher differentiates according to two types of students (although extension to more types is straightforward). Define by $\alpha \epsilon [0, 1]$ a cut-off level in the class distribution of test scores (e.g., $\alpha = 0.25$). Students obtaining test scores above the cut-off level α (i.e., the first quartile of the class distribution) will be treated differently than students obtaining test scores below the cut-off level α. To do so, for the students performing better than α (i.e., better than the first quartile), output test scores are adapted to the average test score of the $(1 - \alpha)$ best students (i.e., 75% best students). Similarly, for students performing worse than cut-off level α (i.e., in the first quartile), output test scores are adapted to the average test score of the α bottom students (i.e., 25% bottom students). As such, this approach accounts for differentiating teaching style along the student intellectual capacities. This second best solution corresponds to common practice in education where students obtain additional instruction time according to their background and their performance at school.

Once the model is estimated for various cut-off points α, it can be compared to the teaching style revealed by the teacher. The latter information, included in the TIMSS 2003 data set, monitors the average time that teachers spent on teaching activities of a teacher (average in the sense that it is estimated over one school year). Using a sensitivity analysis on α, we can compare the modeled teaching style to the revealed teaching style by the teacher. This provides us an indication on which group of students the teacher is focusing. In other words, it reveals the value of α.

4 Empirical results

The TIMSS data provide full information on 1,790 Dutch students who are educated by 82 different math teachers in an equal amount of different classes and schools. The data provide a representative sample of students for the Netherlands. As outlined before, input variables in the analysis include the time for the teaching activities: time allocated to (1) homework, (2) lecturing in front of the class, (3) problems with guidance, (4) problems without guidance, (5) reteaching, (6) tests and quizzes, (7) class room management, and (8) other activities. The inputs sum to the total teaching time t. The output variable represents the test score on math. The model includes seven exogenous variables which have been indicated in the literature as highly detrimental for educational attainments (see, e.g., Astone and McLanahan, 1991; Hanushek, 2003; Hanushek and Welch, 2006 and references therein): (1) the highest education of the mother (as a proxy for the motivational and intellectual

[4]To allow for practical policy implementation, the R code is available upon request.

Table 1: Descriptive Statistics

	Obs.	Mean	St. Dev.	Min.	25 centile	50 centile	75 centile	Max
Output								
Math score	1790	545.34	66.26	290.75	495.09	551.70	594.92	720.95
Input: (time devoted to)								
Homework	1790	23.41	15.41	0.0	13.5	18.8	30.0	75.0
Lecturing	1790	21.43	12.11	6.8	13.5	15.0	30.0	60.0
Problems with guidance	1790	34.20	29.60	0.0	9.4	22.5	60.0	124.2
Problems without guidance	1790	38.82	30.92	0.0	13.5	37.5	60.0	120.0
Reteaching	1790	10.71	7.47	0.0	6.8	7.5	15.0	33.8
Quizzes	1790	12.00	7.65	0.0	7.5	12.0	15.0	40.0
Classroom management	1790	6.901	8.16	0.0	1.8	6.8	7.5	50.0
Other activities	1790	5.008	4.85	0.0	0.0	5.4	7.5	30.0
Background:								
Mother's education	1790	6.06	2.30	1.0	4.0	6.0	9.0	9.0
Language (1 = native)	1790	1.20	0.51	1.0	1.0	1.0	1.0	4.0
Books at home	1790	3.23	1.21	1.0	2.0	3.0	4.0	5.0
Class size	1790	24.42	4.14	12.0	22.0	24.0	28.0	32.0
Teacher is male	1790	1.64	0.48	1.0	1.0	2.0	2.0	2.0
Teacher Education	1790	2.01	0.37	1.0	2.0	2.0	2.0	3.0
Teacher Experience	1790	24.42	9.50	1.0	6.0	16.0	23.0	36.0

influence that parents have), (2) mother tongue of the student, (3) the number of books at home (as a proxy for cultural interests of the parents), (4) the class size (as a proxy for potential student heterogeneity), and (5) sex, experience and education level of the teacher (as a proxy of teacher quality).[5] Summary statistics are presented in Table 1. The optimal teaching style is estimated using the conditional and robust model outlined in Section 3.1.

