



**Teacher characteristics
and their effects on student test scores:
A best-evidence review**

*Johan Coenen, Wim Groot,
Henriette Maassen van den Brink & Chris van Klaveren*

TIER WORKING PAPER SERIES
TIER WP 14/28

Teacher characteristics and their effects on student test scores: A best-evidence review*

Johan Coenen
Wim Groot
Henriette Maassen van den Brink
Chris Van Klaveren

Abstract

There is an abundance of literature that examines how teacher characteristics are related to student test scores. Reviews summarizing the results of these studies do not always consider the strength of the evidence, which might contribute to the fact that the existing research on the effects of teacher characteristics on student test scores remains inconclusive. This review considers the strength of the empirical evidence of the reviewed studies and examines whether teacher characteristics are related to student test scores. This review investigates both acquired and innate teacher characteristics. The literature shows that studies which identify causal effects are scarce.

The literature allows a cautious conclusion that teacher test scores and ability are important predictors of student test scores. Teacher experience is also related to student performance. Especially the first years seem of high importance. Certification in subject, especially in math, is related to higher student test scores. Different roads of teachers to certification do not lead to different student learning gains. Advanced degrees, such as Master degrees, are not associated with higher student test scores. Gender interactions between teachers and students do not seem to be related to student performance. Evidence on national board certified teachers and on race interactions is inconclusive.

1 Introduction

Teacher quality is viewed as an important – if not the most important – determinant of student performance. Researchers, policymakers and schools alike, agree on the importance and relevance of high quality teachers. The concept of teacher quality remains however an intangible indicator. On the one hand, research on teacher value-added concludes that the teacher accounts for a large share in the variation in student test scores. On the other hand,

*We would like to thank Marjolein Coonen for excellent research assistance.

typical observable teacher characteristics, like education and experience, can only explain a fraction of these teacher effects on student outcomes. While there is still a large variation in teacher effects to explain, this review sheds light on the current state-of-the-art of the literature.

Although there are literally hundreds of studies that focus on the importance of teachers for student performance (Hanushek, 2011), there exist only few recent literature reviews concerning the effect of teacher quality on student test scores.

Goe (2007) concludes in her review that experience matters, but only in the first few years. Besides this she highlights the relation between having a math degree and math subject certification and student performance in math. Wayne & Youngs (2003) also did an extensive literature review. Their two main findings are that, (1) students learn more from teachers with higher test scores and higher college ratings. (2) The evidence on the effects of degrees, coursework and certification is inconclusive, with the exception of mathematics: high school students perform better with teachers with certification in, and/or degrees and coursework related to mathematics.

The aim of this literature review is to investigate to what extent observable teacher characteristics can explain these teacher effects on student test scores. In particular, this review focuses on six types of teacher characteristics that are usually distinguished in the literature: teacher education level, teacher certification, teacher test scores and ability, teacher experience, teacher gender and teacher ethnicity. This review aims to give an up to date overview of the literature on teacher characteristics and student test scores, while taking the strength of the underlying studies in account.

We deliberately refer to this review as a best-evidence review to indicate that an overview of the best available empirical literature is given. It is often argued that causal evidence can only be provided by studies that use quasi-experimental or experimental designs (see for instance, Angrist & Pischke (2009); Blundell & Dias (2009)). However, there are only few studies available, which are able to take non-random assignment of teachers and students to classrooms into account. Most studies included in this review are panel studies, while some are (quasi-)experimental studies, with random assignment of students (and not teachers) to classrooms within schools. In both types of studies, teachers might very well be selectively assigned to parallel classes, grade levels and schools, based a.o. on their presumed quality, experience, preferences and degrees.

Even though all studies included in this review control for observable teacher, student and school characteristics, most must make the assumption that this is sufficient to also control for the unobserved selection that may occur. It is often argued that selection of teachers and students to classes and schools based on unobservable characteristics can severely bias estimates of the effect of teacher characteristics on student test scores (see for instance Rothstein (2009)).

This study proceeds as follows. In Section 2, we explain our method of literature search and the inclusion criteria that we used to select the included studies. Furthermore, we provide an overview of the studies that are included in this literature review. Section 3 presents the best-evidence separately for four types of acquired teacher characteristics generally dis-

tinguished in the literature. Section 4 evaluates the evidence on the relationship between teacher gender and ethnicity and student test scores, after which Section 5 discusses the possible effects of publication bias, as we do not include non-published studies in this review. Finally, Section 6 concludes.

2 Method

2.1 Search method

The literature for this review is collected with a three step search method. In the first step, electronic databases were searched for literature on teacher characteristics and student achievement. The databases consulted are: Sage Journals Online, Jstor, PsychLit, EconLit and Google Scholar. At first, relatively general search terms are chosen, such as “teacher effectiveness”, “teacher quality”, “student performance”, “student test score” and “teacher characteristics student achievement”. The literature collected using these general search terms can be classified into the following six categories: teacher education, teacher certification, teacher test scores and ability, teacher experience, teacher gender and teacher ethnicity. In a next step, these specific categories were used as search terms, such as “teacher education student achievement”, “teacher certificate student achievement”, “teacher experience student achievement”. We included only studies that focus on the relationship between teacher characteristics and student test scores.

In a third step, literature is collected by applying the “snowball” principle (i.e. examining the reference list of each study and include all studies that focus on teacher characteristics and student test scores that were not yet included). We then perform the “snowball” principle again for these newly included studies and repeated this process until no new relevant studies were found.

In the last step, we examined websites from authors of included studies, for more relevant studies on teacher characteristics and student test scores. If possible we also searched for other online publication lists in search for more relevant studies. Applying this four step search method resulted in the collection of 93 studies that form the basis for this literature review.

2.2 Inclusion criteria

For this literature review, we impose the following inclusion criteria. Only studies that satisfy these inclusion criteria are included in this review.

1. Studies must make use of data with information on teacher characteristics and standardized student test scores.
2. Studies must account for students’ prior achievement.
3. Studies must account for socioeconomic status.

4. Studies must use (quasi-)experimental design or estimate a panel data model.
5. Studies must have same focus as review.
6. Studies must be published in a peer-reviewed journal or in a thesis.

The first three inclusion criteria are similar to three of the inclusion criteria imposed by Wayne & Youngs (2003). The first criterion is that studies need to make use of data sets which provide them with both information on teacher (-level) characteristics, and standardized test scores of their respective students. This implies that we only include studies which have information on which teachers taught which classrooms. The second criterion recognizes that the student achievement level at the end of a school year is not only influenced by the student's current teachers, but also by prior teachers and their ability. To link student achievement differences to teacher characteristics of their current teacher, it is therefore crucial that studies control for the achievement levels that students bring to the classroom in the beginning of the school year. Following a similar reasoning, we can argue why studies should control for students' socio-economic background characteristics. The potential achievement gains that teachers can establish depend to a large extent on the student characteristics, like their socio-economic background. Furthermore, it may be that teachers with more favorable characteristics select themselves in schools with student populations with on average higher socio-economic status, which would generate biased estimates of teacher effects on student test scores.

The fourth inclusion criterion is that only (quasi-)experimental or panel data studies should be included. Many studies argue that causal evidence can only be provided by studies that use (quasi-)experimental designs (see, among others, Angrist & Pischke (2009); Blundell & Dias (2009)). There are however virtually no studies that provide (quasi-)experimental evidence, based on randomly assigned students, but also randomly assigned teachers. It follows that panel data studies provide the best available evidence on how teacher characteristics are related to student achievement. Panel data studies examine how student achievement gains develop over time and relate these gains to the characteristics of the teachers who taught them. Unfortunately, the estimates of panel data studies cannot be interpreted causally, because selection of students and teachers into schools and classes is not random. Chetty et al. (2013) however find in their study that most of the bias that originates from non-random sorting of students to teachers, is captured by prior student test scores. For this review, this implies that studies which lack randomized experimental data can still provide valid evidence on the effect of teacher characteristics on student test scores, if they employ a sufficient set of control variables in their analysis. Nevertheless, we can distinguish between three types of panel data studies, which provide stronger or weaker evidence:¹

- Panel data studies in which students are randomly assigned to classes (type 1)

¹ There are some studies which make use of the Project STAR data in which both students and teachers were randomly assigned to treatment and control status, but this experiment was aimed at investigating the effect of class size reduction, not at teacher effects on student achievement (Dee, 2004; Nye et al., 2004). These studies are labeled as type 0 in the tables below.

- Panel data studies that consider the possible effects of unobserved heterogeneity (type 2)
- Other panel data studies (type 3).

Panel data studies in which students are randomly assigned to classes also relate student achievement gains to teacher characteristics but, in addition use data where students, (but not teachers), are randomly assigned to classes within schools. Therefore, these panel data studies with randomly assigned students have an advantage over other panel studies, because they control for selection of better students to classes within schools. Nevertheless, both studies cannot ensure that teachers are randomly assigned to classes and schools, and teachers may therefore be selectively distributed over classes and schools based on their (unobserved) teacher skills.

The panel data studies included in this review therefore only control for observable teacher, student and school characteristics, and must assume that this is sufficient to also control for the selective distribution of teachers over classes and schools. Since this assumption cannot be empirically tested, it may be that the studies considered in this review suffer from selection bias and that the conclusion drawn upon these studies are biased.

