CPB Netherlands Bureau for Economic
Policy Analysis

CPB Discussion Paper | 294

Teacher quality and student achievement

Evidence from a Dutch sample of twins

Sander Gerritsen
Erik Plug
Dinand Webbink

Teacher quality and student achievement:
Evidence from a Dutch sample of twins

Sander Gerritsen*
Erik Plug
Dinand Webbink

Abstract

This paper examines the causal link that runs from classroom quality to student achievement using data on twin pairs who entered the same school but were allocated to different classrooms in an exogenous way. In particular, we apply twin fixed-effects estimation to assess the effect of teacher quality on student test scores from second through eighth grade, arguing that a change in teacher quality is probably the most important classroom intervention within a twin context. In a series of estimations using measurable teacher characteristics, we find that (a) the test performance of all students improve with teacher experience; (b) teacher experience also matters for student performance after the initial years in the profession; (c) the teacher experience effect is most prominent in earlier grades; (d) the teacher experience effects are robust to the inclusion of other classroom quality measures, such as peer group composition and class size; and (e) an increase in teacher experience also matter for career stages with less labor market mobility which suggests positive returns to on the job training of teachers.

JEL Codes: I2, J4

Keywords: economics of education, teacher quality, teacher experience, twins

* Sander Gerritsen: Erasmus School of Economics Rotterdam, 3000 DR Rotterdam, The Netherlands, CPB, Netherlands Bureau for Economic Policy Analysis, The Hague. Erik Plug: Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands, Tinbergen Institute, IZA and UCLS. Dinand Webbink: Department of Economics, Erasmus School of Economics, 3000 DR Rotterdam, The Netherlands, Tinbergen Institute and IZA. We thank seminar participants at the University of Amsterdam, Rotterdam, Tinbergen Institute, Aarhus and Torino.

1. Introduction

The quality of teachers is considered to be a crucial factor for the production of human capital. Understanding the determinants of teacher quality is important for improving the quality of education and therefore a key issue for educational policy. A large literature has investigated the contribution of teachers to educational achievements of students, the heterogeneity between teachers and the aspects of teachers that are important (e.g. Hanushek & Rivkin, 2006; Staiger & Rockoff, 2010). A consistent finding in the literature is that teachers are important for student performance and that there are large differences among teachers in their impacts on achievement. However, little evidence has been found that any observable characteristic, save experience, explains the variation between teachers. Teacher experience only seems to matter in the initial years in the profession[1]. Hence, the literature does not yet provide clear policy advice about the type of teachers that are most effective, and therefore should be hired and kept in the education profession based on their observed characteristics.

Estimating the effect of teacher characteristics on student performance is complicated because students, teachers and resources are almost never randomly allocated among schools and classrooms. Unobserved factors correlated with both teacher characteristics and student outcomes might bias estimates using non-experimental data. In fact, recent studies have provided evidence for non-random sorting of teachers (Clotfelter et al. 2006; Feng 2009). Researchers have addressed this issue by using panel methods or exploiting random assignment of students and teachers into schools and classrooms. The most common approach in the literature is to estimate value-added models that focus on gains in student achievement and eliminate confounding by past unobserved parental and school inputs (Hanushek 1971, 1992; Aaronson, Barrow and Sander 2007; Rockoff 2004; Rivkin et al. 2005; Hanushek et al. 2005). Several recent studies exploit multiple years of information for teachers to estimate teacher fixed effects and to link these effects with teacher characteristics (Hanushek et al. 2005; Rockoff 2004; Aaronson, Barrow and Sander 2003; Rivkin et al. 2005). Although the most sophisticated value-added models use a three-way-fixed effects

[1] Recent studies find gains from teacher experience beyond the initial years in the career (Wiswall, 2013, Harris and Sas 2011). Mueller (2013) finds that teacher experience moderates class size effects. He finds a class size effect only for senior teachers.

approach (student, teacher and school fixed effects) concerns remain about non-random assignment of students to teachers and about modeling assumptions. For instance, Rothstein (2010) finds evidence for dynamic sorting which biases the estimated teacher effects[2]. In addition, Wiswall (2013) shows that restrictive modeling assumptions generate the common finding that teacher experience beyond the initial years in the profession is not important.

A second approach in the literature focuses on classes where students are, or appear to be, randomly assigned. For instance, Clotfelter et al. (2006) use a subsample of schools that feature relatively balanced distributions of students across classrooms, based on observable characteristics. Several studies exploit data from the STAR-experiment in which students and teachers were randomly assigned to small and large classes (Krueger 1999; Dee 2004; Nye Konstantopoulos and Hedges 2004; Chetty et al. 2012). Teacher experience is found to be the only observed teacher characteristic that matters which is consistent with studies using value-added models (Staiger & Rockoff 2010). However, the gains from teacher experience are also found after the initial years in the profession.

This paper examines the effect of teacher quality on student achievement using a novel identification strategy that exploits data on twin pairs who entered the same school but were allocated to different classrooms in an exogenous way. Our strategy combines the main components of the two previous approaches from the literature. First, we exploit the exogenous assignment of individual twins to classrooms within the same schools. Second, the longitudinal character of the data enables us to take prior achievements of students into account like in the common value-added models. The assignment of twins to different classrooms can be viewed as a natural experiment that exposes very similar individuals to different schooling conditions. This quasi-experiment allows us to investigate the causal effect of classroom quality on student outcomes using observational data. The variation in classroom conditions to which the twins are exposed can be considered as exogenous if the assignment of twins to different classes is as good as random. This assumption seems quite plausible within the institutional context of this study; Dutch primary education. In many Dutch schools twins are assigned to different classes due to an (informal) policy rule that dictates that twins are not allowed to attend the same class. At school entry schools and

[2] Rothstein (2010) evaluates the most common value-added specifications used for the assessment of teacher performance. He finds that the assumptions underlying common value-added specifications are substantially incorrect and the estimates of teacher effects based on these models cannot be interpreted as causal. See also Guarino et al. (2013) on the validity of value-added measures of teacher performance.

parents do not yet have information about the ability or behavior of twins and the ability or behavior of their class mates. Moreover, because in early childhood twins are more similar than different, it seems not likely that small differences between twins will affect the way they are assigned to different classes. In our empirical analysis we have tested this assumption and did not find evidence of the non randomness of the assignment.

Our research design exploits the exogenous assignment of twins to different classrooms. The treatment in this design is classroom quality, which is a multi-dimensional concept that might include factors such as peer quality, class size and teacher quality. In our empirical analysis we especially focus on the effects of observed teacher characteristics on student outcomes because, in applying our design, teachers seem the most obvious factor differing across classes. Typically, Dutch schools equalize classroom facilities and class composition across classes. In schools with many students and few teachers we expect little variation within twin pairs in class size and peer composition, and much within twin pair variation in teacher quality. Therefore, we will exploit the assignment of twins to different classes in particular to estimate the effects of teacher characteristics on student outcomes. For doing so, we use longitudinal data of a large representative sample of students from Dutch primary education. We have identified twins from the population based sample by using information on their date of birth, family name and school.

Our paper makes two important contributions to the current economic literature. First, we contribute to the literature on teacher quality by introducing an empirical strategy that has never been used to estimate the impact of teacher quality on student outcomes. Previous studies have relied on value-added modeling (e.g. Rivkin et al. 2005, Rockhoff, 2004) or exploited classes where students are, or appear to be, randomly assigned (e.g. Krueger, 1999, Chetty, 2005, Clotfelter et al. 2006). Second, we contribute to the economic literature that exploits data on twins by combining twins with exogenous treatment assignment. Twin differencing has been applied on various topics such as the returns to schooling or the intergenerational effects of schooling (see for instance Ashenfelter and Krueger 1994, Behrman and Rosenzweig, 2002). These studies are based on the assumption that variation within twin pairs is exogenous but it remains unclear why twins differ[3]. As far as we know, there are no twin studies that arguably exploit exogenous variation in twin differences.

[3] Li et al. (2010) exploit a twin design in which parents are forced to send one of their twins to the countryside. The treatment assignment might not be random as it is based on a parental decision.