Below we clarify the descriptives of the input and background variables. The input variables represent the time that teachers devote to several teaching activities. The mean represents how many minutes teachers usually spend on a certain activity per week. The highest education level of the mother is measured on a nine-point scale, and the lowest value means having no or only a few years of primary schooling and the highest value means that the mother is having a university degree. The mother tongue of the student is measured on a four point scale, with one meaning that the student always speaks Dutch at home, and four meaning that the student never speaks Dutch as home. The number of books at home is measured on a five point scale and the lowest value represents having between zero and ten books at home, while the highest value represents having more than 200 books at home. Class size and teacher experience represent the number of students in the class and the years of teaching experience, respectively. The highest education level of the teacher is measured on a three point scale, where the lowest education level is a secondary general

[5]Further robustness analysis with ethnicity at the school delivered similar results.

or vocational education level and the highest education level is having a university degree or higher. We note that most teachers have a higher vocational education level as the highest education level (category 2). Furthermore, the teachers with the lowest education level are teachers in training who are usually under supervision of a more experienced teacher.

4.1 Individual teaching: First best optimum

A. Without accounting for background characteristics

In the first best optimum, teachers adopt a teaching style that maximizes the math test score for each individual student. Let us first consider the unconditional version of the model (i.e., without accounting for student and teacher background). This could represent teachers in apprenticeship or short replacements who do not observe information on student background. In this case, teachers divide their time optimally over the different teaching activities given the number of teaching hours available. This model corresponds to the so-called unconditional robust DEA model.

To illustrate the procedure, Figure 1 shows the revealed (outer ring) and first-best (inner ring) optimal teaching style for a random student. We present both the revealed teaching (as included in the data by the teacher) and the computed first best teaching. For this particular student, the optimal configuration of teaching activities would consist mainly of problem solving with and without guidance (respectively 37 and 38% of the time) and is complemented by lecturing in front of the class (7%), giving homework (8%) and reteaching (6%). Particularly problems with guidance, problem solving without guidance and homework should increase in the (revealed) teaching style to obtain the first best optimum. For this student, we find an inefficiency score of 8%. This result shows that there are students, with less or equal teaching hours, that score better than the student under study.

