The Effects of Ability Tracking of Future Primary School Teachers on Student Performance *

Johan Coenen
Chris Van Klaveren†

Abstract

Because of the Dutch tracking system, primary school teachers in the Netherlands can have a vocational or a higher secondary background. Policymakers and school principles worry that teachers with vocational backgrounds are less capable to teach math and reading. This study therefore examines the effects of ability tracking of future primary school teachers on the students' math and reading performance

We exploit data of 91 schools for all primary school children in grades 3,4 and 5 and identify the tracking effect by exploiting unique information how teachers are assigned to classes based on their teaching abilities.

The estimation results for math (reading) indicate that test scores are .2 (.12) of a standard deviation lower if their teacher had a vocational background. The results for reading are, however, not significant at the 10% confidence level and the tracking estimates appear to be less stable and precise.

JEL-codes: I20 · I21 · I29

Keywords: Teacher · Student achievement · Ability tracking · Test Scores

*We would like to thank Wim Groot, Henriette Maassen van den Brink, Iryna Rud and the participants of the Workshop on Empirical Research in Economics of Education 2013 in Reus for their valuable comments.

†All authors are affiliated with TIER, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands. The corresponding author is E-mail: johan.coenen@maastrichtuniversity.nl. Chris Van Klaveren, TIER, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands. E-mail: cp.vanklaveren@maastrichtuniversity.nl.

1 Introduction

Primary school teachers in the Netherlands have to finish primary school teacher training college (hereafter referred to as teacher college) successfully before they can teach at a primary school. To be admitted at teachers college a higher secondary or intermediate vocational education degree should be attained and as a consequence teachers in primary education may have a vocational or a higher secondary educational background.

Dutch policy makers and school principles frequently argue that primary school teachers with a vocational background are insufficiently capable to instruct children in math and reading. They mention that the math and reading skills of first year students at teacher college with a vocational background are systematically lower compared to first year students with higher secondary education level and, according to them, these math and reading deficiencies are not properly addressed within the teachers college program. That students with a vocational background have lower math and reading skills in the first year is the consequence of the Dutch tracking system in which children are tracked into higher secondary and intermediate vocational education tracks at the age of twelve based on their reading and math ability.[1]

There has been, however, no empirical evidence that children perform worse in math and reading if instruction was received by a teacher with a vocational background. Even though these teachers may indeed have math and reading deficiencies, it is unclear if these deficiencies are properly addressed in the teachers college program. Moreover, it may be that teachers with a vocational background have acquired other relevant teaching skills and experiences, which may compensate for their math and reading deficiencies. This study therefore examines if primary school children perform worse in math and comprehensive reading if they received instruction by a teacher with a vocational background. Hereafter we refer to teachers with a vocational background as vocational teachers (or VT) and to teachers with a higher secondary background as higher secondary teachers (or HT).

Many European countries have tracking systems similar to the Netherlands and also in these countries the possible consequences of educational tracking is heavily debated among parents and policymakers (Hanushek and Woessmann, 2006). Some countries track children into different education levels at a relatively young age. In Germany and Austria, for instance, children are tracked at age 10, in the Czech Republic, Slovakia and Hungary at age 11, and in Belgium at age 12. Many other countries (such as the US, the UK, France, Japan and Switzerland) also track children into different education levels, but rather late at the age of 15 or 16 (see OECD (2004) for an extensive list of tracking ages per country). Hanushek and Woessmann (2006) explain that ability tracking is on the one hand considered as positive because more homogenous classrooms permit a curriculum that is more focused and appropriately spaced. On the other hand they mention that there are many arguments against ability tracking. One of these arguments is that ability grouping eliminates the potential positive peer-effects in heterogeneous classrooms. Using several large European data

[1] Tracking occurs conditionally on the achieved test scores on a national and standardized test (CITO) which is taken at the end of grade 6 and conditionally on the advice given by the school.

sources they find that educational tracking increases the variance of the test score distribution variances and, while less clear, ability tracking seemed to reduce mean performance. The findings of Hanushek and Woessmann (2006) are consistent with the arguments of Dutch policymakers and school principals in the sense that first year students with a vocational background at teacher college will likely have lower math and reading skills than first year students with a higher secondary background.

This study contributes to the discussion on educational tracking because it recognizes that persons can be tracked differently into education levels even though they have attained similar highest education levels. Studies generally focus on how certain outcome variables, such as wages and test scores, are influenced by the highest attained education level, but if persons with similar highest education levels have acquired different skills because of ability tracking then highest education level may not be the variable of interest. This study particularly focuses on ability tracking effects for future primary school teachers, but many other examples can be given. Recently, Agan (2014) showed for the US that the financial college returns for persons who have attained a similar college degree depends on the specific educational path that led to the college degree.

To examine the effects of ability tracking of future primary school teachers on the students' math and reading performance we use data of 91 primary schools which allow us to follow the educational careers of all 3600 children who were in grades 3,4 and 5 from 2010 to 2012. These data contain detailed information on teachers, students and their parents that comes partly from the school registration system and partly from student and parent questionnaires. Moreover, these data contain the test scores on national standardized tests for math and comprehensive reading that were taken at the beginning and/or at the end of each school year.

A unique feature of the data used in this study is that it contains information on how schools assign teachers to classes based on their quality. Studies that measure the relationship between teacher characteristics and student performance (gains) generally estimate multi-level models without taking the possible selective assignment of teachers to classes into account (see, among others, (Wayne and Youngs, 2003; Clotfelter et al., 2006, 2010)), which may impose a bias on the empirical findings. If, for example, vocational teachers are structurally assigned to better performing classes then it may appear that children of vocational teachers perform better but that this effect is purely caused by the selective assignment and not because vocational teachers perform better.

The empirical findings suggest that math (reading) test scores of children of vocational teachers are lower by .2 (.12) of a standard deviation. The tracking effect for reading is, however, not significant at the 10% confidence level and the robustness analysis suggest that the estimates on reading are less stable and precise. The estimated tracking effect for math appears to be robust. A robustness analysis developed by Altonji et al. (2005) shows that the estimated tracking effect for math is not driven by selection on unobserved factors. Moreover, the tracking effects are not driven by experience and gender differences between teachers.