In line with earlier studies on teacher effects we find that teacher experience is the only observed teacher characteristic that matters for student performance ((Hanushek 2011; Staiger & Rockoff 2010, Chetty et al. 2011). Twins that are assigned to classes with more experienced teachers perform better in reading and math. On average one extra year of experience raise test scores by approximately one percent of a standard deviation. The effects of teacher experience are most pronounced in kindergarten and early grades. Our findings are remarkably consistent with the results found by Krueger (1999) and Chetty et al. (2005) using data from the STAR-experiment in which students and teachers were randomly assigned to classes. They also report linear effects of teacher experience and find that the effects of teacher experience reduce after kindergarten. The linear effects of teacher experience contrast 'the consensus in the literature' that only initial teacher experience matters (Staiger & Rockoff 2010), but are in line with recent findings by Wiswall (2013) and Harris & Sas (2011) who also find gains from later experience.

The estimated effects of teacher experience should be interpreted carefully because they do not necessarily reflect the effect of training on the job but could be driven by other intrinsic characteristics of more experienced teachers[4]. For instance, the findings might be driven by sample attrition if less effective teachers are more likely to leave the profession. In the empirical analysis we test for various mechanisms that might explain why experienced teachers are better teachers. We do not find evidence consistent with mechanisms that stress the importance of changes over time such as changes in the quality of teacher education or changes in outside opportunities in the labor market. However, we also find an effect of teacher experience for career stages with less labor market mobility. These estimates suggest positive returns to on the job training of teachers as it is less likely that these estimates will be biased because of selection into or out of the profession. This finding is consistent with recent studies that also have found positive return to teacher experience for later career stages (Wiswall, 2013, Harris & Sas, 2011).

Regardless the story, our estimates show that experienced teachers matter, especially for reading. This implies that policies to maintain experienced teachers in the classroom appear beneficial, especially for younger students. In addition, the matching of experienced teachers with students might be important if schools or parents value the gains in educational achievements of specific students more than the gains of other students.

[4] Rockoff (2004), Kane et al. (2006), Chetty et al. (2011) and Harris & Sass (2011) have previously noted this issue.

Our paper is organized as follows. In Section 2 we describe our empirical strategy and relate this to previous approaches from the literature. Section 3 describes the data. The main estimation results are presented in Section 4. Section 5 provides additional tests about the key assumption of our empirical strategy about the random assignment of twins to classrooms. Section 6 explores the possible mechanisms underlying the estimated effects of teacher experience. In Section 7 we conclude.

2. Empirical strategy

The basic framework in the economic literature that studies the effects of teachers models student achievement as a function of family, peer, community, teacher and school inputs and student ability (Hanushek & Rivkin, 2006). Student achievement at any point in time is seen as a cumulative result of the entire history of all inputs and the individual's initial endowment (e.g. innate ability). A common approach for modeling this so-called educational production function is to assume that the cumulative achievement function is additively separable and linear (e.g. Boardman and Murnane 1979; Todd and Wolpin 2003; Harris & Sas 2011). Estimating the effect of teachers is complicated because in any actual application we will generally not be able to control for all relevant school, family or student characteristics. If some omitted variables are correlated with the relevant teacher characteristics, then the estimated parameters will be biased. The major threat to identification is the non-random sorting of students among schools and classrooms.

Researchers have used two types of empirical strategies for identifying the effects of teacher characteristics. The first, and most common, approach is based on value-added modeling. The second approach exploits situations where students are, or appear to be, randomly assigned. The most common approach in the literature is to include measures of prior achievement and estimate value-added models. These models focus on gains in student achievement or the rate of learning over specific time periods. Recent studies exploit the availability of multiple years of information for teachers for estimating teacher fixed effects which are linked with teacher characteristics (Hanushek 1992; Hanushek et al. 2005; Rockoff 2004; Aaronson, Barrow and Sander 2003; Rivkin et al. 2005). A second approach in the literature identifies teacher effects by exploiting situations in which students are, or appear to be, randomly assigned to classrooms and teachers (e.g. Clotfelter et al. 2006; Chetty et al. 2012). It might be expected that unobserved factors will not bias the estimates due to the random assignment of students.

In this paper we exploit the assignment of twins to different classrooms for estimating the causal effect of class inputs on student achievement. This assignment can be viewed as a natural experiment that exposes very similar individuals to different class room conditions. Our approach is most related to ‘the random assignment studies’ but by including previous test scores as controls we are also able to estimate value-added models. For explaining our empirical strategy we use as a starting point the basic economic framework which relates class inputs to measures of educational performance. Consider the following specification of a traditional educational production function:

$$\text{(E.1)} \quad Y_{ijcs} = \alpha_1 X_i + \alpha_2 Z_j + \alpha_3 SQ_s + \alpha_4 CQ_c + x_i + z_j + sq_s + cq_c + \varepsilon_{ijcs}$$

where indices i, j, c and s stand for pupil i born in family j in classroom c at school s. Observable educational output Y represents the test scores on reading or math. Observable inputs of the educational production function contain individual attributes X, family characteristics Z, and various measures of classroom quality CQ and school quality SQ. For ease of notation, we keep the specification very general in the sense that any class attribute could be represented by CQ. It could be teacher’s experience for example, but also class size or composition. The error term consists of analogous unobservable inputs of the educational production function x, z, sq, cq and an idiosyncratic effect ε which is uncorrelated with all these observable and unobservable determinants. In this paper we are interested in estimating α_4 which represents the structural effect of an observable class input on pupil test scores.

If we use conventional cross section data to estimate equation E.1, least squares estimation might not yield an unbiased estimate of α_4 for multiple reasons. First, there might be a non random assignment of pupils to classes, i.e. $Cov(CQ, x) \neq 0$. This means for example that worse performing pupils are more often assigned to classes with better peers or better teachers. Second, there might be parental influences on the child’s school and classroom, i.e. $Cov(CQ, z) \neq 0$. For instance, higher educated parents may select better schools or class rooms because they may be more involved with their children than lower educated parents. Third, better schools may attract better teachers (and better pupils), i.e. $Cov(CQ, sq) \neq 0$, because differences in quality of school management may cause different schools to attract different teachers. Fourth, schools may use multiple inputs to manipulate classroom environment, i.e. $Cov(CQ, cq) \neq 0$), because schools may decide to compensate classes with high fractions of low ability pupils by reductions in class size and/or extra aide. To sum up, estimating an equation like (E.1) with non experimental data is likely to induce bias due to unobservable characteristics of pupils, parents, and teachers and schools (i.e. school management).

Our empirical strategy for identifying α_4 exploits differences within pairs of twins. If we suppress subscripts and take twin differences, our empirical model can be rewritten as

(E.2) $\Delta Y = \alpha_1 \Delta X + \alpha_2 \Delta Z + \alpha_3 \Delta SQ + \alpha_4 \Delta CQ + \Delta x + \Delta z + \Delta sq + \Delta cq + \Delta \varepsilon$

Identification of α_4 now rests on four assumptions:

(A.1) twins share family background, i.e. $\Delta Z = 0$ and $\Delta z = 0$

(A.2) twins enter the same school, i.e. $\Delta SQ = 0$ and $\Delta sq = 0$

(A.3) twins are exogenously allocated to different classrooms, i.e. $\Delta CQ \neq 0$ but $Cov(\Delta CQ, \Delta x) = 0$

(A.4) observable and unobservable class attributes are unrelated, i.e. $Cov(\Delta CQ, \Delta cq) = 0$

Assumptions (A.1) and (A.2) are satisfied by design. Assumption (A.3) seems also plausible because of the assignment procedures in Dutch education. We will discuss the plausibility of this assumption shortly. However, assumption (A.4) seems not plausible. If we assume that the assumptions (A.1), (A.2) and (A.3) hold we can simplify the empirical model to:

(E.3) $\Delta Y = \alpha_4 \Delta CQ + \Delta cq + \Delta \varepsilon$

Twin fixed-effect estimation will therefore give us the following estimator:

$$\alpha_4^{FE} = \frac{Cov(\Delta Y, \Delta CQ)}{Var(\Delta CQ)} = \alpha_4 + \frac{Cov(\Delta CQ, \Delta cq)}{Var(\Delta CQ)}$$

Hence, this twin fixed effect estimator not only captures the impact of any observable classroom characteristic but also the impact of every unobservable characteristic that is correlated with it. This can be interpreted as the broad impact of classroom quality.

