

The role of teacher quality in education production

Bjarne Strøm^a, Torberg Falch^b

^aDepartment of Economics, Norwegian University of Science and Technology, Norway; ^bDepartment of Teacher Education, Norwegian University of Science and Technology, Norway

Introduction

Most people find it obvious that the quality of teachers matter for students' outcomes. Credible measurement of the quantitative importance of teachers is challenging, and in particular what characterizes strong and weak teachers. The last years have witnessed an explosion of studies of teacher quality, especially from the US, as a result of increased availability of longitudinal administrative data matching students, teachers and schools.

This chapter summarizes recent attempts to measure the effect of teacher quality on student outcomes within an education production framework. We consider to what extent different research strategies are able to provide credible evidence on teacher quality. Estimated teacher effects on student achievement may reflect that some teachers are better at teaching to the test rather than generating true knowledge. Thus, we also discuss to what extent teacher quality measures based on student achievement translate into education and labor market outcomes after students have left school, which illustrates

the potential economic value of increased teacher quality. We discuss to what extent teacher quality is related to individual teacher characteristics and teacher labor market conditions. The chapter also discusses the potential usefulness of value added based teacher quality measures in teacher evaluation systems and pay policies.

Estimating teacher quality

Our point of departure is the education production function as formulated recently in Koedel, Mihaly, and Rockoff (2015), but building heavily on the novel insights provided by Ben-Porath (1967) and Hanushek (1979).

$$y_{it} = f(X_i(t), S_i(t), T_i(t), \alpha_{i0}, \varepsilon_{it}) \quad 22.1$$

y_{it} is an outcome measure for student i at time (grade) t . The outcome is assumed to be generated by the history of inputs relevant for the student where, $X_i(t)$ represents the history of individual and family inputs for student i up to time t . $S_i(t)$ similarly represents all school inputs,

such as resource use and class size, while $T_i(t)$ represents teacher quality. α_{i0} is student i 's initial ability, while ε_{it} is a stochastic idiosyncratic term.

A main question in the literature of teacher quality is how to specify and estimate the production function in order to obtain unbiased estimates of the relationship between student outcomes and teacher quality. In other words, the question is to what extent similar students exposed to different teachers, all else equal, end up having different outcomes. Ideally, one would like to perform an experiment where teachers are randomly allocated across students, while the researcher observes the achievement of students after being exposed to the teacher. However, for obvious reasons, such an experiment is hard to conduct in practice, so studies of the relationship between student outcomes and teacher quality needs to be based on non-experimental data.

We consider two approaches obtaining credible estimates of teacher quality in non-experimental settings. The first approach (the value added approach) requires data for outcome variables (typically test scores or exam results) available for the same student at different points in time. The second approach (the between subject approach) is relevant in cases where test scores are available at a single point in time on several subjects taught by different teachers.

The value added approach

Following [Koedel et al. \(2015\)](#), a tractable and estimable version of the general formulation of the education production function in (1) is the following linear value added model

$$y_{isj,t} = \beta_0 + \beta_1 y_{i,t-1} + \beta_2 X_{is,t} + \beta_4 T_{isj,t} + \varepsilon_{it} \quad 22.2$$

$y_{isj,t}$ is the achievement of student i exposed to teacher j in school s at time (grade) t . Here, $X_{is,t}$, $S_{is,t}$ and $T_{isj,t}$ represents levels of the

individual, family and school inputs and the quality of teacher j experienced by individual i in school s at time (grade) t . This value added model assumes that the inclusion of a lag, $y_{i,t-1}$ in the achievement variable captures all cumulative (dynamic) features of the educational production process. This is consistent with a restriction that the effects of all inputs evolve in a common geometrically declining pattern.

On the other hand, this is a very data demanding strategy as it requires longitudinal data linking individual students to individual teachers within schools and across grades. This type of data has typically become available in some US states. But in many countries, linked student-teacher data are not available or unreliable and hence assessments of teacher quality must still be based on other approaches.

By exploiting the fact that a student has been exposed to different teachers throughout his/her school career, total teacher effects can be identified, holding observed and unobserved school, and to some extent also family and individual variables, constant. One strength of this approach is the possibility to quantify the potential gains from increasing teacher quality due to both observed and unobserved characteristics of the teacher.