An interesting result is that teachers are frequently selectively assigned to classes and this selective assignment has a major impact on the empirical findings. At first sight, vocational teachers appear to perform equally well as higher secondary teachers but once we control for the fact that teachers are selectively assigned to classes we find that vocational teachers perform less well. It implies that vocational teachers are on average assigned to better classes while higher secondary teachers are on average assigned to weaker classes.

This study proceeds as follows. Section 2 shortly describes the empirical literature on teacher ability on student performance. Section 3 describes the Dutch education system. Section 4 describes the data and descriptive statistics. Section 5 discusses the identification strategy and presents the empirical findings. Section 6 shows how the estimation results are potentially influenced by selection on unobservables. Finally, Section 7 concludes.

2 Literature review

Many empirical studies examine how teacher characteristics are associated with student learning outcomes (see, among others, Rivkin et al. (2005); Clotfelter et al. (2006, 2010); Metzler and Woessmann (2012)). Because teacher achievement test scores are the best proxy for teacher ability, we summarize in this section the empirical findings of studies that focus specifically on the association between student test scores and teacher achievement test scores. Following Wayne and Youngs (2003) we distinguish between three teacher test-score categories: (1) licensure examination scores, (2) verbal skill tests and, (3) other test score measures. Table 1, presents an overview of the empirical findings and we note that studies published before 2001 were also mentioned in Wayne and Youngs (2003).

Summers and Wolfe; Summers and Wolfe (1975;1977) were the first who studied if students from the Philadelphia school district performed better if their teachers scored better on the teacher licensure examination. Their findings indicate that primary school children had lower grade point averages if their teacher scored higher on the licensure examinations. For junior high and senior high school samples no significant findings were found. Ferguson (1991) used teacher reading test scores of the statewide teacher testing in Texas in 1986 and found that students scored better reading test scores, but not math test scores, if teachers scored higher reading test scores. In a later study, Ferguson (1998) confirmed his earlier result and found that differences in student test scores gains between elementary and secondary education were associated with the reading test scores differences between primary and secondary school teachers. Clotfelter et al. (2006) use administrative data on North Carolina public schools and find a small but significant effect of teachers licensure test scores on the math achievement of primary school children, while no such effect is found for reading. More specifically, they find that a one standard deviation increase in teacher licensure test score increases the math test scores by 1 to 2 percent of a standard deviation. Using the same data, Clotfelter et al. (2010) examine the effects of teachers licensure test scores on student performance for more subjects. For math and biology they find small but significant effect sizes of respectively .05 and .02. For English a small but significant negative effect is found, and no significant effects are found for economics and civics.

Table 1: Overview of the empirical findings on teacher test scores and student test scores

Study	School type	Teacher test	Subject: GPA	Math	Reading	Other
Clotfelter et al. (2006)	PE	Licensure		+	n.s.	
Clotfelter et al. (2010)	SE	Licensure		+		+ / - / n.s.
Ferguson (1991, 1998)	PE / SE	Licensure		n.s. / n.s.	+ / +	
Summers and Wolfe (1975, 1977)	PE / SE	Licensure	- / n.s.			
Ehrenberg and Brewer (1995)	PE / SE	Verbal skills	+			
Hanushek (1992)	PE	Verbal skills			+ / n.s.	
Murnane and Phillips (1981)	PE	Verbal skills			n.s.	
Ferguson and Ladd (1996)	PE	Other		n.s.	+	
Kukla-Acevedo (2009)	PE	Other		+		
Metzler and Woessmann (2012)	PE	Other		+	n.s.	
Rowan et al. (1997)	SE	Other		+		

Note: PE stands for Primary education; SE stands for Secondary education and can include both middle school and high school in case of US studies.

Only three studies examine the relationship between student achievement and teacher test scores achieved on verbal skill tests. Murnane and Phillips (1981) find that vocabulary test scores of students are not associated with the test scores that their teachers achieved on verbal skill tests. Hanushek (1992) uses the same data but focus on vocabulary and reading test scores of students. Similar to Murnane and Phillips (1981) he finds no association between students' vocabulary test scores and the achieved test scores on verbal skill tests by their teachers, but he does find a positive relationship between the reading test scores of students and the achieved test scores on verbal skill tests by their teachers. Ehrenberg and Brewer (1995) find that teacher test scores achieved on a short verbal facility test explains some between school variation in student achievement gains, even when controlling for teacher experience and graduate education.

Finally, four studies examine the association between student achievement and other teacher test score measures. Ferguson and Ladd (1996) use Alabama personnel records, which include ACT college entrance examination test scores of teachers. This examination tests the skills of teachers on English, reading, math, social studies and natural sciences and for the analysis a grade point average is calculated for each teacher. The empirical results indicate that there is a positive association between the grade point averages achieved by teachers and reading achievement of primary school children. No significant relationship is found between the grade point averages achieved of the teachers and the math test scores achieved by the students.

Rowan et al. (1997) are using the US National Educational Longitudinal Study of 1988 and find that the math performance of primary school children was better when their teachers scored better on a single high school level mathematics test item. Metzler and Woessmann (2012) use Peruvian primary school data to examine if teacher subject knowledge influence the math and reading test scores of their students positively. Their findings suggest that a one standard deviation in subject-specific teacher achievement increases student achievement by .09 of a standard deviation in math. For reading the effects are mostly not significantly different from zero. Kukla-Acevedo (2009) find that teacher grade point averages influences the math achievement of students positively.

The empirical literature generally suggests positive or non-significant effects of teacher test scores on student performance. Most studies make use of panel data to estimate the teacher ability effect. It may be, however, that teachers are selectively assigned to classes based on their ability, and this may impose a biased on the estimated effects. The empirical findings of Table 1 represent lower bound estimates if better teachers with better test scores are structurally assigned to weaker classes. On the other hand, if better teachers are selectively assigned to better classes, then Table 1 represent upper bound estimates. In this study we also make use of panel-data but by using information on whether teachers are assigned to classes based on their ability we can obtain a better estimate of the effects of ability tracking of future teachers on student performance.