As in any quasi-experimental design there are deviations from the ideal experimental design in which a specific treatment is randomly assigned to an experimental group of students. The first, and crucial, issue in this design is whether the assignment of twins is truly exogenous (assumption (A.3)). The second issue in our design is about the treatment variable. What is the treatment considering the fact that classroom quality is multi-dimensional? Both issues should be considered within the institutional context of Dutch primary education. In Dutch primary education parents and pupils are free to choose their school. All schools receive funding from the government based on the number and socioeconomic background of the pupils. Primary school consists of eight grades of which grade 1 and grade 2 are equivalent to kindergarten. Children are allowed to enroll in primary education on their fourth birth day which induces a rolling admission in grade 1. Compulsory education starts at the age of 5. Most schools mix first and second graders. After grade 2 children are reassigned to different classes. The composition of these classes remains quite stable until the end of primary education in grade 8 in which pupils take a nationwide test.

Is the assignment of twins truly exogenous?

The key identifying assumption of our approach is the random assignment of twins to different classrooms. Many schools in Dutch primary education employ a policy of separating twins in different classes.[5] This separation already takes place when the twins enroll in grade 1. It should be noted that this type of policy is only used for twins; there are no such rules for the assignment of singletons. The rolling admission of pupils in grade 1 implies that class size and classroom composition are volatile and only partly observed by parents. After finishing the school year in which pupils enrolled they spend two complete school years in grade 1 and 2. Hence, in total most students spend more than two years in grade 1 and 2. During this whole ‘kindergarten stage’ pupils keep the same teacher(s) and are not reassigned to other classes. This school policy is likely to induce random assignment of individual twins at school entry since in kindergarten there is no (or very little) information on class quality, such as the quality of the class mates, that parents can use to determine which type of class suits their twins best. In addition, twins in early childhood are more similar than different and it seems unlikely that small differences between twins affect the way in which they are assigned to different classes. From this we expect that the assignment of newly entering twins, which creates the classes of grade 1 and 2, can be considered as exogenous. In contrast, the assignment of individual students to classes, even to classes of

[5] The Dutch Society for Parents of Multiples advises parents to follow their own opinion, but believes that separation stimulates the individualization of the twins (Geluk & Hol, 2001). Most schools explicitly put their twin policy on their website. Many schools assign twins to different classes based on the belief that putting them in the same class will harm them, although recent research suggest that this is not the case [see for example Webbink et al. (2007)].

grade 1 and 2, might not be random because of the various sorting mechanisms that have been pointed out in the literature (Clotfelder et al. 2006; Feng 2009). As a consequence, we consider the assignment of classroom components, especially teacher quality, to individual twins in grade 1 and 2 as random, but the assignment of these classroom components to singletons in these classes might not be random. We will therefore run twin-fixed effect regressions and interpret the corresponding estimate as the broad impact of classroom quality.

A reassignment of students in Dutch education takes place in the transition from grade 2 to grade 3. At this stage there will be more information available about the twins and their class mates, although it might be expected that the twins are very similar. This implies that we are not fully sure whether the assignment is still exogenous for third and higher graders. To address this concern we will estimate value-added models which include previous test scores of our twins. We will compare the results of the basic random assignment specifications with the results of the value-added specifications. The potential bias due to non random assignment of twins is expected to be small if these results are very similar. In addition, we will perform several tests on the empirical importance of endogenous classroom assignment.

What is the treatment?

In our design we compare the performance of twins that are assigned to different classes. The treatment in this design is classroom quality, which is a multi-dimensional concept. The literature on class quality typically focuses on the impact of peers, teacher quality and class size. In the Dutch institutional context we expect that teacher quality will be the most important component of this treatment. Due to the rolling admission class size and peer composition are volatile in grade 1 and 2, whereas teacher quality is fixed. In addition, Dutch schools typically equalize facilities, peer composition and class size across classes. In schools with many students and few teachers we expect much within twin pair variation in teacher quality and little within twin pair variation in class size and classroom composition. Therefore, we expect that the assignment of twins to different classes in grade 1 and 2 can mainly be interpreted as an assignment to different teachers. For grade 3 to 8 we also expect that teacher quality is the main component of the treatment. With many students and few teachers in schools we expect only little within twin pair variation in class size and class composition. Most variation in classroom quality will come from differences in teacher quality. A similar interpretation has been used in the literature that investigates the effect of school and teacher quality through the estimation of classroom fixed effects on achievement gains. The resulting classroom differences in average achievement gain have been interpreted as reflecting teacher quality, since the teacher is the most obvious factor differing across classrooms (Hanushek, 1992). In this study we will therefore investigate teacher quality effects in more detail and focus on components such as experience, gender and fulltime or part-time employment of teachers.

3. Data

The data come from the longitudinal biannual PRIMA project (Driessen et al. (2004)). The PRIMA project consists of a panel of approximately 60,000 pupils in 600 schools. The participation in the project is voluntary. The main sample, which includes approximately 420 schools, is called the reference sample, which is representative for the Dutch student population in primary education. An additional sample includes 180 schools for the over-sampling of pupils with a lower socioeconomic background (the low SES sample). After each wave of the project some schools drop out and some new schools are included.[6] This means that the panel structure only holds for a subsample of the dataset.[7] We use all six waves of the PRIMA survey including data on pupils, parents, teachers and schools from the school years 1994-95, 1996-97, 1998-99, 2000-01, 2002-03 and 2004-05. Within each school, pupils in grades 2, 4, 6 and 8 (average age: 6, 8, 10, 12 years) are tested in reading and math. Information on teachers is also collected but the main focus of the project is to follow pupils (and not teachers) during primary education.

Our identification strategy is based on differences within pairs of twins. The PRIMA-data does not contain direct information on twins versus singletons. We have identified twins by matching on family name, date of birth, school and year of the survey. If there are two pupils with exactly the same values on these matching variables they are considered to be twins. In the total sample of the PRIMA-data we have identified 623 records of twin pairs that were assigned to different classrooms and for which (reading) test scores and teacher data are available. Because of the longitudinal character of the data some twin pairs will be observed more than once; the number of unique twin pairs in our data is 495. The total sample of 623 twin pair observations consists of 448 same sex pairs (219 pairs of boys, 229

[6] There are no significant differences between the schools that drop out and the schools that remain in the project (Roeleveld and Vierke, 2003).

[7] Other reasons are pupils that change schools or pupils not being present at the time tests are taken.

pairs of girls), 173 opposite sex pairs and 2 pairs with unknown gender. More twin pairs have been identified in earlier grades than in later grades; 235 pairs in grade 2, 175 pairs in grade 4, 132 pairs in grade 6 and 81 pairs in grade 8. If one of the twins is retained or accelerated we will not observe a pair because of the sampling structure of the PRIMA-project (only grade 2, 4, 6 and 8). This might explain the lower number of pairs in later grades.

Our main dependent variables are scores on tests for languages and arithmetic which were developed as part of the PRIMA-project. The language test for children in second grade, which is equivalent to infant school, measures the understanding of words and concepts. The arithmetic test for these children focuses on the sorting of objects. These tests are taken in class. The test for children in grades 4, 6 and 8 all come from a system for following pupil achievements in primary education developed by the CITO group. The aim of these tests is to observe to what extent students master various elements of the curriculum. The tests for the same grade levels are identical each year. This ensures that the comparison of achievement levels over time is possible. The scores are also comparable between grades. The scales of the raw scores for language and arithmetic have no clear meaning. We have standardized these test scores by wave and grade with the mean and standard error of the reference sample.