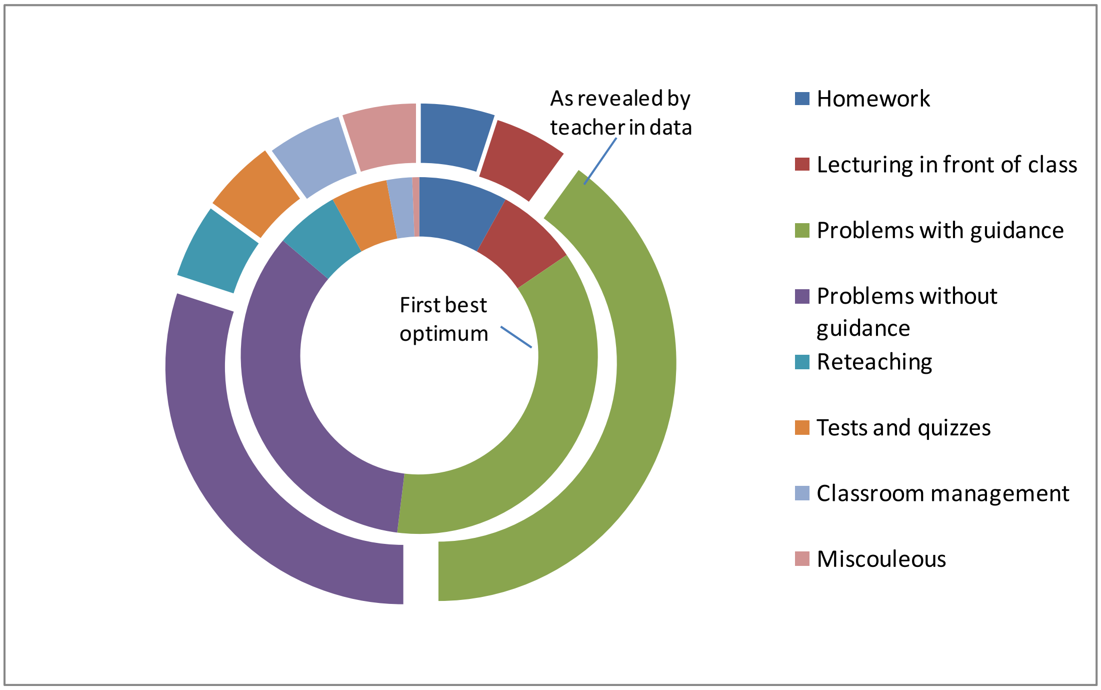

Figure 1: Revealed and first best teaching style for a random student

The aggregated results for all students are presented in Table 2. The first column shows the teaching activities and columns 2 up to 6 provide the summary statistics of the optimal weight distribution for each of the teaching activities we distinguish.[6] We present both the minimum, maximum, mean and quartiles of the underlying student data. Although not extensively discussed in the text, the quartiles are insightful as they present the distribution of the endogenous weights. In the discussion below we focus on the mean values. The last row of the table shows the inefficiency scores. The table shows that lecturing in front of the class, homework and problem solving with and without guidance are dominant in the optimal configuration of teaching activities. That the latter three activities are important drivers of student performance corresponds to the didactics literature suggesting that students are active learners and benefit most from active participation to the class (see Bonwell and Eison, 1991; Domjan, 2009 and references therein).

We find an average inefficiency score of 14 % suggesting that test scores of students could have been 14 % higher if the optimal teaching style was adopted by the teachers. At this stage, it is important to stress that this is not a causal effect. We merely determine the first-best optimal teaching style and then simulate for all students the test scores conditional on this optimal teaching style (causal interpretations are considered as scope for further research).

[6]Results at the student level are available upon request.

Table 2: Unconditional First-Best Estimates

	Min.	**1st Quartile**	**Mean**	**3rd Quartile**	**Max**
Homework	0.00	0.13	0.22	0.30	0.51
Lecturing	0.00	0.13	0.18	0.25	0.86
Problems with guidance	0.00	0.13	0.21	0.43	0.96
Problems without guidance	0.00	0.07	0.29	0.50	0.90
Reteaching	0.00	0.00	0.08	0.14	0.42
Test and Quizzes	0.00	0.00	0.03	0.06	0.28
Classroom management	0.00	0.00	0.02	0.05	0.28
Other activities	0.00	0.00	0.00	0.02	0.12
Inefficiency	1.00	1.08	1.14	1.25	2.05

B. With accounting for background characteristics

Except, maybe, for temporary teachers and teachers in apprenticeship, most teachers observe student background characteristics and account for these in the lecturing style they adopt. Moreover, and as explained in Chapter 3.1, there may be selection effects. Hence, better performing students with more favorable background characteristics and better teachers who are higher educated or more experienced may self-select in better schools (e.g., Van Klaveren, 2009 and references therein). On the basis of an unconditional analysis, we could find that teachers adopt a non-optimal teaching style, while in reality these teachers give instruction to low ability students for whom the adopted teaching style is in fact optimal. More generally, the adopted teaching styles affect the math performance of students, but at the same time both math performance and the adopted teaching style is affected by the underlying ability of the students and teachers. In the analysis, we should therefore condition on characteristics that determine the selection of high ability students to better schools, such as ethnic background, the education level of the parents and the experience and education level of teachers.

We switch from an unconditional to a conditional DEA model and accommodate the optimal teaching style for (1) student characteristics (native language, education of the mother and the number of books at home); (2) school characteristics (class size), and (3) teacher characteristics (sex, experience and education level). By conditioning on the above mentioned student, teacher and school characteristics we compare teaching styles and student performance only in comparable situations. For example, the performance of an immigrant student with a lower educated mother and an experienced and highly educated teacher is compared to other immigrant students with comparable teachers and mothers, but the teaching style adopted by the teacher may vary. Hence, controlling for background and selection characteristics in a DEA setting means that we combine non-parametric DEA with a kernel matching approach, simulate a control group for each student in the sample and determine the optimal configuration of teaching activities.