The fifth inclusion criterion ensures that studies are only included in this review if the study and review focus are similar. It frequently occurs that panel studies satisfy the first two inclusion criteria, present estimates on how teacher characteristics are associated with student achievement gains, but do not examine how teacher characteristics are related to student achievement gains (see, for example, Van Klaveren (2011); Schwerdt & Wuppermann (2011)). Usually teacher characteristics are included as control variables, and not to estimate the effect of teacher characteristics on student achievement. These studies are not included in this review.

The sixth inclusion criterion is that studies must be accepted for publication in international and peer-reviewed journals or must be published as a chapter in a peer-reviewed thesis. The intuition of this inclusion criterion is that reviewers, editors, promoters and committee members recognize analytical and data problems and therefore these problems are better addressed if a study is published and peer-reviewed. This criterion is somewhat controversial though, because it is well known that studies with positive and significant findings are more likely to be published (the so-called publication bias). It implies that due to this sixth inclusion criterion, this review may present empirical findings that are upward biased. On the other hand, this inclusion criterion reduces the probability that empirical results of studies are driven by analytical problems or data problems in studies, because peer-reviewed publications are far more likely to recognize analytical or data deficiencies (Van Klaveren & De Wolf, 2013).

Even though all 93 collected studies focus on teacher characteristics and student test scores, they frequently adopt different identification strategies, apply different empirical estimation methods, or use different data. The strength of the empirical evidence may therefore vary across studies and this may lead to more heterogeneous, and possibly inconsistent research findings.

40 of the 93 collected studies did not satisfy the inclusion criteria formulated in Section 2.2. Of the 53 studies included in the review, 12 studies are not published in a peer-reviewed journal or thesis. In Section 2.2 we pointed out that publication bias may bias the review findings, therefore we also consider the 12 unpublished studies and discuss whether their findings contrast with the findings of published studies.

2.3 Selection of the literature

Since we make use of the best available evidence rather than strictly causal studies, it is also of importance to consider the external validity of the studies. Besides the findings of the studies, we also want to know to what extent they are representative outside the sample and of interest for policy purposes.

Table 1 describes information relevant for the external validity of the studies that satisfy the formulated inclusion criteria and are included in this review. The columns show information about the author(s), the research design used, the observation years of their data, how they measure student test scores, the student grade or grades included in their sample and (when available) the number of students and teachers in their analysis.

Table 1: External validity of all studies included in the literature review: published in peer-reviewed journals

| Study | Method | Observation years | Measure | Grades | Students | Teachers | Country |
|-------------------------------|--------|----------------------|------------|--------|----------|----------|---------|
| Aaronson et al. (2007) | PD | '96-'98 | Gains | 9 | 52957 | 589 | US |
| Bosshardt & Watts (1990) | PD | 1987 | Level, pre | 12 | 2383 | 94 | US |
| Boyd et al. (2006) | PD | '98-'03 | Gains | 3-8 | 960970 | 3766 | US |
| Boyd et al. (2008) | PD | '00&'05 | Gains | 3-8 | 578630 | | US |
| Carrell & West (2010) | EX2 | '00-'07 | Gains | HE | 10534 | 91 | US |
| Clotfelter et al. (2006) | PD | '00-'01 | Level, pre | 5 | 60791 | 3233 | US |
| Clotfelter et al. (2007) | PD | '94-'03 | Both | 3-5 | 180k | | US |
| Clotfelter et al. (2010) | PD | '99-'02 | WSBS | 10 | 137597 | | US |
| Croninger et al. (2007) | PD | '98&'00 | Gains | 1 | 5167 | 1342 | US |
| Darling-Hammond et al. (2005) | PD | '96-'01 | Gains | 4-5 | 132071 | 4408 | US |
| Dee (2004) | EX1 | '85-'89 | Level | K-3 | 8310 | | US |
| Dee (2007) | PD | 1988 | WSBS | 8 | 21324 | | US |
| Dee & Cohodes (2008) | PD | 1988 | WSBS | 8 | 16901 | | US |
| Ehrenberg et al. (1995) | PD | '88&'90 | Gains | 10 | 3029 | | US |
| Goldhaber & Anthony (2007) | PD | '96-'98 | Gains | 3-5 | 611517 | 32448 | US |
| Goldhaber & Brewer (1997) | PD | '88&'90 | Gains | 10 | 5149 | | US |
| Goldhaber & Brewer (2000) | PD | '88,'90,'92 | Level, pre | 12 | 3786 | 2098 | US |
| Hanushek (1992) | EX2 | '71-'75 | Gains | 2-6 | 441 | 22 | US |
| Harris & Sass (2009) | PD | '00-'03 | Gains | 3-10 | +75k | 32000 | US |

| | | | | | | | |
|----------------------------|-----|---------|------------|-------|----------|--------|---------|
| Harris & Sass (2011) | PD | '99-'04 | Gains | 3-10 | +/-250k | 1300 | US |
| Hill et al. (2005) | PD | '00-'03 | Gain, pre | 1&3 | 2963 | 699 | US |
| Holmlund & Sund (2008) | PD | '97-'04 | Level | 10-12 | 16200 | | Sweden |
| Jepsen (2005) | PD | '91-'94 | Gains | 1&3 | 2652 | | US |
| Kane et al. (2008) | PD | '99-'04 | Gains | 4-8 | +/- 300k | | US |
| Krieg (2005) | PD | '02-'03 | Level, pre | 4 | 49415 | | US |
| Kukla-Acevedo (2009) | PD | '00-'02 | Gains | 5 | 3812 | 120 | US |
| Metzler & Woessmann (2012) | PD | 2004 | WSBS | 6 | 4302 | | Peru |
| Monk (1994) | PD | '87-'90 | Level, pre | 10-11 | 2829 | 1091 | US |
| Mullens et al. (1996) | PD | '90-'91 | Gains | 3 | 1043 | | Belize |
| Muñoz & Chang (2007) | PD | '05-'06 | Gains | 9 | 1487 | 58 | US |
| Murnane & Phillips (1981) | PD | '73-'75 | Level, pre | 3-6 | 814 | | US |
| Neugebauer et al. (2011) | PD | 2001 | Level | 4 | 5858 | | Germany |
| Neild et al. (2009) | PD | 2003 | Level, pre | 5-8 | 22853 | 539 | US |
| Nye et al. (2004) | EX1 | '85-'89 | Both | K-3 | 6377 | | US |
| Palardy & Rumberger (2008) | PD | 1999 | Level, pre | 1 | 3496 | | US |
| Rockoff (2004) | PD | '89-'00 | Level | K-6 | +/-10k | 297 | US |
| Rowan et al. (1997) | PD | 1988 | Level, pre | 10 | 5381 | | US |
| Sharkey & Goldhaber (2008) | PD | '90&'92 | Gains | 12 | 486 | 224 | US |
| Sokal et al. (2007) | EX2 | 2006? | Gains | 3-4 | 175 | RA | Canada |
| Summers & Wolfe (1977) | PD | '68-'71 | Gains | 6 | 627 | | US |
| Winters et al. (2013) | PD | '00-'04 | Both | 3-10 | +/-1700k | +/-13k | US |

Note: PD= Panel data study; EX1=Experimental study with both teachers and students randomly assigned;

EX2=Experimental study with students randomly assigned; Level, pre= Test scores in levels with control for pretest score; Both=Level, pre and gains; WSBS=Within Student Between Subject; M=Matching

Table 2: External validity of all studies included in the literature review: not published in peer-reviewed journals