The main explanatory variables in this paper are a set of class input factors. First, we have information on teacher characteristics: experience (measured in years) and gender. In addition, the data provide information whether the class is taught by one fulltime teacher or by two part-time teachers. In case of two teachers we use the experience of the teacher that was present at the time of the survey. Class size is reported by the teacher but is also available from the PRIMA-register. Moreover, we use two measures of the composition of the class: fraction of girls and fraction of native Dutch pupils. The latter is a proxy for the socioeconomic status of twin's class mates.

Table I shows the descriptives for the samples of twins in grade 2, in grade 4-8, and for the total sample of twins. In addition, the last columns of Table I show the descriptives for the total sample of the PRIMA-project. The number of observations differs between grades because of the longitudinal structure of the data and the sampling strategy of the PRIMA-project. Previous test scores are only available for grade 4 and higher, and for pupils (schools) who participated in previous waves of the project. The bottom panel of Table I shows that we have previous test scores for one third of the sample of twins in grade 4 and higher. In addition, information about the peers is not available for classrooms with multiple grades, which is often the case in grade 2. Teacher characteristics have been measured for most classrooms. This implies that the estimation sample for models that exploit the longitudinal character of the data or models that include peer characteristics will be smaller than models that only use cross-sectional data and focus on teacher characteristics. The means of the test scores in our twin sample are negative which means that twins perform below the average of the student population of the reference sample. The means of the test scores for the total PRIMA-sample are also negative as we used the reference sample for the standardization and the total sample also includes the low-SES sample.

Table II shows the variation of the class inputs within pairs of twins, which is the variation that is crucial for our identification strategy. The variance in teacher characteristics within twin pairs is much larger than the variation in class size. At the 95th percentile the difference in teacher experience within a twin pair is 26 years, for class size this difference is 4 pupils.

4. Main estimation results

In this section we show the main results of our empirical analysis. We start by presenting the results for students in grade 2. For these students we are most confident about the assumption that twins are exogenously assigned to classrooms because at school entry there is hardly any information about the student and his/her classmates that might lead to selection into classrooms. Table III shows the twin-fixed effect estimates of teacher characteristics on student performance in reading and math using different specifications[8]. We investigate the effect of teacher experience, gender of the teacher and having one fulltime or two part-time teachers in the classroom. The first two columns of Table III show the estimated effects of having a more or less experienced teacher in the classroom in models without controls. The estimates show that one additional year of teacher experience in the classroom increases performance in reading or math with 1.4 or 1.5 percent of a standard deviation. The other teacher characteristics are included in the models in column (3) and (4). These models also control for split level classrooms, class size and gender of the student. The observed teacher characteristics, gender of teacher and the number of teachers in the class room, do not affect student performance. We also observe that the inclusion of the new variables does not change

[8] To improve the power of our analysis we have used a sample for which missing values for several covariates have been imputed. In case of a missing value on a covariate we have assumed that there is no difference within a pair of twins. For reading we have imputed 31 observations, for math we have imputed 27 observations. Teacher experience has not been imputed. The main estimation results do not change when we use the smaller samples without imputations. Estimations results are available upon request.

the estimated effects of teacher experience in the classroom. Column (5) and (6) additionally controls for differences in classroom composition, in particular the proportion of girls and the proportion of native students in the classroom. Again, we observe that including these controls does not change the effect of teacher experience. Hence, the estimates for students in grade 2 suggest that teacher experience in the classroom is an important determinant of student performance and teacher experience seems to be the only observed teacher characteristic that matters. These findings are consistent with previous results from the literature on teacher quality (see Section 1 and 2).

The returns to teacher experience beyond the initial years in the profession

The recent literature is not consistent about the returns to experience during various stages of the teaching profession. Many studies have found that experience only matters in the initial years in the profession (e.g. Rivkin et al. 2005; Rockoff 2004) and there seemed to be a consensus about this finding in the literature (Staiger & Rockoff 2010; Wiswall 2013). However, recent studies also find gains from teacher experience in later years of the career (Harris & Sas 2011; Wiswall 2013). Moreover, Wiswall (2013) shows that restrictive modeling assumption in previous studies have generated the common finding that experience only matters in the first years of the profession. Using an experience variable with a limited number of categories within a panel setup that includes only a few years of information on teachers seriously reduces the variance that can be exploited in the estimation. He finds high returns to later experience using an unrestricted experience model for student performance in math. For student performance in reading he finds low returns to later experience. Previous studies based on the data from the STAR-project report linear effects of teacher experience (Krueger 1999, Chetty et al. 2012). Our estimates also suggest a linear effect of experience on student achievements (see also Figure 1A and Figure 1B). We have also experimented with higher order terms of experience but we did not find significant results for these specifications[9].

Results for the full sample of students from grade 2 to 8

In the next step of our analysis we use the full sample of twins from grade 2, 4, 6 and 8. As noted in Section 2, for the full sample of twins we are less confident about the assumption

[9] We did not use a specification with a limited number of experience categories because of the restrictive nature of this approach as pointed out by Wiswall (2013).

that students have been randomly assigned to classrooms because students in Dutch primary education are re-assigned to classes after grade 2. It might be expected that teachers, parents and students will have more information after grade 2 about themselves and other students which might lead to non- random selection into classes. This selection might bias the estimated effects for the full sample. To address this issue we not only estimate the ‘random assignment specifications’ from Table III but also estimate value-added specifications in which we control for previous test scores. By combining a value-added specification with our experimental design we aim to mitigate non-random selection into classes. We further investigate the empirical importance of endogenous classroom assignment after grade 2 in Section 5.

The estimation results for the full sample of twins are shown in Table IV. Column (1) to (8) shows the estimated effects using the random assignment specifications that are also used in Table III. Columns (1) to (4) use the full sample of twins, in columns (5) to (8) we only use twins for which previous test scores are available. Columns (9) to (12) show the estimation results for the value-added specifications. A disadvantage of including previous test scores is that we typically loose the first observation (pupils in grade 2) because the previous test score is not available. However, since the random assignment of pupils in grade 2 ensures that there are no initial differences within the twin pairs we can replace the previous test score with a constant, in order to keep the first year of the data[10]. For the full sample of twins we find that one additional year of teacher experience in the classroom increases performance in reading with 0.9 to 1.4 % of a standard deviation and performance in math with 0.6 to 0.9 %. A comparison of the results from the ‘random assignment specification’ with the results from ‘the value-added specification’ can be considered as an important test for the non-random assignment of twins because generally previous test scores are important control variables. We observe in Table IV that the estimates based on the ‘random assignment specification’ are very similar to the estimates based on the ‘value-added specification’ which suggests that the bias from non-random assignment will be limited. Again, including higher order terms of experience does not change the estimated effects. The estimated effect of teacher experience is robust to the inclusion of the other teacher characteristics and other controls. For the other teacher characteristics we find no systematic effects on student performance. Hence, the estimates we have found for the sample of twins in grade 2 are consistent with the estimates based on the whole sample of twins.

[10] Krueger (1999) and Mueller (2013) use a similar approach.

Teacher experience and grade level

Previous studies have reported different returns to experience by grade level. Krueger (1999) and Chetty et al. (2011) find higher effect of teacher experience for kindergarten than for higher grades. We have also investigated whether teacher experience is more important for younger pupils. Table V shows the estimation results for teacher experience by grade level; Panel A shows the results for the random assignment specification based on the total sample, Panel B shows the results for the value-added specification based on the sample for which we observe previous test scores. The estimates show that the effect of teacher experience depends on the grade of the pupil. Teacher experience in class matters most in grade 2 (kindergarten): an additional year of experience in class raises test scores by approximately 1.5 % of a standard deviation. Teacher experience becomes less important in higher grades. In grade 8 we don't find an effect of teacher experience on performance. Hence, teacher experience raises test scores especially for younger pupils. This finding is very similar to the results based on data from the STAR-project. Our finding might also be explained by the fact that students in grade 2 have the same teacher for two years whereas in higher grades this is less likely.