3 The Dutch education system

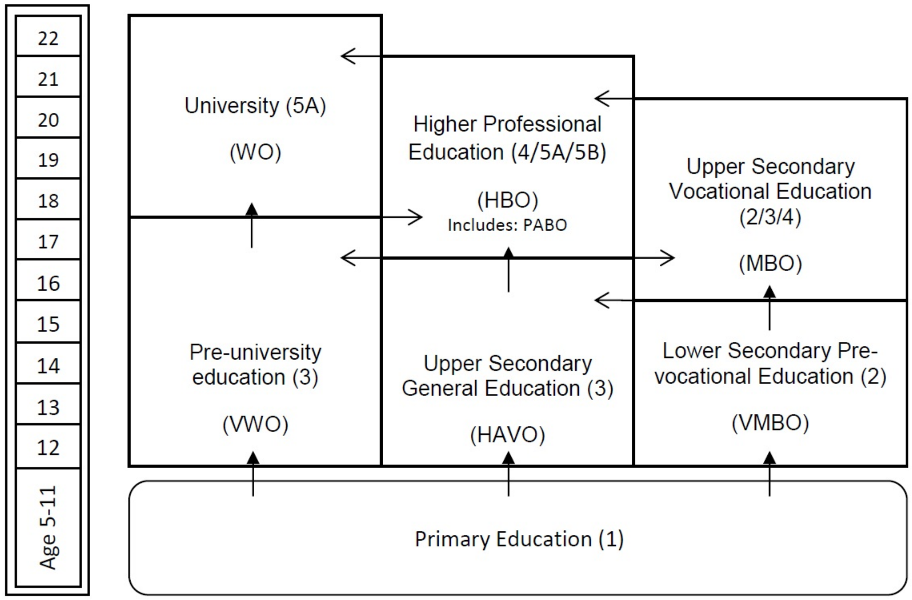
The Dutch education system is graphically illustrated in Figure 1. Children in the Netherlands go to primary education when they become four years old. In the final grade of primary education (grade six) children make a nationwide standardized test. Based upon this test, and an advice of the primary school, children are tracked into three education levels.

The lowest education level children can be tracked to is referred to as lower secondary pre-vocational education (VMBO, four years). This track prepares children for upper secondary education vocational education (MBO, four years). Within upper secondary education vocational education there are four tracks that differ in difficulty. A level one track is the easiest track, while a level four track is the most difficult track. The middle education level is referred to as upper secondary general education (HAVO, five years), which prepares children for higher professional education (HBO, four years). The higher professional education level includes teachers college. The highest education level children can be tracked to is referred to as pre-university education (VWO, six years), which prepares children for an academic university study (four or five years).

The relevant information for this study is that persons can enroll in teachers college, first of all, if they successfully finished an upper secondary general education or pre-university education level (path one). Secondly, students can enroll in teachers college only if they have successfully finished the most difficult track within upper secondary vocational education (path two). It follows that the two educational paths that allow students to start with teachers college are not equal in length. The duration of the first path is 9 years, while the duration of the second path is 12 years and thus considerably longer.[2]

[2]We note that there are two alternative educational paths that enable students to start with teachers college. The first alternative path takes 10 years and students who follow this path start with pre-vocational education, continue with upper secondary general education and then start and successfully finish teachers

Figure 1. *The Dutch Education System*



Note: ISCED level in parentheses.

4 Data and descriptive statistics

This study uses non-representative longitudinal panel data for the Netherlands in which the math and reading performance of all (3600) primary school children of 91 schools. These children were enrolled in grades 3, 4 and 5 in 2010 and together with their teachers they tracked for three school years from February 2010 to June 2012. The data contain detailed information on teachers, parents and the school. The information on children, their parents and the schools comes partly from the school registration system and partly from student, parent and principal questionnaires. Teachers filled in teacher questionnaires each year such that information was obtained on certain background characteristics, such as education level, experience, the education path that led to the highest education level and gender. Some teacher characteristics (birth date and gender) could also be retrieved from the school registration system.

Primary schools in the Netherlands take standardized math and reading tests and the achieved test scores on these tests are used in this study to assess children's math and reading achievement. Normally the math tests are taken biannually, in February and June, while the reading tests are taken yearly in February. During the tracking period schools also took an

college. The second alternative path also takes 10 years and students who follow this path start with pre-university education and then start and successfully finish teachers college. Teachers however rarely follow these alternative paths and for our teacher sample it holds that no teachers followed these alternative paths.

extra standardized reading and math tests in either June or September. It follows that we observe more math test scores than reading test scores for each student. More specifically we observe 8120 math test scores and 5803 reading scores for 3603 children who are taught by 202 teachers of whom 21 percent (42 teachers) are vocational teachers.

Table 2 compares teacher- and class-related background characteristics between vocational and higher secondary teachers. The table shows that primary school teachers are predominantly feminine, which is common in the Netherlands, but vocational teachers are as often male teachers than higher secondary teachers. Vocational teachers have less teaching experience than higher secondary teachers and this is a logical consequence of the Dutch education system in which the educational path to teachers college through vocational education is three years longer than the educational path to teachers college through higher secondary education (see also Section 3). However, vocational teachers have, on average, six years less teaching experience than higher secondary teachers, while the Dutch educational system can only account for a difference of three experience years. On the one hand, this may indicate that vocational teachers gain less teaching experience than higher secondary teachers because they perform less well in primary education. On the other hand, Table 2 shows that higher secondary teachers are on average 5.3 years older than vocational teachers, and observed differences in experience are therefore resulting from the tracking system and additional age difference. It follows that it will be difficult to disentangle the tracking and experience effects in the empirical analysis and this problem is addressed in Section 6.2.

The class related descriptive statistics suggest that vocational and higher secondary teachers teach similar sized classes and that the proportion of boys and Dutch students in these classes is not statistically and significantly different. The positive effect of parents' schooling on their children's schooling has been empirically shown (see Holmlund et al., 2011) and children of higher educated are likely to perform better on math and reading tests. We divided the mother's education level into the four categories: unknown, low (no or only primary school), middle (higher secondary or intermediate vocational education) and high (higher vocational or university education). The descriptive statistics show that there are no significant differences in the education levels of mothers between vocational and higher secondary teachers. If the education level of the mothers is a good predictor for the learning potential of their children then this is an indication that vocational and higher secondary teachers teach children with the same learning potential. In this case, achieved test score differences of children between vocational teachers and higher secondary teachers cannot be attributed to differences in learning potential.

The school-level data is based on register data from the Dutch Ministry of Education and in Table 3 we compare school characteristics for the 91 schools in our sample between vocational and higher secondary teachers. None of reported mean differences are significant at the 95 percent confidence level and vocational and higher secondary teachers therefore appear to teach at very similar primary schools in terms of student population and denomination.