In the value added framework, the approach is to use indicator variables for each teacher the students have been exposed to during the period under investigation. This removes sorting of students across schools and districts as a source of bias. Moreover, if students are followed and tested at different grades and exposed to different teachers over time, the inclusion of fixed student effects also effectively removes potential bias from omitted time-invariant individual and family characteristics, such as innate ability. An early study using this approach was [Hanushek \(1971\)](#), using test scores at the end of third grade and controlling for second grade test scores and a range of control variables. Recent studies, most prominently [Chetty et al. \(2014a,b\)](#) have further developed and refined

the value added approach by exploiting data over cohorts and grades in schools in New York City, and the literature expands rapidly. Recent reviews of the literature can be found in [Jackson, Rockoff, and Staiger \(2014\)](#) and [Koedel et al. \(2015\)](#).

A crucial first question is to what extent these value-added based estimates represents unbiased estimates of the true teacher quality. Or put differently: Do the value added estimates adequately account for the effect teachers have on student's knowledge or do they to some extent reflect unobserved student characteristics or school or teacher behavior not related to true knowledge generation?

A first challenge is that the value added approach may not account sufficiently for endogenous assignment of teachers to classes within schools, as suggested by [Rothstein \(2010\)](#) and followed up in [Rothstein \(2017\)](#).¹ [Kane and Staiger \(2008\)](#) and [Kane, McCaffrey, Miller, and Staiger \(2013\)](#) validates the value added approach by exploiting random assignment of students to teachers in field experiments. They find that the estimated teacher effects based on historical non-experimental data and the value added approach are unbiased predictors of student achievement under random assignment. Although [Chetty et al. \(2014 a\)](#) find that less advantaged students tend to be assigned to teachers with lower value added,

they find that controlling for test scores in previous grades and a small set of other student characteristics is sufficient to account for the bias due to this sorting. They reach this conclusion using two approaches. Their first approach is to estimate bias based on the degree of selection on observable characteristics excluded from the value added model. Their second approach is to exploit changes in teacher assignments across adjacent grades due to teacher turnover in order to assess the bias due to unobservables. Using this quasi-experimental approach, they find that the entry (exit) of a high value added teacher in a cohort induces a higher (lower) than average end-of-year test score for this cohort. They conclude that the degree of bias in the value added measures due to endogenous assignment of teachers within schools is low. [Bacher-Hicks, Kane, and Staiger \(2014\)](#) replicates their analysis using data from Los Angeles. They confirm the [Chetty et al. \(2014\)](#) conclusion that value added estimates are fairly unbiased estimates of teacher effects.²

However, whether assignment of teachers across classes within school after conditioning on past performance and a small set of controls can be viewed as fairly random or systematically related to student characteristics is still a controversial issue. The problem is likely to vary across schools, school districts and countries, see also [Dieterle, Guarino, Reckase, and](#)

¹ An important element in Rothstein's critique of the value added approach is the use of a "placebo" or falsification test. The argument goes like this: If teacher value added represents an unbiased measure of teacher quality, teacher value added in later grades should not be able to predict value added in earlier grades. However, he finds that this is not the case using data from North Carolina schools. [Chetty, Friedman, and Rockoff \(2017\)](#) argue theoretically that the "placebo" test used by [Rothstein \(2017\)](#) is not a valid approach. Put simply, their argument is that prior test scores cannot be used to conduct placebo test because value added measures are themselves estimated from prior test scores and gives a mechanical relationship between prior and current value added measures. See also [Goldhaber and Chaplin \(2015\)](#) for discussion and evidence on the relevance of the Rothstein test.

² [Bacher-Hicks et al. \(2014\)](#) also study to what extent teacher effectiveness vary with student characteristics and find that teacher value added forecast the causal impact on student achievement independently of whether teachers move across schools that differ substantially in terms of student composition. However, they also find that disadvantaged students by race and parental background are disproportionately matched with low value added teachers. This contribute to widening achievement gaps between groups and may have adverse distributional effects.

Wooldridge (2014) for a recent study of the assignment of students to teachers within schools.

A second challenge for the value added approach is whether estimated teacher quality only captures short term effects, for example some sort of “teaching to the test”-effects. Is estimated teacher quality able to predict student performance in the education and labor markets in the future? To investigate this issue, Chetty, Friedman, and Rockoff (2014b) compare outcomes at later points in life between students assigned to high and low value added teachers. They find that students of higher value added teachers have both higher earnings and are more likely to attend college than students of lower value added teachers.