5. Sensitivity tests about non random classroom assignment after grade 2

To further investigate the empirical importance of endogenous classroom assignment after grade 2 we perform several tests. First, we regress test scores in second grade on classroom characteristics in fourth grade. If assignment is random, we should not observe a relationship between test scores and classroom characteristics. Table VI shows the results and provides no evidence for a nonrandom assignment of twins after grade 2. For the models in column (8) that include all variables simultaneously we find two statistically significant effects but the F-test shows that we cannot reject the hypothesis that there is no classroom-effect. Better performing twins in grade 2 are not assigned more often to other type of classes in grade 4 than their (worse performing) twin brothers or sisters on observed class inputs.

As our second test we regress test scores obtained in second, fourth, sixth and eighth grade on class room characteristics that were measured in the second grade. If assignment is random at school entry, but assignment in later years is not, reduced form estimates (assuming that classroom characteristics are correlated across grades) are informative about class input effects. Panel A of Table VII shows the results related to this test. The first two columns show the effect of teacher experience for the full sample; these results have already been shown in Table IV. The next two columns show the results for the sample of twins for which we have test scores in grade 2 and in at least one higher grade. The estimates effects for this sample imply that one additional year of teacher experience in class increases performance with 1.3 percent of a standard deviation. The last two columns of Panel A show the estimated effect of teacher experience in grade 2 on the test scores in grade 2 or in higher grades. Hence, these columns show the reduced form estimates. If assignment to classrooms in higher grades is not random but assignment to grade 2 is random the reduced form estimates are informative about the effect of experience in class. The reduced form estimates indicate that one year of teacher experience in class increases test scores with approximately 1 percent of a standard deviation. These estimates lie in the same ball park as the previous estimates. In sum, these tests don't provide evidence for a non random re-assignment after grade 2, which might threaten the identification of the estimates in Table IV.

Further robustness analyses are shown in Panel B of table VII. Students in split level classrooms have now been excluded from the estimation sample. For these students we only have class composition information from the students in the same grade but not from the students in the other grade. Although this strongly reduces the sample size the estimated effects of teacher experience remain quite similar to the previous estimates.

6. Does the effect of teacher experience reflect on the job training?

The main finding from the previous sections is that students that are allocated to classes with more experienced teachers perform better than students that are allocated to classes with less experienced teachers. We have observed that the effects of teacher experience are not affected by other classroom factors, which suggest that these results come from teachers and their qualities. In addition, the teacher experience estimates do not change when other teacher characteristics are included, suggesting an important role for teacher experience and everything else that is correlated with it. Although we are quite confident that this finding is not driven by non-random selection of students to more experienced teachers, the interpretation of this result is not immediately clear because the randomization that we exploit is about classrooms and not about teachers with different qualities. Hence, it is not clear whether the experience effect reflects the effect of training on the job or whether the experience effect is the result of unobserved teacher qualities that are correlated with obtaining more experience in teaching. As noted in previous studies (Chetty et al. 2011;

Rockoff 2004; Kane et al. 2006; Harris & Sass 2011), the estimated effect of teacher experience might be driven by different mechanisms. First, the estimated effects might be the result of training on the job, which we label as the causal effect of teacher experience. Second, the results might be driven by positive or negative selection of teachers in the education sector. For instance, teachers that are more (less) skilled or motivated and committed might be more (less) likely to stay in the education profession. In fact, Wiswall (2013) finds evidence for negative selection of teachers in American public schools. Third, more experienced teachers might have had a better teaching education and therefore might be more skilled (Corcoran, Evans, and Schwab 2004, Hoxby and Leigh 2004, Bacolod 2007). If the quality of teacher education has deteriorated over time teacher experience will be correlated with the quality of teacher education. Fourth, selection into the teaching profession might have changed over time due to a changing labor market. The increasing demand for higher educated workers will probably have increased the number of alternatives for working in the teaching profession. Over the years the teaching profession might have attracted weaker teachers.

The effect of on the job training (the causal effect of teacher experience) can be isolated from unobserved quality differences across teachers by using multiple years of information of teachers (Rivkin et al. 2005; Wiswall 2013). Unfortunately, we cannot apply this approach because the panel character of our data only relates to students and not to teachers. However, we can empirically explore the plausibility of the various mechanisms by looking at changes in the estimated effects over time or changes in the effect of experience over the teaching career. We start by looking at changes in the estimated effects over time. These changes are informative about the last two mechanisms which both state that older cohorts of teachers had more quality than younger cohorts. If we assume that the causal effect of experience on student performance does not change over time we expect that the estimated effect of teacher experience will decline over time because of older cohorts of teachers leaving the teaching profession. We can put these mechanisms to the test by exploiting the panel character of our data. We have constructed three time periods from the six waves of our panel and included interaction variables between these time periods and teacher experience in our main model[11]. The estimated effects of the interaction variables show whether the experience effect has changed over time. Table VIII shows the estimation results. We do not observe that the estimated effects decline over time as might be expected from the last two

[11] Using all six periods separately would strongly reduce the number of observations for several periods.

mechanisms. Hence, the evidence is not consistent with these explanations of the teacher experience effect.

We further attempt to purge the effect of unobserved quality differences from the effect of on the job training by looking at the experience effect of teachers that differ in mobility. If the effect of teacher experience is driven by unobserved teacher quality that is correlated with obtaining experience, then we expect that the bias from unobserved teacher quality will be smaller for teachers that are less likely to leave the profession. Because teacher mobility is highest in the initial years in the profession we use a model specification that can pick up differences in the effect of teacher experience during the career. First, we have included an interaction variable between the minimum teacher experience of both teachers of the twin pair and the difference in teacher experience. Second, we have constructed three categories for minimum teacher experiences and interacted these variables with the difference in teacher experience. The interaction variables measure whether the estimated effect of teacher experience changes over the career. We expect that the estimates for later career stages will be less likely to be biased by positive or negative selection. Hence, these estimates should be a better approximation of the effect of training on the job than estimates for early career stages. Table IX shows the estimated effects for these specifications. The main result for both subjects is that the estimated effects for these interaction variables are all very small and statistically insignificant. This suggests that the teaching experience effect is quite constant during the teaching career. Hence, we also find returns to teaching experience for career stages in which we expect less bias due to positive or negative selection. The estimates of teaching experience for later career stages are expected to be a better approximation of the effect of on the job training in education. This finding is consistent with recent studies that also find evidence for the importance of teacher experience beyond the initial years in the career (Wiswall, 2013, Harris and Sas, 2011).

7. Conclusion

In this paper we have examined the causal link that runs from classroom quality to student achievement by applying a new identification strategy. This strategy is based on an exogenous assignment of twins to different classrooms. Teacher quality is expected to be the main factor differing across classes. The main findings of this paper are related to teacher quality. We find that teacher experience is the only observed teacher characteristic that matters for student performance. This finding is consistent with previous studies on teacher effects (Hanushek 2010; Staiger & Rockoff 2010, Chetty et al. 2011). Twins that are assigned to classes with more experienced teachers perform better and the effects are most pronounced in kindergarten and early grades. Krueger (1999) and Chetty et al. (2011) report similar results from their analysis using data from the STAR-experiment on class size reduction. Our estimates also suggest a linear effect of experience on student achievements. Until recently there was a consensus in the literature that teacher experience only matters in the initial years in the career (e.g. Rivkin et al. 2005; Rockoff 2004; Staiger & Rockoff 2010). However, recent studies also find gains from teacher experience in later years of the career (Harris & Sass 2011; Wiswall 2013). Moreover, previous studies based on the data from the STAR-project report linear effects of teacher experience (Krueger 1999, Chetty et al. 2012). Hence, our estimates corroborate the recent findings about later returns to experience.

From our analysis we learn that teacher experience is very important but it is not clear how we should interpret this finding because our estimates only show that students do better in classes with more experienced teachers. It remains unclear whether this effect is caused by training on the job or reflects the effects of unobserved teacher quality correlated with attaining more experience in education. We have explored the plausibility of various mechanisms that might explain the robust finding that students in classrooms with more experienced teachers perform better. We do not find evidence consistent with mechanisms that stress the importance of changes over time such as changes in the quality of teacher education or changes in outside opportunities in the labor market. However, we find that teacher experience also matters for career stages with less labor market mobility. As it is less likely that these estimates will be biased by selection into or out of the teaching profession this suggests positive returns to on the job training of teachers. This finding is consistent with recent studies that also find positive returns to teacher experience for later career stages (Wiswall, 2013, Harris & Sas, 2011).