The conditional results are shown in Table 3 and the results differ significantly from the uncon-

Table 3: Conditional First-Best Estimates

	Min.	**1st Quartile**	**Mean**	**3rd Quartile**	**Max**
Homework	0.00	0.01	0.06	0.15	1.00
Lecturing	0.00	0.01	0.03	0.08	1.00
Problems with guidance	0.00	0.05	0.13	0.77	1.00
Problems without guidance	0.00	0.03	0.27	0.87	1.00
Reteaching	0.00	0.00	0.01	0.01	1.00
Test and Quizzes	0.00	0.00	0.01	0.03	0.95
Classroom management	0.00	0.00	0.00	0.00	1.00
Other activities	0.00	0.00	0.00	0.00	0.09
Inefficiency	0.99	1.07	1.11	1.17	1.50

ditional results. Again, solving problems with and without guidance are dominant in the optimal configuration of teaching activities, which is not surprising given the pedagogical and didactics literature (Bonwell and Eison, 1991; Domjan, 2009).[7] However, lecturing in front of the class and homework, that received relatively large weights in the unconditional version, do not receive large weights in the conditional model. Given the nature of the conditional theoretical model this result is not so surprising, since the model assumes that each teacher adopts the best teaching style for each student given his or her background characteristics. Problem solving seems a relatively more 'personal' teaching style, and given that teachers are not (yet) constraint by the fact that they teach a class of students, it is intuitive that they adopt a teaching style where most of the weight is given to activities that can be associated with a personal teaching style.

The inefficiency in the conditional model is 11% and this is lower than the 14% inefficiency of the unconditional model. This is intuitive, because in the conditional model we compare students with a control group of students who have similar background characteristics. For example, in the unconditional model it can happen that we compare the student performance of low ability students to that of high ability students given that these students received at least the same amount of teaching time. However, it is not realistic that teachers can increase the student performance of low ability students to that of high ability students by means of the teaching style they adopt. By conditioning on variables that (partly) capture the ability of the students we take into account that the increase in student performance by the adopted teaching style is limited and given that there is less space for improvement in the conditional setting, it follows that the inefficiency is lower. Nevertheless, the conditional model still shows an inefficiency score of 11%. This means that students score a 11% higher test score when we compare them to other students with similar student, teacher, school characteristics and teaching time received, but with teaching styles that put more weight on problem solving with or without guidance.

[7] Individual results available upon request.

Table 4: Conditional Second Best Estimates

	Min.	**1st Quartile**	**Mean**	**3rd Quartile**	**Max**
Homework	0.00	0.00	0.12	0.01	1.00
Lecturing	0.00	0.00	0.07	0.01	1.00
Problems with guidance	0.00	0.00	0.29	0.87	1.00
Problems without guidance	0.00	0.00	0.46	1.00	1.00
Reteaching	0.00	0.00	0.02	0.00	1.00
Test and Quizzes	0.00	0.00	0.03	0.00	0.98
Classroom management	0.00	0.00	0.01	0.00	0.88
Other activities	0.00	0.00	0.00	0.00	0.001
Inefficiency	1.00	1.00	1.00	1.08	1.61

4.2 Teaching a class of students: Second best optimum

In practice, teachers are unlikely to adopt an optimal teaching style for each individual student because they are teaching a class of students. Hence, the conditional first best optimum seems unrealistic: it assumes that teachers adopt an optimal teaching style for all n students in the class while there are on average 24 students per class (see Table 1) and while the student population per class is rather heterogeneous. Therefore, we will switch now from a conditional first best model to a conditional second best model, where we take into account that teachers adopt an optimal teaching style for a class of students.

Let us first consider a situation where the teacher adopts a uniform optimal teaching style, i.e. the teacher is assumed to maximize the test score of the average student in the class. The intuition behind such a teaching style, is that a teacher can never adopt a teaching style that serve all students well, but by focusing on the average student (s)he at least adopts a teaching style that is beneficial to the majority of students. The conditional second best estimates are reported in Table 4.

When we assume that teachers focus on the average student, we obtain estimates that are rather similar to the first best conditional results: high test scores are associated with teaching styles that put the emphasis on problem solving with and without guidance and homework. So also this view produces results that are consistent with the pedagogical and didactics literature.

Let us assume that the teachers in our sample focus on the average student.[8] Differences between the estimated conditional first and second best optimal teaching styles then give information on whether the educational system succeeds in providing each student his/her optimal teaching style: the more similar both teaching styles are, the more the student is at his/her place in the class. More generally, when student classes are more homogeneous then individual students compare better to the average student in the class and hence a class based second best teaching style will correspond more closely to a student based individual teaching style.