The register data contains information on *all* Dutch primary schools such that it is possible to characterize how the 91 primary schools in our sample differ from the average primary school in the Netherlands. Primary schools in the Netherlands receive subsidy based

Table 2: Comparing Teacher and Class Characteristics between Vocational and Higher Secondary Teachers

	HT (N=160)		VT (N=42)			
	Mean	Std. dev.	Mean	Std. dev.	Difference	Std.Err.
Teacher:						
Male teacher	0.163	(0.370)	0.190	(0.397)	-0.028	(0.068)
School experience (years)	9.213	(9.397)	5.897	(5.721)	3.316**	(1.271)
Total experience (years)	14.383	(11.538)	8.556	(9.027)	5.826***	(1.839)
Age in years	38.807	(11.793)	33.838	(11.490)	4.969**	(2.136)
Class:						
Boys	0.495	(0.151)	0.520	(0.162)	-0.025	(0.023)
Class size	16.948	(6.621)	17.644	(7.031)	-0.696	(1.014)
Dutch	0.615	(0.332)	0.552	(0.380)	-0.063	(0.059)
Education level mother						
Unknown	0.059	(0.189)	0.036	(0.130)	0.023	(0.021)
Low	0.110	(0.150)	0.128	(0.174)	-0.018	(0.025)
Middle	0.687	(0.223)	0.693	(0.196)	-0.006	(0.030)
High	0.144	(0.146)	0.142	(0.135)	0.001	(0.020)

Note: */ **/*** means statistically significant at the 10/5/1 percent level.

on 'weight' indicators which are related to the education level of the parents. Children receive a .3 weight if both parents finished a lower secondary education level as highest education level (i.e. the lowest pre-vocational track) and a 1.2 weight if one parent finished a lower secondary education level and the other parent has a lower education level. The average primary school in the Netherlands has 9 percent children with a .3 weight and 6 percent children with a 1.2 weight. These percentages for the 91 schools in our sample are 16.8 percent for children with a .3 weight and 15.8 percent for children with a 1.2 weight, and it follows that primary school children in our sample have parents with relatively low education levels. Schools located in disadvantaged areas also receive additional government funding. Areas are labeled as disadvantaged if the neighborhood has an above average combination of unemployment, early school leaving, criminality and low income. 42 percent of the schools in our sample are located in a disadvantaged area which is relatively high given that 15 percent of all Dutch primary schools are located in a disadvantaged area.

This study uses unique information on how schools assign teachers to classes based on their quality. This information is given by the school principles and they indicated if better teachers were assigned to better classes, to weaker classes or if there was no selective assignment of teachers based on their ability. Empirical studies generally do not take into account the potential selective assignment of teachers to classes into account (see, among others, (Wayne and Youngs, 2003; Clotfelter et al., 2006, 2010)), which may impose a bias on the empirical findings. In our study, it could for example be that higher secondary teachers are

Table 3: Comparing School Characteristics between Vocational and Higher Secondary Teachers

	HT (N=160)		VT (N=42)			
	Mean	Std. dev.	Mean	Std. dev.	Difference	Std.Err.
Number of children	146.738	(97.740)	146.048	(88.604)	0.690	(15.704)
Disadvantaged children (weight .3)	0.168	(0.181)	0.163	(0.183)	0.005	(0.032)
Disadvantaged children (weight 1.2)	0.158	(0.091)	0.137	(0.068)	0.021*	(0.013)
Disadvantaged area	0.419	(0.495)	0.429	(0.501)	-0.010	(0.087)
Boys	0.502	(0.062)	0.511	(0.057)	-0.009	(0.010)

Note: */ **/*** means statistically significant at the 10/5/1 percent level.

more frequently assigned to weaker classes because schools focus relatively more on primary school children who perform less well. If learning gains are smaller for children who perform less well, the estimation results may, for example, indicate that higher secondary perform less well than vocational teachers, even though this effect is purely caused by the selective assignment of teachers to classes.

Table 4 indicates how schools assign teachers to classes based on their ability. The first table row clearly indicates that better teachers are never assigned to better classes. On the contrary, it happens often that better teachers are assigned to weaker classes, which may indicate that many Dutch primary schools have the tendency to focus more on children who perform less well. If vocational teachers are considered to be lower ability teachers then it follows that the effects of tracking on learning gains may be estimated with bias for schools that selectively assigns their teachers to classes. Unfortunately, but for obvious reasons, there is no information for each individual teacher on whether the principal considers them as a lower or higher ability teacher. Nevertheless we can use the information in Table 4 to control for the potential selective assignment of teachers such that a less biased estimate of the tracking effect on the learning gains of primary school children is obtained.

Table 4: Teacher Assignment to Classes.

	HT (N=160)		VT (N=42)			
	Mean	Std. dev.	Mean	Std. dev.	Difference	Std.Err.
Assigned to better classes	0.006	(0.079)	0.000	(0.000)	0.006	(0.006)
Assigned to weaker classes	0.238	(0.427)	0.190	(0.397)	0.047	(0.070)
Not assigned on teacher quality	0.406	(0.493)	0.571	(0.501)	-0.165*	(0.087)
Assignment policy unknown	0.350	(0.478)	0.238	(0.431)	0.112	(0.077)

Note: */ **/*** means statistically significant at the 10/5/1 percent level.

5 Estimation Strategy and Findings

5.1 Estimation Strategy

We are interested how educational tracking of future primary school teachers (V) affects the standardized math and reading performance of primary school children (Y) and estimate the following student and grade fixed effects model:

$$Y_{it} = \alpha_i + V_{it}\beta + X_{it}'\delta + \mu_g + \varepsilon_{it}. \tag{1}$$

In this equation V indicates if teachers have a vocational background, μ_g and α_i represent grade (g) and pupil (i) fixed effects and X_{it} are time-varying class and teacher covariates. The error term, ε_{it}, is assumed to be normally distributed with mean zero and variance σ_ϵ^2 and all explanatory variables are assumed independent of the error term. Subscript t refers to the periods in which primary school children take national and standardized math and reading tests. Primary school children and their teachers are tracked from February 2010 to June 2012 and there can be in total three test periods for math and two test periods for reading. It follows that children can have the same teacher in three consecutive periods for math and in two consecutive periods for reading. Because achieved test scores in one period are correlated with the achieved test scores in another period we cluster the standard errors at the student level.