What do the value added studies tell us about the numerical contribution of teacher quality to student outcomes? According to the review article by Koedel et al. (2015), the variation in teacher quality is substantial and much of the variation occurs within schools. Hanushek and Rivkin (2012) use several representative estimates of the labor market returns to cognitive test scores as well as a number of estimates of total teacher effects from value added studies to assess the importance of teachers on earnings. They estimate that the economic value of a teacher who is one standard deviation above the mean is \$420 000 for a class of 20 students. Chetty et al. (2014b) suggest that a hypothetical replacement of an average teacher with one in the top 5% will increase students’ lifetime income by around \$1.4 million per classroom taught. These are large gains compared to possible costs connected with such a replacement of teachers. They illustrate that policies that are able to increase teacher quality have a large potential to increase labor productivity.

The between subject approach

A between subject approach has been proposed when the data includes information on several test scores at the same point in time, but from classes with different teachers for the same student. This format allows the researcher to account for student and school fixed effects. Removing the student fixed effect by measuring all variables as deviations from subject specific means (denoted by a bar over the variable), effectively also absorbing all school level effects. Equations of the following format are estimated

$$y_{ij} - \bar{y}_j = \beta(T_{ij} - \bar{T}_j) + (\varepsilon_{ij} - \bar{\varepsilon}_j) \quad 22.3$$

where j denote subject. Clotfelter, Ladd, and Vigdor (2010) notice that while the concern in value added models is that students’ assignment to classrooms (teachers) can be correlated with time varying unobserved determinants of student achievement, the concern in the between subject approach is that assignment of students to classrooms (teachers) is correlated with unobserved student characteristics that differs across subjects.³

Formally, the identification assumption in the between subject approach is that the within student by subject error term, ε_{ij} , is uncorrelated with teacher characteristics T_{ij} . Thus, researchers using this strategy need to argue that this assumption is valid in their specific context. For instance, Clotfelter et al. (2010) try to assess the credibility of the assumption in their analysis of teacher effects in US high schools by measuring to what extent students is assigned to different classes by subject based on their relative ability across subjects. They conclude that the matching of students to classes (teachers) is to a large extent based on general ability and not relative ability across subjects. However, this conclusion is of course limited to their

³ Another difference between the value added and between subject approach is that the latter does not account for the cumulative feature of the learning process as current achievement is only related to current teacher characteristics.

sample and specific institutional context and need not apply in other settings.⁴ Another challenge with the between subject approach is that other variables important for student achievement such as peer groups is also likely to vary across subjects. Although control variables meant to account for across subject differences in peer groups can sometimes be included, teacher effects may to some extent reflect unobserved peer characteristics.

What explains teacher quality variation?

While a lot of research effort in the last 10–15 years has been directed toward the estimation of teacher quality based on the value added approach and the reliability of these quality measures, an increasing amount of studies also investigate the relationship between teacher quality and individual teacher characteristics and teacher labor market conditions.

There are two approaches to investigating factors related to teacher quality. The two-step approach relates estimated overall teacher quality, T_i , to teacher characteristics, C_i , in the following way

$$T_i = \alpha C_i + \varepsilon_i$$

The alternative approach, which is more commonly used, is to estimate the effect of teacher characteristics directly by including teacher characteristics in the education production function 21.2 and 21.3 instead of indicator

variables for teachers. One challenge to this approach is potential endogenous sorting of teachers and students across schools and classrooms within schools.⁵ The possibility that teacher wages and teacher characteristics are jointly determined in the teacher labor market is a further source of endogeneity. Studies using this approach typically include school fixed effects in the model in order to control for time-invariant school or district level variables, in addition to control variables to account for relevant within school or within school district variation. Examples of studies in this tradition are Loeb and Page (2000), Harris and Sass (2011) and Britton and Propper (2016).