If the second best teaching style differs significantly from the first best teaching style, then

[8]Different assumptions are estimated in the next subsection, for which detailed outcomes are available upon request.

students could be better of if they would be in a different (for them more homogeneous) class. To have an idea on how well the first best teaching activities resemble the second best activities we refer to a teaching activity as different if the second best time allocated to the activity does not correspond to the first best time +/- 10% of time (sensitivity analysis using other deviations than 10% gave results that were very similar). Algebraically, the first and second best activities are considered as similar if:

$$\nu_i^{first} x_{io} * 0.9 \leq \nu_i^{second} x_{io} \leq \nu_i^{first} x_{io} * 1.1.$$

In our study a teaching style consists of 8 teaching activities, and based on the rule formulated above, we determine for each activity whether the first and second best activities are different. Then we compute a percentage that indicate to what degree the first and second best activities compare. As an illustration, assume that there are three teaching activities and that we observe for the second best teaching style (0.6655; 0.2612; 0.0733), and for the first best teaching style (0.7320; 0.2508; 0.0172). The weights given to the individual teaching activities add up to one, such that the total teaching time is comparable (i.e. equals one). Applying the rule formulated above gives an indicator vector (1,1,0), where zero (one) indicates that the first best activity does not correspond (correspond) to the second best activity. The degree of correspondence is then 66% since 2 of the 3 first best teaching activities correspond to the second best teaching activities.

In Figure 2 we show the degree of correspondence for all students in the sample. The Figure shows that students do not obtain their first best optimal lecturing style. For 78% of the students, the second best lecturing style differs more than 50% from the first best lecturing style (i.e., out of the 8 teaching activities, at least 4 do not correspond between first and second best).

This poor match between first and second best optima could indicate that Dutch students could be better of if they would be in more homogeneous classes. This corresponds to the intuition and previous literature for two reasons. Firstly, the relatively high ethnic and social segregation in Dutch education is well documented (e.g., Dronkers, 1995; Karsten et al., 2006; Ritzen et al., 1997). Students are to a large extent segregated in schools following their background, and less following their abilities. Secondly, the underlying data for the estimates correspond to the second year of secondary education where all students take a similar 'basic' track. Therefore, student diversity in terms of abilities is still very large. The results may suggest that a common track does not allow for an optimal teaching style: if first and second best lecturing styles would correspond better, student test scores would be higher.

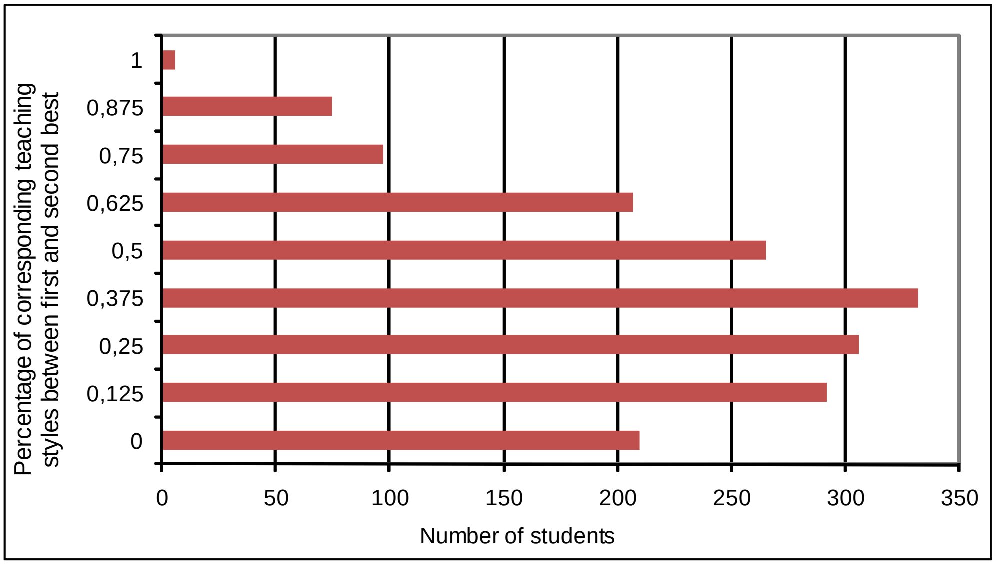

Figure 2: Comparison between first and second best teaching style

In the comparison above we have assumed that teachers focus on the average student in the class, but we do not know whether this is the case. Therefore in the next section we try to discover on who teachers are focusing in the class when they decide on the teaching style they adopt.