The tracking effect is, under certain assumptions, represented by β only when a student fixed effect model is estimated, because the estimation parameter then captures how the math and reading performance of primary school children changes because of a switch from a higher secondary teacher to a vocational teacher (or vice versa). The tracking effect can be interpreted as a causal estimate under the assumption that (1) vocational and higher secondary teachers self-select in comparable primary schools with comparable pupil populations, (2) primary school classes remain intact over the different periods, (3) vocational teachers differ in their vocational background from higher secondary teachers but otherwise have similar observed and unobserved teacher characteristics, and (4) the assignment of teachers to classes does not depend on ability or vocational background. We now shortly discuss the validity of these assumptions.

Section 4 shows that vocational and higher secondary teachers are teaching comparable pupil populations at comparable primary schools and this supports the first assumption. The inclusion of a student fixed effect therefore is sufficient to control for observed and unobserved pupil differences between vocational and higher secondary teachers. The second assumption is satisfied for 82 percent of the primary school classes in our sample and therefore we examine if the empirical findings depend on whether classes remain intact.

Having a vocational or a higher secondary educational background is not randomly imposed on primary school teachers, and therefore it is unlikely that the third assumption is satisfied. Table 2 confirms this, in the sense that vocational teachers are, on average, younger and have less teaching experience, which makes it difficult to distinguish tracking effects from age or experience effects. It follows, first of all, that it is important to include all observed teacher characteristics (in X_{it}) that potentially influence the learning performance of pupils. In Subsections 6.1 we, moreover, examine to what extent unobserved differences between vocational and higher secondary teachers can explain the estimated tracking effects. Because the literature indicates that teacher experience negatively affects the math and reading performance of pupils only when teachers have a few years of experience (see, for instance, Rivkin et al., 2005) we examine in Subsection 6.2 how the estimated tracking effects depends on observed experience years.

The fourth assumption, assumes that teacher are not assigned to classes based on their ability or vocational background. To test this we use the information shown in Table 4 and create an indicator variables that indicates 1 if teachers are selectively assigned to classes and zero otherwise. We then estimate the following model student and grade fixed effects model:

$$Y_{it} = \alpha_i + (\beta_0 + \beta_1 I) \cdot V_{it} + X_{it}'\delta + \mu_g + \varepsilon_{it}. \tag{2}$$

The difference with the model presented in Equation 1 is that V is now interacted with the indicator variable which shows if teachers are selectively assigned to classes. The parameter of interest in this study is β_0, because this parameter represents the tracking effect while taking into account that teachers in some primary schools are selectively assigned to classes based on their ability. β_0 thus represents , the tracking effect without selection bias. We note that by estimating Equation 2 we can also be more certain that that assumptions 1 and 2.

5.2 Empirical Findings

Table 5 shows the estimation results for math and reading when we estimate the student and grade fixed effects model indicated in Equation 1. The table presents only the tracking effects, the estimated coefficient for the teacher characteristics included and the coefficient for class size. For both subjects we estimate one model in which we include only grade fixed effects (model I) and one model in which we include grade and student fixed effects (II). It is interesting to compare the estimated coefficient for V of both models because model I. compares the performance of children between vocational and higher secondary teachers, and model II. captures how the math and reading performance of primary school children changes when primary school children switch from a higher secondary teacher to a vocational teacher (or vice versa).

For mathematics we find a non-significant tracking effect in both models which suggest that the math performance of primary school children is not negatively affects by a tracking effect. For reading the grade fixed effect model indicates that primary school children of vocational teachers perform significantly worse than primary school children of higher secondary teachers, but if this negative and significant effect disappears if the tracking effect is estimated by using only variation in test scores of primary school children who switched from a higher secondary teacher to a vocational teacher (or vice versa). Based on the estimation results of Table 5 we cannot conclude that there is a negative tracking effect and, thus, that primary school children of vocational teachers perform worse than primary school children of higher secondary teachers.

With respect to the teacher and class characteristics presented in Table 5 we conclude, first of all, that the significant estimation coefficient on class size indicates that children who switch to larger classes perform less well in math, but not in reading. A more robust result is that primary school children of female teachers perform worse than those of male teachers. This result is interesting because it remains robust even when we both grade and student fixed effects are included, such that the result suggest that the learning gains of children who switched classes and teacher gender were smaller if they were taught by a female teacher. Because we are in this study interested in the tracking effect and because most primary school teachers are female (see Table 2) we examine in Subsection 2 if the estimated tracking effect is different for the subsample of female teachers.

Table 5: Grade and Student Fixed Effects Results

	Math		**Reading**	
	I	II	I	II
V	0.007	-0.028	-0.094***	-0.044
	(0.023)	(0.022)	(0.030)	(0.043)
Female	-0.051**	-0.068**	-0.060*	-0.121**
	(0.025)	(0.030)	(0.034)	(0.061)
Experience (years)	0.002	-0.020***	-0.008	-0.016
	(0.005)	(0.007)	(0.007)	(0.013)
Age (years)	0.003	-0.002	0.023	0.029
	(0.011)	(0.009)	(0.014)	(0.020)
Class size	-0.000	-0.013***	0.003	-0.002
	(0.002)	(0.004)	(0.002)	(0.004)
Student controls	Yes	.	Yes	.
School controls	Yes	.	Yes	.
Student fixed effects	No	Yes	No	Yes
Grade fixed effects	Yes	Yes	Yes	Yes
Adjusted R^2 / Within R^2	0.359	0.320	0.219	0.118
N	8120	8120	5803	5803

Note: */ **/*** means statistically significant at the 10/5/1 percent level. Standard errors are clustered at the student level and are reported in parenthesis. Student controls are gender, mother's education level, ethnicity and age. School conatrols are school size, school located in disadvantaged area, weights (.3 and 1.2) and denomination of school (public, catholic, protestant, Islamic.

The estimation results in Table 5 do not take into account that teachers may be selectively assigned to classes and therefore we estimate equation (2) while including grade and student fixed effects. The estimation results are presented in Table 6. In this study we are particularly interested in the constant tracking effect, which is the β_0-estimate in Equation 2. The interaction effects represent, respectively, if children of vocational teachers perform different than children of higher secondary teachers if teachers were assigned selectively to classes based on their perceived ability and if children of vocational teachers perform different than children of higher secondary teachers if the assignment policy is unknown.