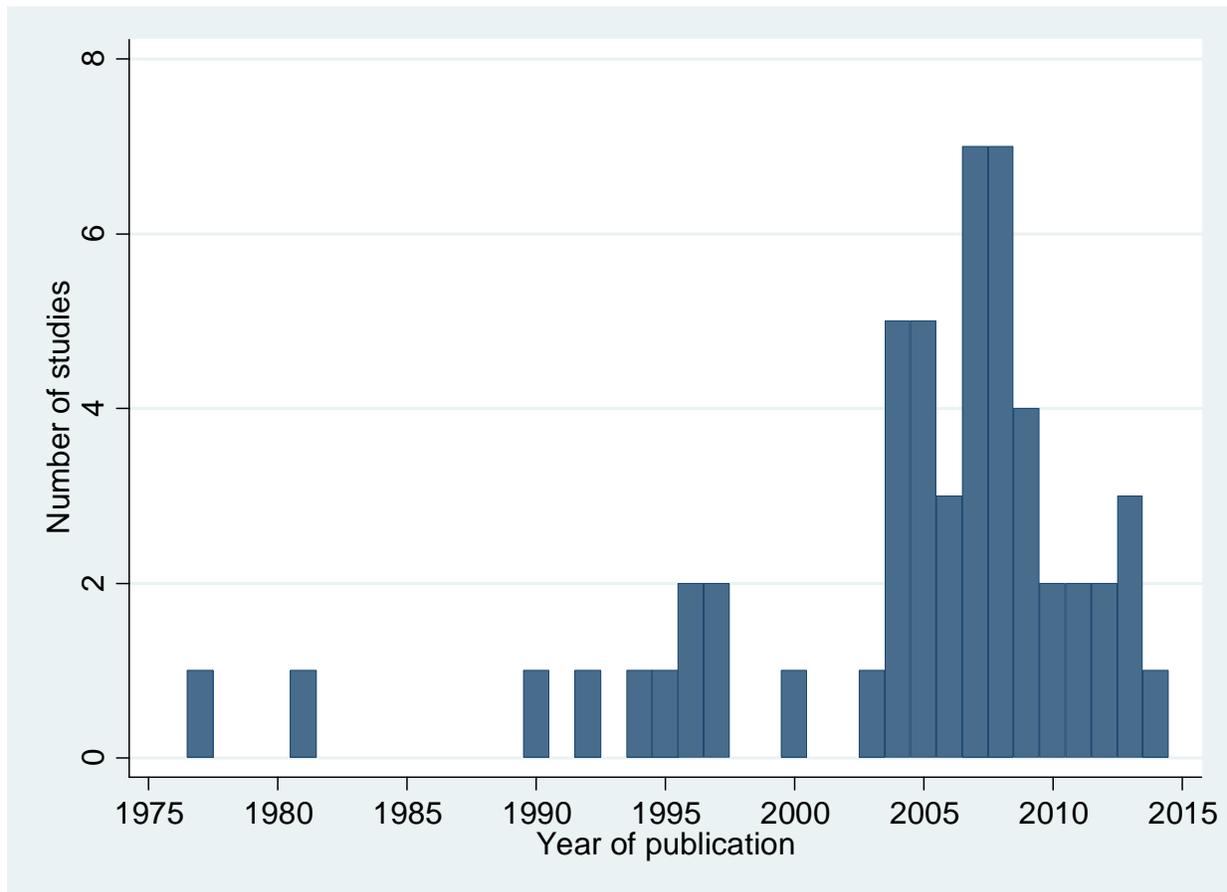
| Study | Method | Observation years | Measure | Grades | Students | Teachers | Country |
|------------------------------|--------|-------------------|------------|--------|----------|----------|-------------|
| Ammermüller & Dolton (2006) | PD | '95,'99,'01,'03 | Gains | 4,8 | +/-8k | | US&UK |
| Antecol et al. (2012) | EX2 | '01-'02 | Level | 1-5 | 1900 | 100 | US |
| Betts et al. (2003) | PD | '97-'99 | Gains | 2-9? | | | US |
| Cantrell et al. (2008) | EX2 | '03-'04 | Both | 2-5 | 3790 | 99 | US |
| Cavalluzzo (2004) | PD | '00-'02 | Level, pre | 9-10 | 103k | 2109 | US |
| Coenen & Van Klaveren (2013) | PD | '10-'11 | Level, M | 3-5 | 2586 | 174 | Netherlands |
| Coenen & Van Klaveren (2014) | PD | '10-'12 | Both | 3-5 | 3602 | 202 | Netherlands |
| Constantine et al. (2009) | EX2 | '03-'05 | Level | K-5 | 2600 | 174 | US |
| Decker et al. (2004) | EX2 | '01-'02 | Level | 1-5 | +/-2000 | | US |
| Goldhaber & Brewer (1996) | PD | 1990 | Level, pre | 10 | 6196 | | US |
| Muralidharan & Sheth (2013) | PD | '05-'09 | Level, pre | 1-5 | 66660 | | India |
| Sanders et al. (2005) | PD | '99-'02 | Both | 4-8 | 37k | 122 | US |

Note: PD= Panel data study; EX1=Experimental study with both teachers and students randomly assigned; EX2=Experimental study with students randomly assigned; Level, pre= Test scores in levels with control for pretest score; Both=Level, pre and gains; WSBS=Within Student Between Subject; M=Matching

From the tables above we can conclude that most of the studies investigate teacher effects in the US. Only 5 from the 41 published studies do not use US data. From the unpublished studies 4 out of 12 are not based (exclusively) on US data. As was mentioned before, most studies use panel data, while some make use of experimental data. Of the experimental studies, all but two have students randomly assigned to classes, but not teachers. The included studies in general test the effect of teacher characteristics on student test scores either by test scores in level, test scores in level while controlling for prior achievement or by test scores gains between period t and t-1. Some alternative approaches include within student between subject test score differences, and test scores in level with matching as an alternative for including prior achievement.

Figure 1 presents a frequency histogram of the publication years of studies listed in Table 1.

Figure 1: Number of studies on the influence of teacher characteristics on student test scores, per year



3 Acquired teacher characteristics and student test scores

In this section, we discuss the literature on how acquired teacher characteristics affect student test scores, based on the available evidence. We address these characteristics separately in four subsections: (1) Education, in which we discuss to what extent teacher degrees, and the quality of the undergraduate college they attended, matter for student achievement. (2) Certificates, in which we evaluate the existing literature on three different types of studies: the effect of being certified in the subject a teacher teaches, the effect of alternative pathways into teaching, like Teach For America and the effect of National Board Certified Teachers (NBCT) on student performance. (3) Test scores and ability, in which the effect of teacher ability, usually proxied by teacher test scores, on student test scores is examined. (4) Experience, in which we discuss the extensive literature on the effects of teacher experience on student learning outcomes. Innate teacher characteristics, like gender and race, we will cover in the next section.

3.1 Education

Studies which investigate how teachers' own education affect student test scores tend to focus on the highest obtained degree, but other elements of teachers' own educational career are also studied. Some studies are interested in the quality of the colleges teachers attended, as a proxy for ability, while other studies look at the coursework teachers have taken during their training in teacher college.

It is not straightforward how teacher education can affect student learning. It is for instance unclear why having a Master or even a PhD degree instead of a Bachelor degree would make teachers teach better. These teachers with different highest degrees should all be capable to teach in terms of content, and didactic skills may be more important. It also depends on the country (and in the case of the US, even on the state), whether having a Master degree is an additional degree, which may distinguish those teachers from teachers who only acquired a bachelor degree, or a strict requirement to be fully certified as a teacher.

In Table 3, we present the main findings of studies which focus on the highest attained degree. In Table 4, we summarize findings of studies which investigate the effect of college quality of the teacher.

Table 3: Findings of studies on teacher education level

| Study | Type | Type of Education | Subjects | Results | |
|---------------------------|------|--|--|--------------------------------|--|
| | | | | Relation | Effect |
| <i>Published</i> | | | | | |
| Hanushek (1992) | 1 | Master | Reading/ vocabulary | 0 | |
| Murnane & Phillips (1981) | 1 | Master | Vocabulary | 0 | |
| Harris & Sass (2011) | 2 | AD newly acquired | Math/reading | El: 0 Mi: +/- | |
| Clotfelter et al. (2010) | 2 | AD | Math/English /biology/ELP | 0 | |
| Dee & Cohodes (2008) | 2 | GD | Math/reading/ science/social studies | 0 | |
| Aaronson et al. (2007) | 2 | B major, M, PhD | Math | 0 | |
| Clotfelter et al. (2007) | 2 | M, AD, PhD | Math/reading | M: 0 / - A: -/- PhD: -/0 | M:-.004-.008 A:-.025-.052 PhD: -.093 |
| Croninger et al. (2007) | 2 | AD, Elementary education degree vs. other degree | Math/reading | 0, 0/+ | .078 |
| Clotfelter et al. (2006) | 2 | Master | Math/reading | - / 0 | -.023 |
| Jepsen (2005) | 2 | B, B+, M | Math/reading | 0 | |
| Muñoz & Chang (2007) | 3 | Master or higher M, PhD, B/M | Reading | 0 | |
| Goldhaber & Brewer (2000) | 3 | major S, B major E, M major E | Math/science | 0,0,+/0,- /0,0 | |
| Rowan et al. (1997) | 3 | B, M major in math | Math | + | .015 |
| Goldhaber & Brewer (1997) | 3 | B, M major in math | Math | +, + | |
| Monk (1994) | 3 | M,M+ | Math/science | 10: -/-,0/- 11: 0/0,0/- | |
| <i>Not published</i> | | | | | |
| Betts et al. (2003) | 3 | Master, PhD (H) | Math/reading | E: +/0 H: 0/+ | |
| Goldhaber & Brewer (1996) | 3 | B, M major in subject | Math/Science/ English/History | + / +/ 0/0 | |

Note: effect sizes are given when interpretable, available in the studies and statistically significant. Different estimated coefficients are separated by commas; estimations of the same variables for different subjects are separated by a slash sign (/). B=Bachelor, M=Master, AD=Advanced degree (higher than bachelor), GD=Graduate Degree, Major S=Major in subject, Major E=Major in education, El=Elementary school, Mi=Middle school.

The main findings of the literature on the effects of having an advanced degree are that students from teachers with a master degree do not perform better or worse than students from teachers with only a bachelor degree. This conclusion is rather persuasive, since a wide range of both older and more recent studies among different grade levels and different subjects do not find students from teachers with an advanced degree to perform significantly better or worse than students from teachers which only obtained a bachelor degree (for instance: Hanushek (1992); Clotfelter et al. (2010); Croninger et al. (2007)).

While having advanced degrees in general does not appear to be a relevant signal in predicting teacher quality, subject-specific bachelor and master degrees with majors in math or science are found to be associated with higher student math test scores (Goldhaber & Brewer, 1997, 2000; Rowan et al., 1997). The results of these studies can however not be interpreted as causal, since the data and design do not allow causal inference from these studies. It could be the case that the timing of when a teacher acquires an advanced degree is of importance. Harris & Sass (2011) find that a recently acquired advanced degree is related to higher student performance in math, but also with lower student performance in reading.

The two unpublished studies find some positive associations between having a master or a PhD degree, or having a bachelor/master with a major in subject, but the evidence is not very consistent over different grades or subjects.