The main finding of our paper is that experienced teachers are very important for student performance. Although we are unable to isolate the effect of on the job training from the effect of unobserved teacher quality this finding has important policy implications. More focused policies that maintain experienced teachers in the classroom appear beneficial, especially for younger students.

References

Aaronson, D., Barrow, L., Sander, W., 2007, Teachers and student achievement in Chicago public high schools, *Journal of Labor Economics*, 25, 95-135.

Ashenfelter, O. and Krueger, A. (1994). Estimating the returns to schooling using a new sample of twins. *American Economic Review*, 84: 1157–1173.

Bacolod, M. P. (2007). Do Alternative Opportunities Matter? The Role of Female Labor Markets in the Decline of Teacher Quality, *Review of Economics and Statistics*, 89: 737–751.

Boardman, A.E., Murnane R., 1979, Using panel data to improve estimates of the determinants of educational achievement, *Sociology of Education*, 52, 113-121.

Behrman, J.R. and M.R. Rosenzweig, 2002, Does Increasing Women's Schooling Raise the Schooling of the Next Generation?, *American Economic Review*, 92 (1): 323-334

Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Whitmore Schanzenbach, D., Yagan, D. (2011). How does you kindergarten classroom affect your earnings? Evidence from project STAR. *Quarterly Journal of Economics*, 126 (4): 1593-1660.

Clotfelter, C.T., H. F. Ladd, J.L. Vigdor, 2006. Teacher-Student Matching and the Assessment of Teacher Effectiveness, *Journal of Human Resources*, 41(4): 778-820.

Corcoran, S. P., W. N. Evans, R. M. Schwab (2004), Changing Labor-market Opportunities for Women and the Quality of Teachers, 1957–2000, *American Economic Review*, 94 (2004), 230–235.

Dee, T. S., 2004, Teachers, Race and Student Achievement in a Randomized Experiment. *The Review of Economics and Statistics*, 86 (1): 195-210

Driessen, G., Van Langen, A., & Vierke, H. (2004). *Basisrapportage PRIMA-cohortonderzoek, Vijfde meting 2002-2003 (Report on PRIMA-longitudinal research project, Survey 2002-2003).* Nijmegen.

Feng, Li., 2009, Opportunity wages, classroom characteristics, and teacher mobility, *Southern Economic Journal*, 75, 1165-1190.

Geluk, A., & Hol, J., 2001, Samen of apart? Meerlingkinderen naar school, peuterspeelzaal of kinderopvang [Together or Apart? Multiples to school, playground or daycare; Brochure]. Bergen, the Netherlands: NVOM (Dutch Society for Parents of Multiples).

Guarino, C., Reckase, M.D., Wooldridge, J.M., 2012. Can Value-Added Measures of Teacher Performance Be Trusted?, IZA Discussion Papers 6602, Institute for the Study of Labor (IZA).

Hanushek, E.A., 1971, Teacher characteristics and gains in student achievement: Estimation using micro data, *American Economic Review*, 60 (2), 280-288.

Hanushek, E.A., 1992, (1992): “The Trade-Off Between Child Quantity and Quality,” *Journal of Political Economy*, 100, 84–117.

Hanushek, E.A., 2011, The Economic Value of Higher Teacher Quality, *Economics of Education Review*, 30(2): 466-479.

Hanushek, E.A., Rivkin, S.G., 2006, Teacher quality. In: Hanushek, E., Welch, F. (Eds.), Handbook of Economics of Education, vol 2. Elsevier.

Hanushek, E.A., Kain J.F., O’Brien, D.M., Rivkin, S.G., 2005, The market for teacher quality, NBER Working Paper 11154.

Harris, D.N., Sass, T.R., 2011, Teacher training, teacher quality and student achievement, *Journal of Public Economics*, 95, 798-812.

Hoxby, C. M., and A. Leigh. (2004), Pulled Away or Pushed Out? Explaining the Decline of Teacher Aptitude in the United States, *American Economic Review*, 94: 236–240.

Kane, T., Rockoff J.E., Staiger D.O., 2006, What does certification tell us about teacher effectiveness? Evidence from New York City, Working Paper 12155.

Krueger, A.B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics*, 114 (2): 497-532.

Li, H., M. Rosenzweig, J. Zhang. (2010). Altruism, Favoritism, and Guilt in the Allocation of Family Resources: Sophie's Choice in Mao's Mass Send-Down Movement, *Journal of Political Economy*, 118(1): 1-38.

Mueller, S., 2013, Teacher experience and the class size effect – Experimental evidence, *Journal of Public Economics*, 98, 44-52.

Nye, B., S. Konstantopoulos, L.V. Hedges, 2004, How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26 (3): 237-257

Rivkin, S.G., Hanushek, E.A., and Kain, J.F. (2005). Teachers, schools and academic achievement. *Econometrica*,Vol. 73, No. 2: 417-458.

Rockoff, J.E. (2004). The impact of individual teachers on student achievement: evidence from panel data. *The American Economic Review*, Vol. 94, No. 3, Papers and proceedings of the one hundred sixteenth annual meeting of the American Economic Association, pp. 247-252.

Roeleveld, J. and H. Vierke, 2003, *Uitval en Instroom bij de derde meting van het PRIMA-cohortonderzoek*, Amsterdam/Nijmegen: SCO-Kohnstamm Instituut / ITS.

Rothstein, J. (2010), Teacher Quality in Educational Production: Tracking, Decay and Student Achievement, *Quaterly Journal of Economics*, 125(1): 129-174

Staiger, D.O., Rockoff, J.E., 2010, Searching for effective teachers with imperfect information, *Journal of Economic Perspectives*, 24 (3): 97-117

Todd, P.E., Wolpin, K.I., 2003, On the specification and estimation of the production function for cognitive achievement, *Economic Journal*, 113 (485), F3-33.

Webbink, D., Hay, D., Visscher P.M. (2007). Does sharing the same class in school improve cognitive abilities of twins? *Twin Research and Human Genetics*, Vol. 10, No. 4: 573-580

Wiswall, M., 2013, The dynamics of teacher quality, *Journal of Public Economics*, 100, 61-78.

Figures

Figure 1A. Student performance in reading by teachers experience within pairs of twins