The empirical results suggest negative tracking effects for math and reading if we control for the effect of selective teacher assignment to classes. The tracking effect for math is significant and is one-fifth of a standard deviation, which is substantial. The tracking effect for reading is just not significant at the 10 percent confidence level, but the estimated coefficient is also substantial with approximately one-eighth of a standard deviation. These findings, more intuitively mean that children of vocational teachers perform less well than children of higher secondary teachers, and that this effect did not show up in Table 5 because higher secondary teachers are assigned to weaker classes, while vocational teachers are assigned to better classes. Improving the learning gains of weaker students is apparently more important for schools than improving the learning gains of better students which is likely related to the fact that the primary schools in our sample have relatively many disadvantaged pupils (see Section 4).

Table 6: Results analysis mathematics and reading with selective teacher assignment controls

	Math	Reading
V (Constant tracking effect)	-0.198***	-0.124
	(0.034)	(0.080)
V·selective assignment	0.310***	0.222*
	(0.085)	(0.128)
V·assignment policy unknown	0.219***	0.073
	(0.047)	(0.103)
Female	-0.048	-0.104
	(0.034)	(0.064)
Experience (years)	-0.026***	-0.024*
	(0.009)	(0.014)
Age (years)	0.009	0.043**
	(0.011)	(0.022)
Class size	-0.014***	-0.003
	(0.004)	(0.004)
Student fixed effects	yes	yes
Grade fixed effects	yes	yes
Within R^2	0.324	0.119
N	8120	5803

Note: */ **/*** means statistically significant at the 10/5/1 percent level. Standard errors are clustered at the student level and are reported in parenthesis.

6 Robustness checks

6.1 Can unobserved teacher characteristics explain the negative tracking effects for math?

In the empirical analysis in Section 5 we tried to take into account the selective assignment of teachers to classes, but we cannot be sure that the tracking effect is not driven by selection on unobserved characteristics. In this subsection we follow Altonji et al. (2005, 2008) and Schwerdt and Wuppermann (2011) and examine the extent to which the estimated tracking effect is the result of unobserved characteristics. More specifically, we take the negative significant math effect of -.198 from Table 6, as we believe that this is the most reliable tracking estimate, and examine whether the effect can be explained by unobserved factors under the assumption that the true tracking effect is zero.

We first rewrite Equation 2 into:

$$Y_{it} = \alpha_i + \beta V_{it} + X_{it}' \delta + \eta_i, \tag{3}$$

where $\beta_0 V_{it}$ is now presented as a separate term in Equation 3, and where $\beta_1 I \cdot V_{it}$ and μ_g are now captured by $X_{it}'\delta$. We furthermore conveniently refer to the tracking effect as β instead of β_0. We start by assuming that assume that selection on observables occurs to the same extent as selection on unobservables:

$$\tfrac{Cov(Y,X'\delta)}{Var(X'\delta)} = \tfrac{Cov(Y,\eta)}{Var(\eta)}. \tag{4}$$

Following Altonji et al. (2005), we first estimate how the probability of having a vocational teacher depends on the explanatory variables in our model:

$$V_i = X_i'\gamma + \varepsilon_i. \tag{5}$$

Obviously, $I \cdot V_{it}$ is not included as explanatory variable in this regression, but the variables that indicate how teachers are assigned to classes are included. Then we replace V_i in Equation 3 by $X_i'\gamma + \varepsilon_i$ in equation 5, such that we have (after rearranging terms):

$$Y_i = X_i'(\delta + \beta\gamma) + \beta\varepsilon_i + \eta_i. \tag{6}$$

By construction, we have that $\varepsilon \perp X$ and that $\eta \perp X$ such that

$$plim\,\hat{\beta} = \beta + \tfrac{Cov(Y,\eta)}{Var(\varepsilon)}, \tag{7}$$

Where β represents the true tracking effect and where $\frac{Cov(Y,\eta)}{Var(\varepsilon)}$ represents the bias due to selection on unobservables. $Var(\varepsilon)$ can be obtained by estimating equation 5 and $Cov(Y,\eta)$ can be calculated by using equation 4:

$$Cov(Y,\eta) = \tfrac{Cov(Y,X'\delta)}{Var(X'\delta)} \cdot Var(\eta). \tag{8}$$

We can, however, only calculate $Cov(Y, \eta)$ if we have a consistent estimator of β (see also Appendix A in Schwerdt and Wuppermann, 2011) and for this reason we estimate equation 6 under the assumption that the true effect is zero ($\beta = 0$).

Altonji et al. (2005) consider the situation in which selection on observables occurs to the same extent as selection on unobservables as a upper bound estimate. We note that in our case an upper bound estimate means a less negative effect that is closer to zero. The case when there is no selection on unobservables is considered to be a lower bound estimate. Altonji et al. (2005) argue that the upper bound estimate is a rather conservative and restrictive bound, because it implies that the included control variables in the analysis are a random draw of the full set of relevant control variables which determine the effect of interest. Even though it is likely that selection on unobservables does not occur to the same extent as selection on unobservables, we simply assume that the true effect may lie somewhere between the upper and lower bound effect. We rewrite Equation 4 as

$$\frac{Cov(Y,X'\delta)}{Var(X'\delta)} = \phi \cdot \frac{Cov(Y,\eta)}{Var(\eta)}, \tag{9}$$

such that ϕ represents the ratio between selection on observables and selection on unobservables. To determine the extent to which selection on unobservables can explain our estimated effect of -.198 we calculate the true tracking effect for different ϕ-values. Table 7 shows the potential bias that can be generated by selection on unobservables. Column 6 shows that the estimated tracking effect can be fully explained by selection on unobserved factors if$\phi = 1.597$. Columns two, three and four show the amount of bias and the true tracking estimate for more realistic values of ϕ. Column 5 shows the situation in which selection on observables occurs to the same extent as selection on unobservables situation and Altonji et al. (2005) label this situation as the upper bound estimate. The table shows that the true tracking effect would be -.074 even in the extreme upper bound case in which 63 percent of the estimated tracking effect is driven by selection on unobservables.