Table 4: Findings of studies on undergraduate college quality of teachers

| Study | Type | Type of Education | Subjects | Results | |
|---------------------------|------|----------------------|--------------------------|----------|--------|
| | | | | Relation | Effect |
| <i>Published</i> | | | | | |
| Clotfelter et al. (2010) | 2 | CQ: very competitive | Math/English/biology/ELP | + | .019 |
| Boyd et al. (2008) | 2 | CQ: competitive | Math | + | .014 |
| Clotfelter et al. (2006) | 2 | CQ | Math/reading | 0 | |
| Murnane & Phillips (1981) | 3 | CQ | Vocabulary | 0 | |
| Summers & Wolfe (1977) | 3 | CQ | General | + | |

Note: effect sizes are given when interpretable, available in the studies and statistically significant. Different estimated coefficients are separated by commas; estimations of the same variables for different subjects are separated by a slash sign (/). CQ=College Quality

The results of studies on teachers' undergraduate college quality are mixed. These studies make use of the Barron's ranking of college selectivity and usually divide colleges in 4 categories: very competitive, competitive, not competitive and unranked. Some studies find students from teachers who attended more competitive colleges to perform better, but usually only in math (Boyd et al., 2008; Clotfelter et al., 2010; Monk, 1994). Boyd et al. (2008) find that students from teachers who have attended a competitive college perform 1.4% of a standard deviation better compared to students from other teachers. Clotfelter et al. (2010) also find a small effect of college quality on grade 10 student outcomes. They conclude that students performed 1.9% of a standard deviation higher when their teacher attended a very

competitive college. On the other hand, in another paper with grade 5 students, the same authors do not find a significant relation between college quality and student test scores.

3.2 Teaching certificates

Especially in the US, teaching certification is a highly relevant topic. Unlike in most European countries, in the US, a substantial amount of students are taught by a teacher without a certification. Therefore, there exists an abundant literature on teacher certification, which we here broadly divide in 3 categories: (1) studies which examine differential learning outcomes of students from traditionally certified and alternatively certified teachers, (2) studies on the effect of teacher certification in subject and (3) studies which investigate whether NBCT teachers perform better in the classroom.

In Table 5, the main findings from studies on traditional versus alternative pathways into teaching are shown.

Table 5: Findings on teacher certification: alternative pathways into teaching

| Study | Type | Type of Certification | Subjects | Relation | Results Effect |
|-------------------------------|------|---|--------------|----------------------------|------------------------------|
| <i>Published</i> | | | | | |
| Kane et al. (2008) | 2 | Fellows, TFA, intern. recruits, other uncertified | Math/reading | 0/-,+/ 0,-/0,0 | -.012, .024, -.023 |
| Boyd et al. (2008) | 2 | IE, fellows, TFA, temp, other Temp or emerg. | Math | -,-,0,-,- | -.019-.032 |
| Sharkey & Goldhaber (2008) | 2 | certification, private school certification | Math/science | + /0, 0 | |
| Croninger et al. (2007) | 2 | Regular or alt. vs none, temp, prov, emerg, prob | Math/reading | 0 | |
| Boyd et al. (2006) | 2 | IE, fellows, TFA, temp, other | Math/ELA | M:-,-,0,-,- E:0,-,-,-,- | M:-.018-.033 E:-.018-.050 |
| Jepsen (2005) | 2 | Not fully certified | Math/reading | 0 | |
| Palardy & Rumberger (2008) | 3 | Full certification | Math/reading | + | |
| Darling-Hammond et al. (2005) | 3 | Uncertified, alt., emerg or temp, out-of-field, no-test certified | Math/reading | -/- | |
| Darling-Hammond et al. (2005) | 3 | TFA | Math/reading | +/-/0 | |
| Goldhaber & Brewer (2000) | 3 | Probationary, emergency in subject, private school certification | Math/science | 0, 0, -/0 | -.010 |
| <i>Not published</i> | | | | | |
| Constantine et al. (2009) | 1 | Alt. vs regular | Math/reading | 0 | |
| Decker et al. (2004) | 1 | TFA | Math/reading | + /0 | .150-.260 |

Note: effect sizes are given when interpretable, available in the studies and statistically significant. Different estimated coefficients are separated by commas; estimations of the same variables for different subjects are separated by a slash sign (/). IE=Individual Evaluation; TFA= Teach For America; Alt.=Alternative; Temp, prov, emerg, prob=Temporary, provisional, emergency or probational certification

In most studies, there are no significant differences in student performance between regularly and alternatively certified teachers, although some studies find negative learning outcomes for students from alternatively certified teachers.

Kane et al. (2008) investigated whether the teacher certification status matters for student achievement. Using six years of panel data of both students and teachers, they compare teacher effectiveness of regular certified teachers, regular uncertified teachers and three types

of alternatively certified teachers in New York City. They find hardly any differences in the average achievement impacts of certified, uncertified and alternatively certified teachers. They also suggest that teacher classroom performance during their first two years of teaching is a more reliable indicator of their future effectiveness.

Boyd et al. (2006) study whether different pathways into teaching in New York City affect student achievement. They investigate whether teachers who enter through new routes into teaching with reduced coursework are as effective as other teachers. They find that teachers from new routes attain higher achievement gains than temporary license teachers. Compared to university trained teachers, they often attain smaller achievement gains (from 2 to 5% of a standard deviation) in both mathematics and English language arts. Most differences disappear over time as the cohort matures. In another study, Boyd et al. (2008) find very similar results.

The two unpublished studies do not have very different results, although Decker et al. (2004) do find a positive effect of Teach For America certified teachers on their students' performance.

Decker et al. (2004) evaluate the effectiveness of Teach For America (TFA) teachers. By randomly assigning students to TFA teachers and other teachers in the same schools and in the same grades. They find that students of TFA teachers performed 15% of a standard deviation better in math, than students of other teachers. For reading, they find no significant differences. In another analysis, they restrict the sample of control teachers to become more comparable to TFA teachers in their average amount of teaching experience. In this specification, they find that students of TFA teachers scored 26% of a standard deviation higher math test scores than students of the other teachers. For reading, they find again no significant differences.

Constantine et al. (2009) investigate the effectiveness of several AC (Alternative route to Certification) teachers in comparison to TC (Traditional route to Certification) teachers. They use schools throughout the US which recently employed AC teachers and had grade levels with at least both one AC teacher and one relatively inexperienced TC teacher. Students were randomly assigned to either a TC teacher or an AC teacher. Within schools and grade levels, the matching of each TC and AC teacher constituted a mini-experiment. They find no significant differences in performance in math and reading between students of AC and students of TC teachers. Furthermore, they find no evidence from their study that more coursework in teacher training is associated with higher student achievement. There were no significant differences between both low-coursework AC and TC teachers, and high coursework AC and TC teachers. Lastly, they also find no evidence of an association between a teachers' coursework content and student test scores. They conclude that the prospective teacher's road to certification is on average unlikely to provide information about teacher effectiveness with regard to student learning outcomes.

In Table 6, the main findings from studies on teacher certification in subject are shown.

Table 6: Findings on teacher certification: in subject

| Study | Type | Type of Certification | Subjects | Results | |
|----------------------------|------|--|---|------------------------------|-----------|
| | | | | Relation | Effect |
| <i>Published</i> | | | | | |
| Clotfelter et al. (2010) | 2 | Subject | Math/English /biology/ELP | + / + / 0 / 0 | .110/.103 |
| Boyd et al. (2008) | 2 | None, Math/ Science/special ed/other | Math | - , 0 | -.042 |
| Dee & Cohodes (2008) | 2 | Subject | Math/reading /science/ social studies | + / 0 / 0 / + | .116/.081 |
| Sharkey & Goldhaber (2008) | 2 | Not subject-certified | Math/science | 0 | |
| Neild et al. (2009) | 3 | special ed, other field, not certified vs primary subject certified | Math/science | 0 / +, - / - 0 / -, - / - | |
| Goldhaber & Brewer (2000) | 3 | Not certified subject vs certified other subject or not certified | Math/science | - / 0 | -.010 |
| Goldhaber & Brewer (1997) | 3 | Certified, in math | Math | - , + | |

Note: effect sizes are given when interpretable, available in the studies and statistically significant. Different estimated coefficients are separated by commas; estimations of the same variables for different subjects are separated by a slash sign (/).

Most studies find no effects from certification alone. Certification in subject on the other hand is often found to be related to higher student test scores in these subjects, especially for math performance. Some recent studies show that students perform better in math when their teacher is certified in math (Clotfelter et al. (2010); Dee & Cohodes (2008); Neild et al. (2009)). For other subjects, there is no clear relation found over multiple studies.

Clotfelter et al. (2010) for instance investigate how several teacher credentials affect student test scores. Using data on high school students from ninth and tenth grade, they employ a within student between subject design to identify the effect of among others being certified in subject. They find 11% of a standard deviation higher math test scores for students from subject-certified teachers in these subjects. Besides for math, Clotfelter et al. (2010) furthermore find a similar positive effect of 11% of a standard deviation of subject-certified teachers in English on student English test scores. They find no such effect for the subjects biology and economic, legal and political systems (ELP).

Dee & Cohodes (2008) find a similar coefficient for students from math certified teachers. They perform 12% of a standard deviation better than teachers without certification in math. They also find a positive effect of 8% of a standard deviation for subject certified teachers in social sciences. For subject certified teachers in reading and science, they do not find significant differences compared to students from teachers who are not certified in reading and science.