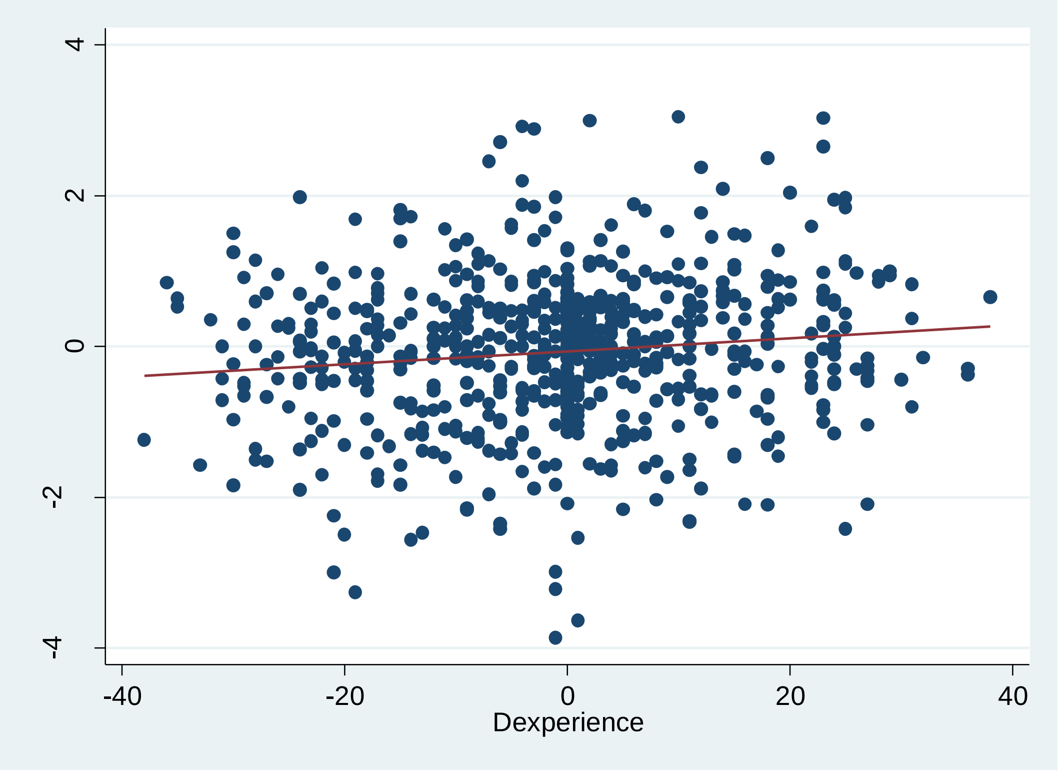
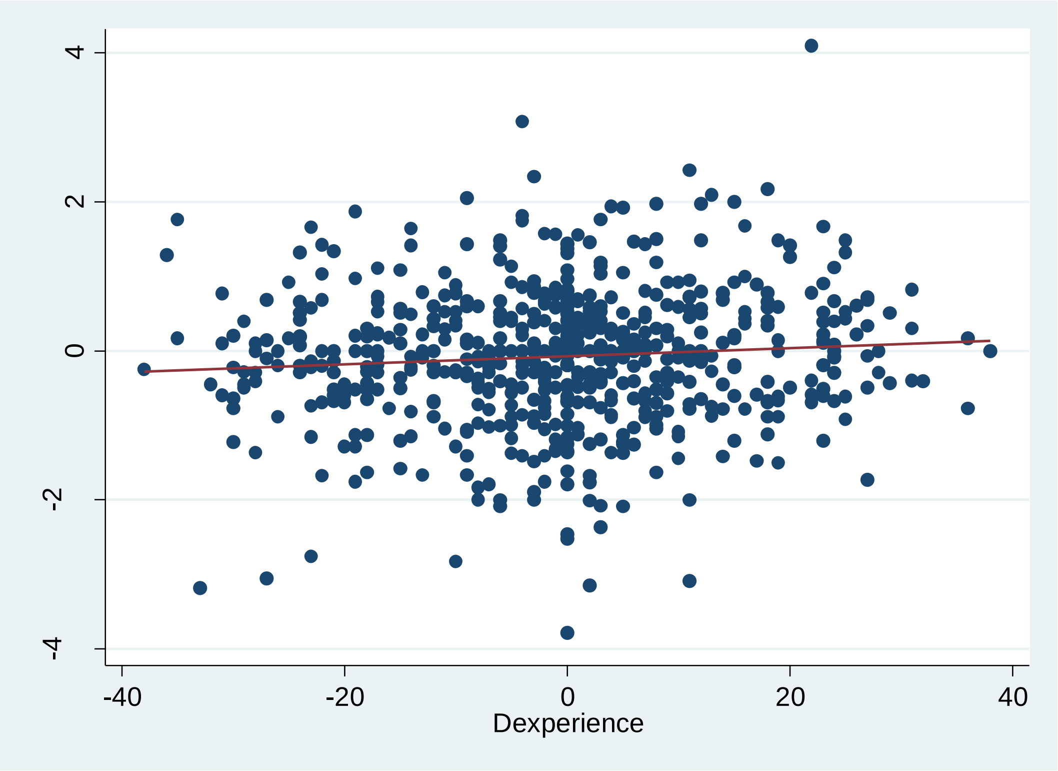


Figure 1B. Student performance in math by teachers experience within pairs of twins



Tables

Table I: Descriptive statistics of estimation samples (for reading)

	Twin samples						PRIMA-sample	
	Grade 2		Grades 4, 6 and 8		Total		Total	
Variables	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**
twin characteristics								
reading score	-0,39	1,07	-0,16	1,07	-0,25	1,08	-0,15	1,02
math score	-0,33	0,94	-0,11	1,05	-0,20	1,02	-0,12	1,01
girl	0,50	0,50	0,49	0,50	0,49	0,50	0,50	0,50
teacher & class room								
experience in education (years)	15,36	10,43	16,60	11,30	16,13	10,99	18,21	10,58
female	0,98	0,15	0,63	0,48	0,76	0,43	0,66	0,47
multiple classroom teachers	0,52	0,50	0,40	0,49	0,45	0,50	0,50	0,50
split level classroom	0,83	0,38	0,21	0,41	0,44	0,50	0,47	0,50
class size	24,07	4,50	23,54	4,95	23,74	4,79	24,39	5,78
observations	470		776		1246		330350	
previous test scores twins & class composition (for subsample in grades 4-8, excluding mixed grades)								
reading score T-2	-	-	-0,28	0,93	-	-	-0,12	1,01
math score T-2	-	-	-0,13	0,98	-	-	-0,05	0,98
girl share in class (%)	-	-	50,49	10,02	-	-	50,06	11,72
native share in class (%)	-	-	60,63	35,92	-	-	64,99	34,17
observations			276				97557	

Table II: Distribution of twin differences in class room characteristics in total estimation sample for reading (# Twin pairs=623)

	Percentiles						
	5th	25th	50th	75th	95th	mean	sd
Δ reading	-1,78	-0,64	-0,05	0,56	1,56	-0,07	1,02
Δ math	-1,64	-0,60	-0,08	0,50	1,47	-0,08	0,93
Δ girl	-1,00	0,00	0,00	0,00	1,00	0,02	0,53
Δ experience (in years)	-26,00	-10,00	0,00	8,00	24,00	-0,78	14,57
Δ female (teacher)	-1,00	0,00	0,00	0,00	1,00	0,01	0,46
Δ multiple class room teachers	-1,00	0,00	0,00	0,00	1,00	0,03	0,67
Δ split level classrooms	-1,00	0,00	0,00	0,00	1,00	0,02	0,41
Δ class size	-4	-1	0	1	4	-0,06	2,81

Table III: Twin-fixed effect estimates of teacher quality effects on student test scores in grade 2

	(1)	(2)	(3)	(4)	(5)	(6)
Independent variables:	reading	math	reading	math	reading	math
teacher experience	0.014***	0.015***	0.014***	0.015***	0.014***	0.014***
	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
female teacher			-0.174	0.290	-0.106	0.271
			(0.314)	(0.291)	(0.342)	(0.281)
two teachers			0.065	0.025	0.065	0.020
			(0.098)	(0.076)	(0.099)	(0.076)
split level classroom			-0.228	-0.103	-0.231	-0.151
			(0.183)	(0.180)	(0.192)	(0.189)
girl (in opposite sex twin pair)			0.162	0.169	0.156	0.160
			(0.120)	(0.115)	(0.119)	(0.112)
class size			0.005	0.033	0.004	0.033
			(0.032)	(0.027)	(0.031)	(0.028)
% girls in class					-0.006	-0.001
					(0.004)	(0.003)
% natives in class					0.004	-0.004
					(0.003)	(0.004)
# twin pairs	235	236	235	236	235	236
R-squared	0.035	0.052	0.049	0.074	0.061	0.079

Notes: Each column shows the results of an OLS-regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 Covariates have been imputed, see footnote 8.