For lower values of ϕ (i.e. Columns 2, 3 and 4) the disturbing effect of selection on unobservables becomes smaller and, logically, the true tracking effect becomes more negative. To show that the impact of selection on unobservables is linearly dependent of ϕ, which can be readily seen in Equation 9, we present the bias-estimate ratio in the last row of Table 7. For ϕ=1 this ratio equals .626 and if selection on unobservables equals half the selection on observables then the bias/estimate ratio becomes .313. We conclude that the estimated tracking effect may be partly driven by selection on unobservables, but the estimates in Table 7 suggest that there likely is a substantial and large true tracking effect.

Table 7: Potential bias by selection on unobservables

	ϕ				
	0.25	0.50	0.75	1	1.597
Bias ($\frac{Cov(Y,\eta)}{Var(\varepsilon)}$)	-.031	-.062	-.093	-.124	-.198
True tracking effect (β)	-.167	-.136	-.105	-.074	0
Estimated effect ($\hat{\beta}$)	-.198	-.198	-.198	-.198	-.198
Ratio bias/$\hat{\beta}$	.156	.313	.469	.626	1

6.2 To what extent is the estimated tracking effect driven by teacher experience and gender differences?

The empirical literature shows that teacher experience negatively affects the math and reading performance of pupils, but only when teachers have a few years of teaching experience (see, for instance, Rivkin et al., 2005). At the same time, vocational teachers have structurally less experiences than higher secondary teachers because the educational path to teachers college has been considerably longer for vocational teachers (i.e. three years). In this subsection we examine if the tracking effect for teachers with less than 5 years of teaching experience is different than the tracking effect for teachers with more teaching experience. Less than 5 years of teaching experience was chosen because the empirical literature often refers to 5 teaching years

For this purpose we re-estimate Equation 2 and interact V with a variable that indicates if teachers have less than 5 experience years. The robustness analysis hence departures from the empirical finding that teaching experience only matters for teachers with a few experience years, and if the constant tracking effect disappears when the interaction terms are included in the model then the constant tracking effect appears to be an experience effect.

A second issue is that the children who switched classes and who were taught in the new class by a teacher who had a different gender than the previous teacher performed less well in those classes were there was a female teacher (see Table 5). In this study we are mainly interested if there is a tracking effect and because most primary school teachers are female (see Table 2) we examine if the tracking effect is different for male and female teachers.

Table 8 show the estimation results for math when the teacher experience and gender interactions are included in the estimation model. The second column of the table show the estimation results for math from Table 5 so that it is easy to see how the constant tracking effect changes because of the inclusion of the interaction effects. The estimated constant tracking effect in Column 3 is still negative, significant and about the same size as the estimated tracking effect in the full sample. This suggest that the observed constant tracking effect is not the result of experience differences between vocational and higher secondary teachers. A result worth mentioning is that the variable V interacted with teacher experience, indicates that there is no tracking effect for teachers with less than 5 years of teaching experience.

The constant tracking effect for male teachers (-0.421) is more negative than the constant tracking effect for female teachers (-0.172).[3] For this study it is relevant that we find a significant constant and negative tracking effect for both female and male teachers. The tracking effect for male teachers differs much from the estimated effect for female teachers and this is because most primary school teachers are female teachers such that the estimated tracking effect for male teachers is not so precisely estimated and less stable. That the effect for male teachers is less precisely estimated can also be seen from the fact that the standard error of the constant tracking effect in the third column, which represents the constant tracking effect for males, is more than twice as large than the standard errors of the constant tracking in the other columns.

Table 8: Estimation Results for Math with Teacher Experience and Gender Interactions

	Experience Years:		
	Table 5	$\leq$5 years	Female
V (Constant tracking effect)	-0.198***	-0.216***	-0.421***
	(0.034)	(0.034)	(0.077)
$V\cdot$[$\leq$5 years teaching experience | Female]	.	0.244***	0.249***
	.	(0.055)	(0.068)
$V\cdot$selective assignment	0.310***	0.299***	0.365***
	(0.085)	(0.083)	(0.090)
$V\cdot$assignment policy unknown	0.219***	0.189***	0.259***
	(0.047)	(0.053)	(0.051)
Female	-0.048	-0.086**	-0.174***
	(0.034)	(0.036)	(0.043)
Experience (years)	-0.026***	-0.020**	-0.026***
	(0.009)	(0.009)	(0.009)
Age (years)	0.009	0.019	0.031**
	(0.011)	(0.013)	(0.013)
Class size	-0.014***	-0.014***	-0.014***
	(0.004)	(0.004)	(0.004)
Student fixed effects	yes	yes	yes
Grade fixed effects	yes	yes	yes
Within R^2	0.324	0.329	0.326
N	8120	8120	8120

Note: */ **/*** means statistically significant at the 10/5/1 percent level. Standard errors are clustered at the student level and are reported in parenthesis.

Table 9 show the estimation results with teacher experience and gender interactions included for reading. While the constant tracking effect for reading was substantial but just

[3] We note that both constant tracking effects are significant at the 95 percent confidence level.

not significant at the 10 percent confidence level in Table 6, it becomes significant when V is interacted with the dummy variable that indicates if the teacher has less than 5 years of teaching experience. Similar to the estimation results reported in Table 8 shown for Math, the results suggest that there is no tracking effect for teachers with less than 5 years of teaching experience.

The constant tracking effect for male teachers becomes highly significant and negative when we interact 'being a vocational teacher' with teacher gender. At the same time, the estimation results show that there is a negative but not significant tracking effect for female teachers (-0.059). On the one hand this indicates that the constant negative tracking effect for reading is driven by male teachers, but on the other hand we should keep in mind that the tracking effects for male teachers are less stable and less precisely estimated.

We therefore conclude that the empirical results suggest that there is a substantial tracking effect for math. The results for reading also point to a negative tracking effect, but these results are less robust and precise.

Table 9: Estimation Results for Reading with Teacher Experience and Gender Interactions

	Experience Years:		
	Table 6	≤5 years	Female
V (Constant tracking effect)	-0.124	-0.165**	-0.697***
	(0.080)	(0.080)	(0.122)
V·[≤5 years teaching experience | Female]	.	0.278***	0.638***
	.	(0.096)	(0.111)
V·selective assignment	0.222*	0.220*	0.372***
	(0.128)	(0.124)	(0.129)
V·assignment policy unknown	0.073	0.025	0.191*
	(0.103)	(0.104)	(0.103)
Female	-0.104	-0.139**	-0.406***
	(0.064)	(0.069)	(0.078)
Experience (years)	-0.024*	-0.020	-0.025*
	(0.014)	(0.014)	(0.013)
Age (years)	0.043**	0.062***	0.106***
	(0.022)	(0.024)	(0.024)
Class size	-0.003	-0.004	-0.004
	(0.004)	(0.004)	(0.004)
Student fixed effects	yes	yes	yes
Grade fixed effects	yes	yes	yes
Within R^2	0.119	0.123	0.129
N	5803	5803	5803

Note: */ **/*** means statistically significant at the 10/5/1 percent level. Standard errors are clustered at the student level and are reported in parenthesis.