Boyd et al. (2008) find only negative effects on math performance from having a teacher with no certification, while they find no significant differences between students who had

teachers certified in math, science, special education or in another subject. According to their study, students from teachers without any certification perform 4% of a standard deviation lower in math. Another recent study finds no significant relation between subject certified teachers and student test scores in math and science (Sharkey & Goldhaber, 2008).

In Table 7, the main findings from studies on National Board Certified Teachers (NBCT) are shown.

Table 7: Findings on teacher certification: NBCT

| Study | Type | Type of Certification | Subjects | Results | |
|----------------------------|------|---|---------------------------|-----------------|-------------------------|
| | | | | Relation | Effect |
| <i>Published</i> | | | | | |
| Clotfelter et al. (2010) | 2 | Precertification, application, certification | Math/English /biology/ELP | +,+,+ | .022/.049/.049 |
| Harris & Sass (2009) | 2 | NBCT | Math/reading | 0 | |
| Clotfelter et al. (2007) | 2 | NBCT | Math/reading | + / + | .018-.061/ .012-.038 |
| Clotfelter et al. (2006) | 2 | NBCT | Math/reading | 0 / + | .030 |
| Goldhaber & Anthony (2007) | 3 | Future certification, current certification vs no certification | Math/reading | + / +, 0 / + | .050 / .040, .020 |
| <i>Not published</i> | | | | | |
| Cantrell et al. (2008) | 1 | Non-applicants, unsuccessful applicants vs certified | Math/language | 0, - / - | -.173 / -.134 |
| Sanders et al. (2005) | 2 | NBCT certified vs others | Math/reading | 0 | |
| Cavalluzzo (2004) | 2 | failed/withdrawn vs no certification | Math | +,+,- | .074 / .019 / -.026 |

Note: effect sizes are given when interpretable, available in the studies and statistically significant. Different estimated coefficients are separated by commas; estimations of the same variables for different subjects are separated by a slash sign (/). NBCT=National Board Certified Teacher.

Some studies find that students from National Board Certified Teachers perform better, others find no significant differences between teachers who are certified, unsuccessful applicants or non-applicants. Clotfelter et al. (2006, 2007, 2010) find mainly positive effects of having a teacher who is a national board certified teacher. In three different studies, they find mostly significant positive relations between NBCTs and student test scores, with the exception of math test scores for fifth grade students (Clotfelter et al., 2006). Goldhaber & Anthony (2007) also suggest positive outcomes for students with a NBCT. They distinguish between future certification: teachers who were not certified when they were teaching the students in the data, but were certified at the time of the study, and current certification, teachers who were already certified when they taught the students. In this way, they examine whether National Board Certification is merely a signal of high quality teachers, or if the process of obtaining the certification has a quality effect on the teachers as well, i.e. makes them better teachers. For math, they find evidence for both a signal and a quality effect of

the certification. Students from teachers who would receive certification later performed 5% of a standard deviation better in math, while they performed 4% of a standard deviation better if their teacher already was National Board Certified. For reading on the other hand, they find a smaller coefficient of 2% of a standard deviation for students with a currently certified teacher.

The findings of research that was not published in a peer-reviewed journal are similarly inconclusive. Cantrell et al. (2008) perform an experiment in which they randomly assign Los Angeles region classrooms of elementary students to either NBC applicants or comparison teachers of the same school. Furthermore, they also use similar data on pairs of teachers which are not randomly assigned to students, to compare estimated teacher impacts on classrooms with and without randomly assigned teachers. They find that in the experimental setting, certified teachers were not more effective than non-applicants, while teachers who unsuccessfully applied were less effective compared to both certified teachers and non-applicants. Students from unsuccessful applicants performed 17% of standard deviation worse in math and 13% of a standard deviation worse in language in comparison to students from certified or non-applicant teachers. The results of their analysis without randomization show similar, although slightly smaller estimates. This suggests that non-experimental value-added estimates can be relevant for predicting student achievement in an experimental setting. Their findings do suggest that the NBPTS is capable of distinguishing between high quality and low quality teachers. However, when certified teachers are compared to the large majority of teachers who never applied for NBCT certification, there are no significant differences found in their students' test scores.

Cavalluzzo (2004) does find differential effects for four different subgroups of teachers. She finds that students from certified teachers perform 7% of a standard deviation better in math than students from non-certified teachers (who did not apply). Students from teachers who were in the application process also performed slightly better, but only by 2% of a standard deviation. Students from failed or withdrawn teachers performed 3% of a standard deviation worse than students from non-certified teachers who never applied. In another non-peer reviewed report, Sanders et al. (2005) do not find evidence that students from NBC teachers perform better than students from other teachers.

3.3 Teacher test scores and ability

Teacher test scores and other measures of teacher ability are other frequently studied teacher characteristics. The reasoning of why a relation between teacher ability and student test scores could be expected is straightforward: smarter teachers should be more capable to teach students, because they have a higher ability themselves. The most investigated teacher test scores are licensure test scores, while other types of teacher test scores and other measures of teacher ability which have been investigated include verbal skills tests, GPA scores, SAT scores and similar tests taken by both teachers and students.

In table 8, we show the main findings of the literature on teacher test scores and ability.

Table 8: Findings on teacher test scores

| Study | Type | Teacher test | Subjects | Results | |
|------------------------------|------|--|------------------------------|------------------|---------------------|
| | | | | Relation | Effect |
| <i>Published</i> | | | | | |
| Hanushek (1992) | 1 | Verbal skills | Reading/ vocabulary | + / 0 | |
| Murnane & Phillips (1981) | 1 | Verbal skills | Vocabulary | 0 | |
| Metzler & Woessmann (2012) | 2 | Similar tests teachers and students | Math/reading | + | .090 |
| Clotfelter et al. (2010) | 2 | Licensure | Math/English /biology/ELP | + / - / + / 0 | .047/-.022/ .016 |
| Kukla-Acevedo (2009) | 2 | Overall GPA/Math GPA | Math | + / 0 | .034-.084 |
| Boyd et al. (2008) | 2 | Math, verbal SAT | Math | + , - | .041/-.033 |
| Clotfelter et al. (2007) | 2 | Elementary/early childhood education test | Math/reading | + | .015-.068 |
| Clotfelter et al. (2006) | 2 | Licensure | Math/reading | + / 0 | .012 |
| Hill et al. (2005) | 3 | Content knowledge test math, reading | Math | + , 0 | |
| Rowan et al. (1997) | 3 | Math quiz test score | Math | + | .020 |
| Mullens et al. (1996) | 3 | Math score end primary school | Math | + | |
| Summers & Wolfe (1977) | 3 | Licensure | General | - | |
| <i>Not published</i> | | | | | |
| Coenen & Van Klaveren (2014) | 2 | Tracking based on test at age 12 | Math/reading | + / 0 | .198 |

Note: effect sizes are given when interpretable, available in the studies and statistically significant. Different estimated coefficients are separated by commas; estimations of the same variables for different subjects are separated by a slash sign (/).

The majority of the studies on teacher test scores and ability find that students from teachers with higher test scores also perform better themselves. The evidence suggests that both overall test scores and specific math test scores can be a good predictor of teacher ability and consequently student test scores. Reading and verbal skills tests on the other hand are mostly unrelated to student outcomes. While there are some studies in which math and overall test scores are also positively related to student reading outcomes (see for instance Clotfelter et al. (2007); Metzler & Woessmann (2012)), there is more consistent evidence on the existence of a positive association between teacher test scores and student math outcomes.

Already decades ago there were studies on the relation between teacher test scores and student test scores. Murnane & Phillips (1981) use data from a welfare reform experiment with only black families, and thus, black students in the sample. They find no relation

between vocabulary test scores of students and the test scores of their teachers on verbal skill tests. Hanushek (1992) using the same data set, also investigates the association with reading test scores of students. Like Murnane & Phillips (1981), he does not find an association between the verbal skills test scores of teachers and vocabulary test scores of their students, however he does find a positive association between verbal skill tests of teachers and the reading test scores of their students.

Clotfelter et al. (2006) conclude that, besides experience, licensure test scores are a second consistent predictor of student achievement. One standard deviation higher teacher licensure test scores are related to 2 percent of a standard deviation higher student math test scores in elementary education. Clotfelter et al. (2007) include teacher test scores from the Elementary education test and the Early Childhood Education test. At least one of these tests was required for all elementary school teachers in North Carolina. In a linear specification they find that one standard deviation higher teacher test scores are related to 1% of a standard deviation higher student test scores in math and to 0.3% of a standard deviation in reading. In another specification they found non-linear effects. From 4% of a standard deviation lower math test scores for students from a teacher with more than 2 standard deviations below average test score, to 3% of a standard deviation higher math test scores for students from a teacher with more than 2 standard deviations above average test score.

In another study, Clotfelter et al. (2010) also suggest a positive relation between licensure test scores and high school student performance of respectively 5% and 2% of a standard deviation for math and biology. On the other hand, they find a negative relation between licensure test scores and student test scores for English.

Boyd et al. (2008) employ math and verbal SAT scores from teachers as an indicator of teacher ability. They find that higher math SAT scores are positively related to student test scores, but verbal SAT scores are negatively related to student performance. With respectively 4% and 3% of a standard deviation, their estimated effect sizes for math and verbal SAT scores are quite comparable.