4.3 Who are teachers teaching for? Revealed Teaching Style

In the TIMSS questionnaire, teachers indicate the teaching activities they undertake to educate their class (i.e., the inputs of the DEA model). This set of teaching activities corresponds to the revealed teaching style. If we compare the second best teaching style for different student groups in the class to the revealed teaching style, we can discover on which students in the class the teacher is targeting his or her teaching style.

Let α be a parameter that indicates the fraction of students the teacher is targeting in his lecturing style. Hence, similar to before, $\alpha \epsilon [0, 1]$ defines a cut-off level in the class distribution of test scores. For the students performing better than α, output test scores are adapted to the average test score of the $(1-\alpha)$ best students. We allow α to vary between 0.05 and 1, and compute for each of the values of α the conditional second best model. To see for which value of α the conditional second best teaching style corresponds best to the revealed teaching style, we perform a grid search (a technique which is extensively used in, e.g., cross validation bandwidth selection). As such, we find the α parameter for each teacher that minimizes the distance between the revealed and second best teaching style.

So, suppose that by varying α, we find that the revealed teaching style corresponds best to a conditional second best optima with $\alpha^* = 0.40$. We then learn that the teacher has targeted his teaching style on the 40 % least performing students in order to maximize their test scores.

Again, we use the following rule to express to what degree the revealed and second best teaching activities compare:

$$\nu_i^{second} x_{io} * 0.9 \leq \nu_i^{revealed} x_{io} \leq \nu_i^{second} x_{io} * 1.1.$$

The results are graphically presented in Figure 3. We find a positive relationship between the α parameter and the degree to which the revealed and second best teaching activities compare.[9] The revealed teaching style matches best to a conditional second best optimum if $\alpha^* = 70$. It seems that teachers are targeting their teaching style on the 70 % least performing students. This is a rather intuitive result, because the teaching style is optimal for the majority of students, and is focused on those students who need attention the most. The choice for this particular teaching style may come from the teachers incentive not to 'loose' students in the class during the school year. By adopting a teaching style that suits most of the students in the class, and by differentiating among students in the homework and problem solving activities, a teacher seems to optimize the performance of the students.

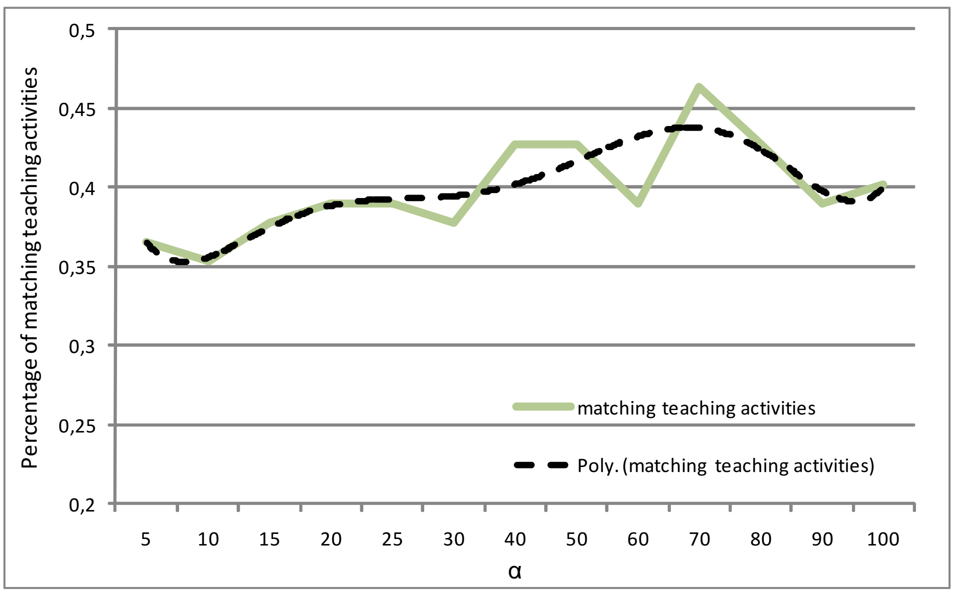

Figure 3: Comparison between second best and revealed teaching style

5 Conclusion

In this study we examine what configuration of teaching activities maximizes student performance and, furthermore, on which students in the class the teacher is targeting his or her teaching style. For

[9]As the grid is determined on 0.05 (i.e., α increases by 0.05) some discontinuity arises in the figure. Therefore, we also present the polynomial trend line of order 6. The trend line hides the discontinuity of the grid and shows a clear trend in the results.

this purpose we formulate a nonparametric efficiency model that is rooted in the Data Envelopment Analysis (DEA; Charnes et al., 1978) literature. To illustrate the nonparametric model, we use data on the math performance of Dutch students who are in their second year of secondary education, which is taken from the Trends in International Mathematics and Science Study (TIMSS) 2003 data.