7 Conclusion

Dutch Primary school teachers must successfully finish teacher college before they are allowed to teach at a primary school. Because of the educational tracking system in the Netherlands, teachers have vocational or higher secondary backgrounds and policymakers and school principals have frequently argued that teachers with vocational backgrounds are less capable to teach math and reading. This claim has however never been empirically supported. This study examines the effects of ability tracking of future primary school teachers on primary school children's math and reading performance

From a broader and scientific perspective this study is interesting because many countries have tracking systems and the consequences of educational tracking are heavily debated among parents, scientists and policymakers (Hanushek and Woessmann, 2006) The consequences of educational tracking are, however, rarely empirically examined. This study examines the effect of ability tracking of future primary school teachers on their children's performance and thereby contributes to the discussion on the potential effects that educational tracking can have.

The empirical findings suggest that math (reading) test scores of children of vocational teachers are lower by .2 (.12) of a standard deviation and we refer to this result as a negative tracking effect. The tracking effect for reading is, however, not significant at the 10% confidence level and the robustness analysis suggest that the estimates on reading are less stable and precise. Robustness analyses show that the tracking estimate for math is not driven by selection on unobserved factors or by experience and gender differences between teachers. The robustness analysis for reading point towards a negative tracking effect but the estimates are not robust or precise enough.

An interesting result is that the negative tracking effects are found after controlling for selective assignment of teachers to classes based on their ability. At first sight, vocational teachers appear to perform equally well as higher secondary teachers but once we control for the fact that teachers are selectively assigned to classes we find that vocational teachers perform less well. Moreover, they are on average assigned to better classes while higher secondary teachers are on average assigned to weaker classes. The effect of having a VE teacher on math test scores is large, in comparison to effect sizes in the literature on the effect of other teacher characteristics on student test scores, like experience and teacher licensure test scores.

Our study shows that ability tracking of future primary school teachers may negatively affect the math and reading performance of primary school children. The question, however, is what policy conclusions can be drawn from this result? We could argue that teacher colleges should admit students only when they successfully finished higher secondary education, but this does not improve the situation for primary schools because there currently is a shortage of primary school teachers. Another possibility is that math and reading deficiencies of future and currently employed vocational teachers are better addressed, for example by offering courses. It is however unclear if and to what extent addressing math and reading deficiencies will be effective because vocational teachers were generally tracked into vocational education *because* their math and reading skills were not sufficient to be tracked into higher education levels. An option which is, in our opinion, worth considering is that primary schools assign more than one teachers to a single class to ensure that each subject is given at the appropriate level. More practically, it is possible to assign a higher secondary and a vocational teacher to a grade three and four class. The higher secondary teacher can specialize in teaching reading and math and vocational teachers can specialize in other subjects. This situation is not so different from secondary education in which specialized teachers tend to give only certain subjects.

References

Agan, A., 2014. Disaggregating the returns to college. Forthcoming Working Paper.

Altonji, J. G., Elder, T. E., Taber, C. R., 2005. Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. Journal of Political Economy 113 (1).

Altonji, J. G., Elder, T. E., Taber, C. R., 2008. Using selection on observed variables to assess bias from unobservables when evaluating swan-ganz catheterization. The American Economic Review 98 (2), 345–350.

Clotfelter, C. T., Ladd, H. F., Vigdor, J. L., 2006. Teacher-student matching and the assessment of teacher effectiveness. Journal of Human Resources 41 (4), 778–820.

Clotfelter, C. T., Ladd, H. F., Vigdor, J. L., 2010. Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects. Journal of Human Resources 45 (3), 655–681.

Ehrenberg, R. G., Brewer, D. J., 1995. Did teachers' verbal ability and race matter in the 1960s? coleman revisited. Economics of Education Review 14 (1), 1–21.

Ferguson, R. F., 1991. Paying for public education: New evidence on how and why money matters. Harv. J. on Legis. 28, 465.

Ferguson, R. F., 1998. Can schools narrow the black-white test score gap. The Black-White test score gap, 318–374.

Ferguson, R. F., Ladd, H. F., 1996. How and why money matters: An analysis of alabama schools. Holding schools accountable: Performance-based reform in education, 265–298.

Hanushek, E. A., 1992. The trade-off between child quantity and quality. Journal of Political Economy, 84–117.

Hanushek, E. A., Woessmann, L., 2006. Does educational tracking affect performance and inequality? differences-in-differences evidence across countries. The Economic Journal 116 (510), C63–C76.

Holmlund, H., Lindahl, M., Plug, E., 2011. The causal effect of parents' schooling on children's schooling: A comparison of estimation methods. Journal of Economic Literature 49 (3), 615–651.

Kukla-Acevedo, S., 2009. Do teacher characteristics matter? new results on the effects of teacher preparation on student achievement. Economics of Education Review 28 (1), 49–57.

Metzler, J., Woessmann, L., 2012. The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation. Journal of Development Economics 99 (2), 486–496.

Murnane, R. J., Phillips, B. R., 1981. What do effective teachers of inner-city children have in common? Social Science Research 10 (1), 83–100.

OECD, 2004. PISA Learning for Tomorrow's World: First Results from PISA 2003. Vol. 659. OECD Publishing.

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

Rowan, B., Chiang, F.-S., Miller, R. J., 1997. Using research on employees' performance to study the effects of teachers on students' achievement. Sociology of Education, 256–284.

Schwerdt, G., Wuppermann, A. C., April 2011. Is traditional teaching really all that bad? a within-student between-subject approach. Economics of Education Review 30 (2), 365–379.

Summers, A. A., Wolfe, B. L., 1975. Equality of educational opportunity quantified: A production function approach.

Summers, A. A., Wolfe, B. L., 1977. Do schools make a difference? The American Economic Review 67 (4), 639–652.

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