Kukla-Acevedo (2009) states that teacher qualifications and experience cannot be seen independently of each other and therefore investigates the joint relationship between experience and qualifications, like teacher GPA, to determine whether experience has a consistently positive effect on student achievement. Of all the indicators of teacher qualifications and experience she examines, overall GPA is the only indicator which consistently has a positive effect on student test scores in math. This positive relation is not constant over time, because teacher overall GPA and experience have a combined effect. Teachers with a lower GPA reduce the gap in teaching effectiveness once they become more experienced.

Metzler & Woessmann (2012) make use of Peruvian elementary education data, which features for both students and their teachers test scores from similar (but not identical) tests for math and reading. Their results suggest that a one standard deviation rise in teacher test scores is associated with a rise of 10% of a standard deviation of student test scores. The authors conclude that teacher achievement test scores is the only teacher characteristic that is a consistent predictor of student achievement.

The results of the unpublished study from Coenen & Van Klaveren (2014) are in line with the main findings of the published studies: they do not use direct teacher test score data, but use the tracking in secondary education in the Netherlands which is a direct consequence of the final test students take at the end of primary education. They find that students from teachers who were in the lowest secondary education track (i.e. lower test scores at age 12) perform 20% of a standard deviation lower in math scores, compared to students from teachers who were in the intermediate or highest secondary education track. For reading they also find a large, but non-significant effect.

3.4 Teacher experience

The literature on the effect of teacher experience on student outcomes includes several different assumptions concerning the functional form of an effect of teacher experience on student outcomes. Many studies include experience as a linear variable, assuming that each additional year of experience makes teachers better in teaching, and subsequently could lead to higher positive effects on student learning. Another approach which is often used in studies, categorical variables for different experienced teachers, allows for heterogeneous experience effects. Yet other studies only distinguish between the first years of experience and all later years, based on earlier research which indicates that only the first years of experience make a difference.

In Table 9 we show the findings of the literature on teacher experience.

Table 9: Findings on Teacher Experience

| Study | Type | Experience | Subjects | Results Relation | Effect |
|---------------------------|------|--|--|-------------------------------|---------------------------|
| <i>Published</i> | | | | | |
| Nye et al. (2004) | 0 | More than 3 years | Math/reading | M3, R2 : + M1,2/ R1,3:0 | .189/.142 |
| Carrell & West (2010) | 1 | Novice vs experienced | Contemp. courses/follow on courses | -/+ | -.690/.700 |
| Hanushek (1992) | 1 | Linear | Reading/vocabulary+ | | |
| Murnane & Phillips (1981) | 1 | <8,=>8<15, =>15 years | Vocabulary | + / - / + | |
| Harris & Sass (2011) | 2 | Categories | Math/reading | E: +/+ M: +/+ H: -/- | .040-.160 -.040- -.160 |
| Clotfelter et al. (2010) | 2 | To 5 years, most in 1st 2 | Math/English /biology/ELP | + | .048-.061 |
| Kukla-Acevedo (2009) | 2 | Interaction teacher GPA | Math | + | |
| Boyd et al. (2008) | 2 | 1st 3-5 years | Math | + | max .060 |
| Clotfelter et al. (2007) | 2 | To 27 years, half in 1st 2 | Math/reading | + | .092/.118 .067/.096 |
| Aaronson et al. (2007) | 2 | Linear | Math | 0 | |
| Croninger et al. (2007) | 2 | More than 2 years | Math/reading | 0 / + | .060 |
| Boyd et al. (2006) | 2 | 2 years vs 1, 3 years vs 2, more years vs 3 years | Math/ELA | + / +, + / +, 0 | .076/.047, .030/.018 |
| Clotfelter et al. (2006) | 2 | To 27 years, half in 1st 2 | Math/reading | + | .100/.080 |
| Jepsen (2005) | 2 | Linear | Math/reading | 1: 0 / 0 3: + / + | .021/.020 |
| Rockoff (2004) | 2 | 1st 2 years, linear | Math/reading | + / 0, + / + | .100 .150/.180 |
| Muñoz & Chang (2007) | 3 | linear | Reading | 0 | |
| Goldhaber & Brewer (2000) | 3 | linear | Math/science | 0 / + | |
| Monk (1994) | 3 | linear | Math/science | 10: 0 / 0 11: + / - | |
| Bosshardt & Watts (1990) | 3 | linear | Other/econ. | - | -.080/-.100 |
| Summers & Wolfe (1977) | 3 | linear | General | + / - | |
| <i>Not published</i> | | | | | |
| Betts et al. (2003) | 3 | novice vs experienced | Math/reading | 0 | |
| Goldhaber & Brewer (1996) | 3 | linear | Math/Science/ English/History | 0 | |

Note: effect sizes are given when interpretable, available in the studies and statistically significant. Different estimated coefficients are separated by commas; estimations of the same variables for different subjects are separated by a slash sign (/).

The main findings of the literature are that there is in fact a positive relation between teacher experience and student outcomes. There does not seem to be a linear effect of teacher experience: most studies find significant student learning gains for the first couple of years of experience, but hardly later on in the teacher's career.

Nye et al. (2004) make use of experimental data on grades K to 3 from Project STAR. They distinguish between teachers with 3 years experience or less and teachers with more than 3 years of experience. They only find significantly positive results for grade 3 students in math and for grade 2 students in reading, while almost all the other estimates are positive but not significant.

Boyd et al. (2006) find in their study that teacher experience seems to matter, but only in the first years of the teacher's career. They find that students from teachers with two years experience have 7.6% of a standard deviation higher achievement gains in math compared with students from teachers with one year experience. Additionally, students from teachers with three years experience have 3% of a standard deviation higher achievement gains in math, compared to students from teachers with two years of experience. Beyond three years of experience, they do not find significantly higher achievement gains. Importantly, they distinguish between actual teacher experience effects and selection effects through teacher attrition. After taking into account that relatively weak teachers are more likely to quit teaching after the first years, they no longer find a significant effect of experience on student achievement.

Boyd et al. (2008), using a different sample and different cohorts, find similar results, although somewhat smaller: the maximum gain of one additional year of experience is found in the second year compared to the first year of experience. This gain equals 6% of a standard deviation. Until the fifth year of experience the authors find statistically significant additional achievement gains.

Clotfelter et al. (2006) conclude that teacher experience consistently predicts improved student achievement. The effect of a very experienced teacher on an average student's math (reading) test scores is approximately 10 percent (8 percent) of a standard deviation compared to a teacher with 0 years of experience. In contrast to Boyd et al. (2006, 2008), they find experience effects from the first years of experience to more than 27 years of experience. Effect sizes are smallest for the first two years of experience; respectively 5 and 3.5% of a standard deviation for math and reading. For math test scores, they also find larger returns for teacher experience for students with a higher SES and for students who are more able in math.

In another paper, Clotfelter et al. (2010) find comparable estimates of the effect of teacher experience on student performance in both mathematics and English. Some other studies however do not find an impact of teacher experience (f.i. Goldhaber & Brewer (1997, 1996); Aaronson et al. (2007); Muñoz & Chang (2007)).

Carrell & West (2010) look at the effect of teacher quality on student achievement in post-secondary education. Using a design with students randomly assigned to either inexperienced lecturers or experienced associate or full professors, they find mixed results: students from the less experienced teachers perform 70 percent of a standard deviation better in introductory

courses, while students from more experienced teachers perform 69% of a standard deviation better in more advanced courses.

In contrast to published studies, the outcomes for the 2 studies which have not been published in a peer-reviewed journal find no significant relation between experience and student outcomes.

4 Teacher gender and race

In this section, we will discuss the literature on innate teacher characteristics and student test scores. Teacher gender and race are on the one hand less policy relevant, because they can obviously not be changed by teachers. On the other hand, it can guide policy into optimal teacher classroom assignments. Research on teacher gender and race has one advantage over research on teacher qualifications, credentials and experience: gender and race are exogenous variables. This implies that intra-school non-random assignment of teachers to classes can only take place if schools structurally assign better (or worse) students to female (or male) teachers, or to white (or black/Asian/Hispanic) teachers. Selective assignment of teachers to classrooms is not uncommon. Clotfelter et al. (2005) for instance show in their study that novice teachers are more often assigned to black students than to white students.

In Section 4.1 we will discuss the literature on teacher gender. In Section 4.2, we will focus on research which investigates the influence of teacher race on student performance.

4.1 Teacher gender

Research on the effect of teacher gender on student outcomes has especially in the last few decades become very relevant. The teaching profession has since become to a large extent a female profession, with the percentage of female teachers in primary education nowadays as high as 85%. If male and female teachers have differential effects on male and female students, than this majority of female teachers can contribute to increasing or closing the gender education gap.

The literature has pointed to two potential mechanisms why gender interactions between teachers and students could affect student performance. An objective same-gender effect and a subjective same-gender effect. An objective effect, i.e. on standardized test scores, could take place, because same-gender teachers would function as good role models for their students of the same gender (see for instance: Driessen (2007); Holmlund & Sund (2008)). A subjective effect implies no actual quantifiable learning effect, but higher grading by teachers who favor (or disfavor) students from the same gender.

In table 10, the main findings of the literature on gender interactions is shown.