Table IV: Twin-fixed effect estimates of teacher quality effects on student test scores in grades 2 to 8

	Random assignment specification								Value-added specification			
	Total sample				Sample with previous test scores				Sample with previous test scores			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Independent variables:	reading	math	reading	math	reading	math	reading	math	reading	math	reading	math
teacher experience	0.009***	0.006**	0.011***	0.006**	0.013***	0.008***	0.014***	0.008***	0.013***	0.009***	0.014***	0.009***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
female teacher			0.164*	0.132*			0.075	0.122			0.072	0.201**
			(0.090)	(0.078)			(0.107)	(0.095)			(0.103)	(0.090)
two teachers			-0.034	-0.001			-0.002	0.070			-0.015	0.061
			(0.062)	(0.056)			(0.072)	(0.060)			(0.071)	(0.059)
class room that mixes grades			0.142	0.116			0.125	0.147			0.098	0.115
			(0.110)	(0.116)			(0.114)	(0.132)			(0.112)	(0.125)
girl (in opposite sex twin pair)			0.163*	-0.121			0.229**	-0.099			0.208**	-0.097
			(0.092)	(0.089)			(0.097)	(0.086)			(0.093)	(0.080)
class size			-0.005	0.008			-0.020	0.018			-0.021	0.017
			(0.016)	(0.016)			(0.020)	(0.020)			(0.019)	(0.019)
% girls in class			-0.001	0.002			-0.002	0.004*			-0.002	0.004
			(0.003)	(0.002)			(0.003)	(0.002)			(0.003)	(0.002)
% natives in class			0.006*	-0.003			0.005*	-0.003			0.005	-0.002
			(0.003)	(0.003)			(0.003)	(0.003)			(0.003)	(0.003)
previous test score (t-2)									0.267***	0.340***	0.246***	0.348***
									(0.070)	(0.061)	(0.071)	(0.059)
# twin pairs	623	611	623	611	451	447	451	447	451	447	451	447
R-squared	0.016	0.008	0.039	0.022	0.034	0.019	0.063	0.041	0.065	0.082	0.089	0.105
Controls:												
Teacher/Class characteristics	no	no	yes	yes	no	no	yes	yes	no	no	yes	yes
Previous test scores	no	no	no	no	no	no	no	no	yes	yes	yes	yes

Notes: Each column shows the results of an OLS-regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Covariates have been imputed, see footnote 8.

Table V: Twin fixed effect estimates of teacher experience on student test scores by grade

	grade 2		grade 4		grade 6		grade 8	
	reading	math	reading	math	reading	math	reading	math
Panel A: random assignment specification, total sample								
teacher experience	0.014***	0.014***	0.009**	0.005	0.011**	0.001	0.004	-0.004
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.009)	(0.007)
# twin pairs	235	236	175	173	132	128	81	74
R-squared	0.061	0.079	0.074	0.090	0.077	0.075	0.112	0.120
Panel B: value-added specification, sample with previous test scores								
teacher experience	0.014***	0.014***	0.013**	0.010	0.013**	0.002	0.007	-0.007
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.012)	(0.009)
# twin pairs	235	236	82	86	81	77	53	48
R-squared	0.061	0.079	0.180	0.275	0.260	0.262	0.235	0.568

Notes: Each column within a panel shows the results of an OLS-regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Panel A includes all covariates as in column (3) and (4) in table IV. Panel B also includes previous test scores.

Table VI: Twin-fixed effect estimates of grade 2 reading scores on grade 4 classroom characteristics

Classroom characteristics in fourth grade:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
teacher experience	-0.009							-0.011*
	(0.006)							(0.006)
female teacher		-0.240						-0.174
		(0.213)						(0.238)
two teachers			0.015					0.053
			(0.155)					(0.182)
split level classroom				0.248				0.275
				(0.176)				(0.191)
class size					-0.015			-0.017
					(0.040)			(0.044)
% girls in class						-0.006		-0.007
						(0.004)		(0.006)
% natives in class							0.003	-0.003
							(0.008)	(0.010)
p-value F-test: no class room effect								0.307
# twin pairs	87	87	88	88	87	112	108	79
R-squared	0.018	0.016	0.000	0.012	0.002	0.009	0.001	0.052

Notes: Each column shows the results of an OLS-regression.. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table VI continued: Twin-fixed effect estimates of grade 2 math scores on grade 4 classroom characteristics

Classroom characteristics in fourth grade:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
teacher experience	-0.008							-0.009
	(0.006)							(0.006)
female teacher		-0.193						-0.108
		(0.131)						(0.119)
two teachers			0.027					0.035
			(0.139)					(0.153)
class room that mixes grades				-0.165				-0.001
				(0.172)				(0.165)
class size					-0.014			-0.012
					(0.037)			(0.040)
% girls in class						0.005		-0.000
						(0.006)		(0.007)
% natives in class							-0.008	-0.018**
							(0.006)	(0.008)
p-value F-test: no class room effect								0.177
# twin pairs	92	93	93	93	92	117	113	85
R-squared	0.019	0.012	0.000	0.007	0.002	0.006	0.014	0.072

Notes: Each column shows the results of an OLS-regression.. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table VII: Various twin-fixed effect estimates of teacher experience on student test scores for grades 2, 4, 6, 8

PANEL A:	Total sample		Sample for which teacher experience in grade 2 is available		Reduced form	
	(1)		(2)		(3)	
	reading	math	reading	math	reading	math
teacher experience	0.011***	0.006**	0.013***	0.013***	0.011**	0.009**
	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)	(0.005)
# twin pairs	623	611	301	299	301	299
R-squared	0.039	0.022	0.038	0.079	0.030	0.060
PANEL B: sample without split level classrooms	Total sample: Random assignment specification		Longitudinal sample: Random assignment specification		Longitudinal sample: Value added specification	
	(4)		(5)		(6)	
	reading	math	reading	math	reading	math
teacher experience	0.013***	0.005	0.017***	0.005	0.017***	0.006
	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)
# twin pairs	299	289	186	184	186	184
R-squared	0.055	0.072	0.070	0.103	0.128	0.191

Notes: Each column shows the results of an OLS-regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In columns (2) and (3) the sample is used for which teacher experience from grade 2 is available. In columns (4)-(6) the sample is used that excludes split level classrooms. In column (4) and (5) the full specification is used as in columns (3) and (4) in table IV. In column (6) also previous test scores have been included.

Table VIII: Twin-fixed effect estimates of teacher experience on student test scores by PRIMA survey year (1994-2004)

	reading		math	
Independent variables	(1)	(2)	(3)	(4)
teacher experience	0.012	0.008	0.009	0.004
	(0.008)	(0.012)	(0.007)	(0.008)
teacher experience*year	0.000		0.000	
	(0.001)		(0.001)	
teacher experience*dummy=1 if survey years are 1998 or 2000		0.003		0.005
		(0.012)		(0.009)
teacher experience*dummy=1 if survey years are 2002 or 2004		0.003		0.001
		(0.012)		(0.009)
# twin pairs	623	623	611	611
R-squared	0.043	0.045	0.031	0.030

Notes: Each column shows the results of an OLS-regression using the random assignment specification. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.In columns (1) and (3), the omitted category of teacher experience is survey year 1994 (i.e. year=0 if survey year=1994, year=2 if survey year=1996 and so on). In columns (2) and (4) the omitted category of teacher experience is the dummy that equals 1 for survey years 1994 or 1996. In all columns the full specification is used as in columns (3) and (4) of table IV.

Table IX: Twin-fixed effect estimates of teacher experience on student test scores by minimum experience of teachers of both twins

	reading		math	
Independent variables:	(1)	(2)	(3)	(4)
teacher experience	0.012***	0.011***	0.007**	0.005*
	(0.004)	(0.004)	(0.003)	(0.003)
teacher experience*minimum teacher experience of both teachers	-0.000		-0.000	
	(0.000)		(0.000)	
teacher experience*dummy=1 if minimum teacher experience between 5 and 16 years		-0.000		0.003
		(0.007)		(0.006)
teacher experience*dummy=1 if minimum teacher experience 17 years or more		0.000		-0.007
		(0.010)		(0.009)
# twin pairs	623	623	611	611
R-squared	0.046	0.046	0.023	0.026

Notes: Each column shows the results of an OLS-regression using the random assignment specification. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In columns (2) and (4) the omitted category is the dummy that equals 1 if the minimum years of experience of both teachers lie between 0 and 4. In all columns the full specification is used as in columns (3) and (4) of table IV.

Publisher:

CPB Netherlands Bureau for Economic Policy Analysis
P.O. Box 80510 | 2508 GM The Hague
T (070) 3383 380

November 2014 | ISBN 978-90-5833-666-8