Introducing Data Envelopment Analysis into the literature of teacher quality and performance literature has several advantages. First of all, the non-parametric nature is attractive as, by definition, it does not require an a priori specified educational production function. Often there is no *a priori* information on the underlying production technology, and as has been shown by Yatchew (1998) and Rothstein (2008) models that assume some parametric educational production function are prone to model misspecification errors. Second, by addressing a conditional version of the non-parametric model, we can take into account that better students and teachers may self-select in better schools. In essence, conditioning on covariates in a non-parametric DEA setting is comparable to a kernel matching approach, since for each student an appropriate control group of students is simulated to determine the optimal configuration of teaching activities. A third advantage lies in the endogenous weight specification that allows different teaching activities to be complements rather than substitutes, in contrast to previous literature (e.g., Rouce and Krueger, 2004; Banerjee et al. 2007; Barrow et al., 2009; Schwerdt and Wuppermann, 2009 and Van Klaveren, 2009). Complementing lecturing activities correspond better to the real life activities of the teachers.

In the analyses we distinguish between a first best and a second best model. The first best model assumes that teachers adopt an optimal teaching style for each student *individually*, while the second best model assumes that teachers adopt an optimal teaching style for the average student in the class. We find for both models that high test scores are associated with teaching styles that put the emphasis on problem solving with or without guidance and homework, which is consistent with the pedagogical and didactics literature.

Despite the fact that both models put the emphasis on the same teaching activities, we find that a second best teaching style differs significantly from a first best teaching style. This result suggests that students could perform better if they would be in a different (for them more homogeneous) class. This corresponds to the intuition and previous literature for two reasons. First, the relatively high ethnic and social segregation in Dutch education is well documented (e.g., Dronkers, 1995; Karsten et al., 2006; Ritzen et al., 1997). Students are to a large extent segregated in schools following their background, and less following their abilities. Secondly, the underlying data for the estimates correspond to the second year of secondary education where all students take a similar 'basic' track. Therefore, student diversity in terms of abilities is still very large. The results may suggest that a common track does not allow for an optimal teaching style: if first and second best lecturing styles would correspond better, student test scores would be higher.

In the second best model we assume that teachers focus on the average student in the class, but we did not verify this. However, by comparing the teaching activities that teachers undertake to educate their class, i.e. the revealed teaching style, with the second best teaching style we can discover on which students in the class the teacher is targeting his or her teaching style. To be more precise, we estimate the second best model for different fractions of the least performing students in the class and then determine for which fraction the distance between the revealed and second best teaching style is minimized.

We find that the revealed teaching style is most similar to a conditional second best optimum if the teacher is targeting his or her teaching style on the 70 % least performing students. This is a rather intuitive result, because the teaching style is optimal for the majority of students, and is focused on those students who need attention the most. The choice for this particular teaching style may come from the teachers incentive not to 'loose' students in the class during the school year.

References

Aaronson, Daniel, Lisa Barrow and William Sander (2007), 'Teachers and student achievement in the chicago public high schools', *Journal of Labor Economics* **25**, 95–135.

Aslam, M. and G. Kingdon (2007), What can teachers do to raise pupil achievement, working paper series, The centre for the study of african economies.

Astone, N.M. and S.S. McLanahan (1991), 'Family structure, parental practices and high school completion', *American Sociological Review* pp. 309–320.

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

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

Bonwell, C.C. and J.A. Eison (1991), 'Active Learning: Creating Excitement in the Classroom. ERIC Digest.', *ASHE-ERIC Higher Education Reports, The George Washington University* .

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

Cazals, C., J.P. Florens and L. Simar (2002), 'Nonparametric frontier estimation: a robust approach', *Journal of econometrics* **106**(1), 1–25.

Charnes, A., W.W. Cooper and E. Rhodes (1978), 'Measuring the efficiency of decision making units', *European journal of operational research* **2**(6), 429–444.