Table 10: Findings on gender interactions

| Study | Type | Gender | Subjects | Results | |
|------------------------------|------|-----------------------------------|------------------------------------|----------------------|--------------------------|
| | | | | Relation | Effect |
| <i>Published</i> | | | | | |
| Sokal et al. (2007) | 1 | Same-gender, boys | Reading | 0 | |
| Winters et al. (2013) | 2 | Teacher: Female | Math/reading | E: 0 / 0 M/H: +/+ | .013-.028 / .007-.016 |
| Neugebauer et al. (2011) | 2 | Same-gender | Math/German /science | 0 | |
| Clotfelter et al. (2010) | 2 | Male T - Female S, Other comb. | Math/English/ biology/ELP | - , 0 | -.105 |
| Holmlund & Sund (2008) | 2 | Same-gender | Math/Swedish /English | 0 | |
| Dee (2007) | 2 | Same-gender | Math/science /English/history | + | .045 - .050 |
| Krieg (2005) | 3 | Same-gender | Math/reading/ writing/listening | 0 | |
| Ehrenberg et al. (1995) | 3 | Gender and race | Math/science /English/history | 0 | |
| <i>Not published</i> | | | | | |
| Antecol et al. (2012) | 1 | FT-FS, FT math major-FS, FT-MS | Math/reading | -/0, 0, 0 | -.08- -.200 |
| Coenen & Van Klaveren (2013) | 2 | Same-gender | Math | 0 | |
| Muralidharan & Sheth (2013) | 2 | Same-gender | Math/language | + | .032-.036 |
| Ammermüller & Dolton (2006) | 2 | Same-gender | Math/reading/science | Mixed | |

Note: effect sizes are given when interpretable, available in the studies and statistically significant. Different estimated coefficients are separated by commas; estimations of the same variables for different subjects are separated by a slash sign (/). FT=Female Teacher; MT=Male teacher; FS=Female Student; MS=Male Student

Most of the studies on gender effects find no relation between teacher gender and student outcomes: gender interactions do not seem to matter for student performance.

Holmlund & Sund (2008) investigate whether same gender teachers affect student performance of students from academic secondary education in Sweden. They find that the gender gap in subjects with a higher proportion of female teachers is larger, after they correct for differences in different subjects in teacher gender composition. They argue however, that this finding is caused by non-random assignment of students to teachers, and by teacher selection into different subjects. Therefore, using either teacher turnover or student mobility, they identify the within student and subject effect. They find practically no significant effects from having a same gender teacher in this specification with student fixed effects.

Sokal et al. (2007) also find no effect of a same gender teacher (i.e. male teacher) on the reading performance of boys. They conducted a 10 week long intervention of additional reading instruction for boys, who were low achievers in reading in Canadian schools. Al-

though they do not find an effect of teacher gender on reading performance, they do show how the intervention has positively affected boys' perception of reading.

Neugebauer et al. (2011) do not find any relation between teacher gender and student performance. They conclude that boys do not benefit from male teachers and girls do not (significantly) benefit from female teachers, neither on blind test scores, nor on grades given by the teachers.

Dee (2007) does find evidence in favor of a positive same-gender effect. He investigates whether same-gender teacher assignment influenced student achievement, teacher perceptions of student performance, and student engagement. He analyzes this with a within student between subject design. He differences two separate equations, in which the student's performance in a subject is a function of observed student characteristics, and the subject's teacher gender. His within-student comparisons indicate that assignment to a same-gender teacher significantly improves student test scores of boys and girls, but also the teacher's perception of student performance, and student engagement with the taught subject.

The evidence from studies which are not published in peer-reviewed journals is mixed. In a developing country context, with primary education students in India, Muralidharan & Sheth (2013) find positive same-gender teacher effects on both math and reading. On the other hand, using Dutch primary education data, Coenen & Van Klaveren (2013) do not find an effect of assignment to a same-gender teacher. Ammermüller & Dolton (2006) find mixed results in their study. They use data on reading, math, and science, from PIRLS and TIMSS of grade 4 and grade 8 students from the US and the UK. They find a positive same-gender teacher effect for grade 8 boys, for math in the US, and for science in England. Furthermore, using student fixed effects, they also find positive same-gender teacher effects in math, for both boys and girls of grade 8, but only in England, not in the US, and only in 2003, not in 1995 or 1999. Antecol et al. (2012) use data from a randomized experiment and find that female students perform significantly worse in math with a female teacher. When they also control for whether the female teacher has a major in math, they find a positive, although non-significant estimate. For boys, they find no differences in math or reading performance based on the gender of their teacher.

4.2 Race

There are also studies which investigate whether teacher-student interactions of the same or a different race can influence student outcomes. The mechanisms behind such effect are similar as with gender interactions: (1) objective effects, because same race teachers are supposedly better role models for students and (2) subjective effects, because positive discrimination might occur in teacher grading of students of the same race and/or negative discrimination in teacher grading of students of another race. Only studies which focus on the first category fall into the focus of this review, since we concentrate on studies between teacher characteristics and actual student test scores, not grades or other subjective assessments by teachers.

In Table 11, the main findings of studies on the relation between teacher race and student achievement are shown.

Table 11: Findings on race interactions

| Study | Type | Race | Subjects | Results | |
|--------------------------|------|-----------------------------------|----------------------------------|----------|-------------------------|
| | | | | Relation | Effect |
| <i>Published</i> | | | | | |
| Dee (2004) | 0 | Same race (W/B) | Math/reading | + / + | |
| Clotfelter et al. (2010) | 2 | White T - Black S, White | Math/English | +,+,- | .036, .055, -.083 |
| | | T - Hisp. S, Black T - White S | /biology/ELP | | |
| Muñoz & Chang (2007) | 3 | Race dummy (minority=1) | Reading | 0 | |
| Ehrenberg et al. (1995) | 3 | Gender and race | Math/science/ English/history | 0 | |

Note: effect sizes are given when interpretable, available in the studies and statistically significant. Different estimated coefficients are separated by commas; estimations of the same variables for different subjects are separated by a slash sign (/). T=Teacher; S=Student

The main findings are mixed. Some studies find positive same-race effects, others find both beneficial and detrimental effects from teacher-student interactions of different races. Yet, other studies find no significant influence of race interactions whatsoever.

Dee (2004) uses experimental data from Project STAR in which teachers and students are randomly assigned to classrooms. He finds positive same-race effects for both black and white teacher-student interactions, for math and reading. Asian and Hispanic interactions are not taken into account, because of too small sample sizes.

Clotfelter et al. (2010) find that both black and Hispanic students perform better when they have a white teacher. Black students perform 5.5% of a standard deviation better, while Hispanic students perform 3.6% of a standard deviation better. In contrast, they also find that white students perform 8% of a standard deviation worse when their teacher is black. The studies from Ehrenberg et al. (1995) and Muñoz & Chang (2007) did not find any effects of same-race teacher student interactions.

5 An overview of the results

In Table 12, the main outcomes of this literature review are shown. In the first three columns, a general tentative conclusion about the results per topic is given for respectively math, reading and other subjects. In the last three columns we show the number of studies which find positive results, negative results or non-significant results for each topic. The general tentative conclusion is based on the number of studies with positive, non-significant and negative findings, combined with the strength of the evidence provided by the respective studies.

Table 12: Overall tentative conclusions for math, reading and other subjects, and number of studies with positive, non-significant and negative

| Topic | Math | Reading | Other | # + | # 0 | # - |
|-----------------------|-------|---------|-------|-----|-----|-----|
| Education level | 0 | 0 | 0 | 5 | 15 | 5 |
| College quality | + | 0 | + | 3 | 2 | 0 |
| Alternative pathways | 0 | 0 | 0 | 5 | 10 | 6 |
| Subject certification | + | 0 | 0 | 7 | 6 | 0 |
| NBCT certification | + / 0 | + / 0 | n.a. | 6 | 5 | 0 |
| Teacher test scores | + | + / 0 | n.a. | 11 | 7 | 3 |
| Experience | + | + | + / 0 | 17 | 11 | 6 |
| Same-gender teacher | 0 | 0 | 0 | 3 | 6 | 0 |
| Teacher female | 0 | 0 | 0 | 1 | 4 | 0 |
| Race | + / 0 | + / 0 | 0 | 2 | 2 | 1 |

Note: n.a.=non applicable, when there are not enough relevant studies to draw an overall conclusion

The evidence of the studies in this review suggest that college quality, certification in subject, teacher test scores and experience affect student test scores positively. On the other hand, education level, pathways into teaching certification and teacher gender do not seem to matter for student test scores. There exists some evidence of beneficial effects of NBCT certification and teacher-student race interactions, but there are at least an equal amount of studies which do not find evidence for this.

6 Discussion and conclusion

The literature allows a cautious conclusion that certain teacher characteristics are associated with student performance. Firstly, teacher test scores and ability appear to be an important predictor of student test scores. Secondly, teacher experience is also related to student performance, although most studies find significant test score gains only during the first years of experience of teachers. Thirdly, certification in subject, especially in math, is related to higher student test scores. On the other hand, different pathways into teaching do not seem to affect student performance. Traditionally and alternatively certified teachers do not establish different learning gains. Furthermore, advanced degrees, such as master degrees, are not associated with higher student test scores. Gender interactions between teachers and students do not seem to be related to student performance. Evidence on national board certified teachers and on race interactions remains mixed and inconclusive.