Cherchye, L., W. Moesen, N. Rogge and T.V. Puyenbroeck (2007), 'An introduction to benefit of the doubt composite indicators', *Social Indicators Research* **82**(1), 111–145.

Clotfelter, Charles T., Helen F. Ladd and Jacob L. Vigdor (2006), Teacher-student matching and the assessment of teacher effectiveness, NBER Working Papers 11936, National Bureau of Economic Research, Inc.

Clotfelter, Charles T., Helen F. Ladd and Jacob L. Vigdor (2007), Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects, NBER Working Papers 13617, National Bureau of Economic Research, Inc.

Daraio, C. and L. Simar (2005), 'Introducing environmental variables in nonparametric frontier models: a probabilistic approach', *Journal of productivity analysis* **24**(1), 93–121.

Daraio, C. and L. Simar (2007), 'Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach', *Journal of Productivity Analysis* **28**(1), 13–32.

De Witte, K. and M. Kortelainen (2009), 'Blaming the exogenous environment? Conditional efficiency estimation with continuous and discrete exogenous variables', *MPRA Paper* .

Domjan, M. (2009), *The principles of learning and behavior: active learning edition*, Wadsworth Pub Co.

Dronkers, J. (1995), 'The existence of parental choice in the Netherlands', *Educational Policy* **9**(3), 227.

Fried, H.O., C.A.K. Lovell and S.S. Schmidt (2008), *The measurement of productive efficiency and productivity growth*, Oxford University Press, USA.

Hanushek, E.A. (1997), 'Assessing the effects of school resources on the student performance: An update', *Educational Evaluation and Policy Analysis* .

Hanushek, E.A. (2003), 'The failure of input-based schooling policies', *Economic Journal* pp. 64–98.

Hanushek, E.A. and F. Welch (2006), *Handbook of the Economics of Education*, North-Holland.

Hanushek, Eric A. and Steven G. Rivkin (2006), *Teacher Quality*, Vol. 2 of *Handbook of the Economics of Education*, Elsevier, chapter 18, pp. 1051–1078.

Hanushek, Eric, John F. Kain and Steven G. Rivkin (2005), 'Teachers, schools, and academic achievement', *Econometrica* **73**(2), 417–458.

Hjalmarsson, L., S. Kumbhakar and A. Heshmati (1996), 'Dea, dfa and sfa: a comparison', *Journal of Productivity Analysis* **7**(2), 303–327.

Jeong, S.O., B.U. Park and L. Simar (2010), 'Nonparametric conditional efficiency measures: asymptotic properties', *Annals of Operations Research* pp. 1–18.

Karsten, S., C. Felix, G. Ledoux, W. Meijnen, J. Roeleveld and E. Van Schooten (2006), 'Choosing segregation or integration?: The extent and effects of ethnic segregation in Dutch cities', *Education and Urban Society* **38**(2), 228.

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

Murnane, R.J. and B.R. Phillips (1981), 'Learning by doing, vintage, and selection: Three pieces of the puzzle relating teaching experience and teacher performance', *Economics of Education Review* **1**(4), 453–465.

Ritzen, J.M.M., J. van Dommelen and F.J. De Vijlder (1997), 'School finance and school choice in the Netherlands', *Economics of Education Review* **16**(3), 329–335.

Rothstein, J. (2008), Teacher quality in educational production: Tracking, decay, and student achievement, NBER Working Papers 14442, National Bureau of Economic Research, Inc.

Rouce, C.E. and A.B. Krueger (2004), 'Putting computerized instruction to the test: a randomized evaluation of a "scientifically based" reading program', *Economics of Education Review* **23**(4), 411–430.

Schwerdt, Guido and Amelie C. Wuppermann (2009), 'Is traditional teaching all that bad? a within-student between subject approach', *working paper* .

Smith, J.B, V.E. Lee and F.M. Newmann (2001), Instructions and achievement in chicago elementary schools, Report, consortium on chicago school research, Chicago.

Van Klaveren, C. (2009), 'Lecturing style teaching and student performance', *Working Paper* .

Wayne, A.J. and P. Youngs (2003), 'Teacher characteristics and student achievement gains: A review', *Review of Educational Research* **73**(1), 89–122.

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

Yatchew, A. (1998), 'Nonparametric regression techniques in economics', *Journal of the Economic Literature* **36**, 669–721.

TIER WORKING PAPER SERIES
TIER WP 10/23

ISBN 978-94-003-0022-4