Although the research on teacher quality has contributed to our knowledge of which teacher characteristics improve learning outcomes of their students, there remains a gap between the estimated teacher effects on student outcomes and the extent in which underlying observable teacher characteristics can account for these effects. Apparently easily measurable characteristics like education, credentials and experience can explain only a small part in the variation of teacher quality and the resulting effects on student test scores.

Another problem remains the lack of (quasi-)experimental designs, or in the absence of this, information about teacher and student assignment in schools. This makes the estimation of causal effects of teacher characteristics on learning performance difficult, since non-random assignment of teachers and/or students to schools and classes can lead to serious selection bias in estimates. In this review we acknowledge the lack of experimental studies and therefore consider the relative strength of the available evidence from mostly panel data studies. Still we cannot consider the results from these studies as causal.

Future research should aim to incorporate additional information about teachers, besides their credentials, experience, gender and race. Less observable information like teacher motivation, teaching styles, being able to motivate students might be of performance. Future research should also be aimed at either using (quasi-)experimental data, or at collecting information which enables researchers to adjust their estimates for selective assignment of teachers and students to classrooms.

References

- Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), 95–135.
- Ammermüller, A., & Dolton, P. J. (2006). Pupil-teacher gender interaction effects on scholastic outcomes in England and the USA. Tech. rep., ZEW Discussion Papers.
- Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.
- Antecol, H., Eren, O., & Ozbeklik, S. (2012). The effect of teacher gender on student achievement in primary school: Evidence from a randomized experiment.
- Betts, J. R., Zau, A., & Rice, L. (2003). *Determinants of student achievement: New evidence from San Diego*. Public Policy Institute of California San Francisco.
- Blundell, R., & Dias, M. C. (2009). Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources*, 44(3), 565–640.
- Bosshardt, W., & Watts, M. (1990). Instructor effects and their determinants in precollege economic education. *Journal of Economic Education*, (pp. 265–276).
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy*, 1(2), 176–216.
- Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008). The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. *Journal of Policy Analysis and Management*, 27(4), 793–818.
- Cantrell, S., Fullerton, J., Kane, T. J., & Staiger, D. O. (2008). National board certification and teacher effectiveness: Evidence from a random assignment experiment. Tech. rep., National Bureau of Economic Research.
- Carrell, S. E., & West, J. E. (2010). Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Political Economy*, 118(3), 409–432. URL <http://ideas.repec.org/a/ucp/jpolec/v118y2010i3p409-432.html>
- Cavalluzzo, L. C. (2004). *Is National Board Certification an effective signal of teacher quality?*. CNA Corporation Alexandria, VA.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2013). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. Tech. rep., National Bureau of Economic Research.

- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. (2005). Who teaches whom? race and the distribution of novice teachers. *Economics of Education review*, *24*(4), 377–392.
- 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. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, *26*(6), 673–682.
- 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.
- Coenen, J. B., & Van Klaveren, C. (2013). Better test scores with a same-gender teacher? Tech. rep., Top Institute for Evidence Based Education Research.
- Coenen, J. B., & Van Klaveren, C. (2014). The effects of ability tracking of future primary school teachers on student performance. Tech. rep., Top Institute for Evidence Based Education Research.
- Constantine, J., Player, D., Silva, T., Hallgren, K., Grider, M., Deke, J., et al. (2009). An evaluation of teachers trained through different routes to certification. *Final Report for National Center for Education and Regional Assistance*, *142*.
- Croninger, R. G., Rice, J. K., Rathbun, A., & Nishio, M. (2007). Teacher qualifications and early learning: Effects of certification, degree, and experience on first-grade student achievement. *Economics of Education Review*, *26*(3), 312–324.
- Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). Does teacher preparation matter? evidence about teacher certification, teach for america, and teacher effectiveness. *education policy analysis archives*, *13*(42), 1–51.
- Decker, P. T., Mayer, D. P., & Glazerman, S. (2004). *The effects of Teach for America on students: Findings from a national evaluation*. University of Wisconsin–Madison, Institute for Research on Poverty.
- Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics*, *86*(1), 195–210.
- Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources*, *42*(3), 528–554.
- Dee, T. S., & Cohodes, S. R. (2008). Out-of-field teachers and student achievement evidence from matched-pairs comparisons. *Public Finance Review*, *36*(1), 7–32.

- Driessen, G. (2007). The feminization of primary education: effects of teachers sex on pupil achievement, attitudes and behaviour. *International Review of Education*, 53(2), 183–203.
- Ehrenberg, R. G., Brewer, D., & Goldhaber, D. (1995). Do teachers' race, gender, and ethnicity matter? evidence from the nels. *Industrial and Labor Relations Review*, 48(3), 547–561.
- Goe, L. (2007). The link between teacher quality and student outcomes: A research synthesis. *National Comprehensive Center for Teacher Quality*.
- Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? national board certification as a signal of effective teaching. *The Review of Economics and Statistics*, 89(1), 134–150.
- Goldhaber, D., & Brewer, D. (1996). Evaluating the effect of teacher degree level on educational performance. In W. J. Fowler Jr (Ed.) *Developments in school finance*. Washington, DC: National Center for Education Statistics. US Department of Education.
- Goldhaber, D. D., & Brewer, D. J. (1997). Why don't schools and teachers seem to matter? assessing the impact of unobservables on educational productivity. *Journal of Human Resources*, (pp. 505–523).
- Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? high school teacher certification status and student achievement. *Educational evaluation and policy analysis*, 22(2), 129–145.
- Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy*, (pp. 84–117).
- Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, 30(3), 466–479.
- Harris, D. N., & Sass, T. R. (2009). The effects of nbpts-certified teachers on student achievement. *Journal of Policy Analysis and Management*, 28(1), 55–80.
- Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of public economics*, 95(7), 798–812.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers mathematical knowledge for teaching on student achievement. *American educational research journal*, 42(2), 371–406.
- Holmlund, H., & Sund, K. (2008). Is the gender gap in school performance affected by the sex of the teacher? *Labour Economics*, 15(1), 37–53.
- Jepsen, C. (2005). Teacher characteristics and student achievement: Evidence from teacher surveys. *Journal of Urban Economics*, 57(2), 302–319.

- Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? evidence from new york city. *Economics of Education Review*, *27*(6), 615–631.
- Krieg, J. M. (2005). Student gender and teacher gender: What is the impact on high stakes test scores. *Current Issues in Education*, *8*(9), n9.
- 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.
- Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of education review*, *13*(2), 125–145.
- Mullens, J. E., Murnane, R. J., & Willett, J. B. (1996). The contribution of training and subject matter knowledge to teaching effectiveness: A multilevel analysis of longitudinal evidence from belize. *Comparative Education Review*, *40*(2), 139–157.
- Muñoz, M. A., & Chang, F. C. (2007). The elusive relationship between teacher characteristics and student academic growth: A longitudinal multilevel model for change. *Journal of Personnel Evaluation in Education*, *20*(3-4), 147–164.
- Muralidharan, K., & Sheth, K. (2013). Bridging education gender gaps in developing countries: The role of female teachers. Tech. rep., National Bureau of Economic Research.
- 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.
- Neild, R. C., Farley-Ripple, E. N., & Byrnes, V. (2009). The effect of teacher certification on middle grades achievement in an urban district. *Educational Policy*, *23*(5), 732–760.
- Neugebauer, M., Helbig, M., & Landmann, A. (2011). Unmasking the myth of the same-sex teacher advantage. *European Sociological Review*, *27*(5), 669–689.
- Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational evaluation and policy analysis*, *26*(3), 237–257.
- Palardy, G. J., & Rumberger, R. W. (2008). Teacher effectiveness in first grade: The importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis*, *30*(2), 111–140.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, *94*(2), 247–252.

- Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy*, 4(4), 537–571.
- 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*, (pp. 256–284).
- Sanders, W. L., Ashton, J. J., & Wright, S. P. (2005). *Comparison of the Effects of NBPTS Certified Teachers with Other Teachers on the Rate of Student Academic Progress. Final Report.* Arlington, VA: National Board for Professional Teaching Standards.
- Schwerdt, G., & Wuppermann, A. C. (2011). Is traditional teaching really all that bad? a within-student between-subject approach. *Economics of Education Review*, 30(2), 365–379.
- Sharkey, N. S., & Goldhaber, D. (2008). Teacher licensure status and student achievement: Lessons from private schools. *Economics of Education Review*, 27(5), 504–516.
- Sokal, L., Katz, H., Chaszewski, L., & Wojcik, C. (2007). Good-bye, mr. chips: male teacher shortages and boys reading achievement. *Sex roles*, 56(9-10), 651–659.
- Summers, A. A., & Wolfe, B. L. (1977). Do schools make a difference? *The American Economic Review*, 67(4), 639–652.
- Van Klaveren, C. (2011). Lecturing style teaching and student performance. *Economics of Education Review*, 30(4), 729–739.
- Van Klaveren, C., & De Wolf, I. (2013). Systematic reviews in education research: When do effect studies provide evidence? Tech. rep., Top Institute for Evidence Based Education Research.
- Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational research*, 73(1), 89–122.
- Winters, M. A., Haight, R. C., Swaim, T. T., & Pickering, K. (2013). The effect of same-gender teacher assignment on student achievement in the elementary and secondary grades: Evidence from panel data. *Economics of Education Review*.

TIER WORKING PAPER SERIES

TIER WP 14/28

© TIER 2014

ISBN 978-94-003-0098-9