Individual teacher characteristics

Teacher experience

According to the review of the earlier research in Palayo and Rees (2010), teacher productivity in terms of student achievement gains improves in the first few years of the career but stops improving thereafter. Using data from different states and school districts in the US, Papay and Kraft (2015), Wiswall (2013) and Harris and Sass (2011) indicate that earlier findings were partly a result of restrictive model assumptions, failure to take account of endogenous teacher attrition, and dynamic matching of better students to teachers with higher productivity as teachers gain experience. They find that improvement in teacher quality continues far longer into the

⁴ Gronqvist and Vlachos (2016) use cross section data on students' achievement in Swedish compulsory schools to estimate teacher effects and argue that since formal tracking is forbidden in Sweden, the between-subjects strategy is well suited there. Their evidence also suggest that the matching of teachers and students is unrelated to variables characterizing teachers' cognitive ability, while they cannot completely rule out the possibility of systematic relationship between matching and teachers non-cognitive ability. Slater, Davies, and Burgess (2012) use the between subject approach to obtain measures of variation in teacher effectiveness in England using matched data for students, schools and teachers from 33 schools. They also include prior subject specific student performance.

⁵ Hanushek, Kane, and Rivkin (2004), Clotfelter, Ladd, and Vigdor (2006) and Dieterle et al. (2014) contain US evidence on teacher-student matching across school districts, schools and within schools. Falch and Strøm (2005) and Bonesrønning et al. (2005) provide evidence on teacher sorting across schools and school districts in Norway.

career than the earlier results indicate. On-the-job training throughout much of the career seems to improve teacher quality.

Teacher skills, formal training and education

Earlier research reviewed in Palayo and Rees (2010) emphasize that teacher test scores and attending selective educational institutions positively affect student performance in the US, while formal teacher credentials in terms of advanced degrees and certification does not. Recent studies have extended this research by looking at a broader set of teacher skills, exploring data from other countries, and using more credible estimation strategies.

Rockoff, Jacob, Kane, and Staiger (2011) conduct a survey of new math teachers in New York City in order to measure the association between a set of observable teacher characteristics and teacher quality using the two-step approach. While relying on a small sample, their investigation indicates that a much broader set of observational credentials than those traditionally used when recruiting teachers are positively associated with student achievement gains.

Harris and Sass (2011) studies the relationship between teacher education and training using the value added format and exploits variation over time within teachers in these variables to assess the contribution to student achievement using data from Florida public schools while conditioning on teacher, school and student fixed effects. While their general results support earlier research findings that formal teacher training and graduate coursework did not affect student outcomes, some heterogeneity exist across subjects and grade levels. Moreover, it is an open question to what extent the results applies outside the Florida school system and in

other institutional contexts. A recent paper by Bastian (2019) using data from North Carolina gives a more comprehensive and nuanced picture of the effectiveness of teachers with graduate degrees.

Clotfelter et al. (2010) use the between subject approach for high schools in North Carolina. They find that subject specific teacher credentials (licensure and certification) in math and English language increase student achievement. According to their results, the achievement gain of having a teacher with strong teacher credentials compared to a teacher with weak teacher credentials is 0.23 standard deviations. Moreover, they judge that the teacher credentials account for more than one-fifth of the overall distribution of teacher quality in their sample. While they also find that teacher test scores are predictive of student test scores, the effect of licensure and certification variables are more important than teacher test scores. However, it is also in this case a question of the external validity to other contexts. Slater et al. (2012) find for English schools that teacher credentials are not associated with their measures of total teacher quality based on the between subject approach.

Gronqvist and Vlachos (2016) use matched student-teacher register data from upper secondary schools in Sweden and the between subject approach to study the effect of teacher cognitive and noncognitive skills.⁶ Their conclusion is that teacher skills have negligible effects on average student achievement, but that the average effect hides important heterogeneities. First, cognitive and non-cognitive teacher skills seems to be complementary, i.e., students having a teacher with high cognitive skills and high social interaction skills increase student achievement, while having a teacher with low social interaction skills

⁶ Teacher skills are measured by three measures: From military drafts they have measures of cognitive skills and social interaction skills for Swedish men. In addition, for all teachers GPA from upper secondary school which may be associated with both cognitive and noncognitive skills are available. All teacher skill measures are obtained before teachers started tertiary education.

and high cognitive skills decrease achievement. Second, while increased cognitive skills seems to increase the achievement gap between high and low-achieving students, increased social interaction skills decrease the gap.

A recent paper by [Hanushek, Piopiunik, and Widerhold \(2019\)](#) investigates the relationship between student achievement and teacher cognitive skills across 31 countries using a between subject approach and test scores in math and reading from the PISA surveys. To measure country differences in teacher skills, they exploit information from adult test scores in numeracy and literacy from the PIAAC survey. They show that relative teacher skills vary substantially across countries. The results suggest that student achievement is strongly associated with teacher cognitive skills. The results imply that one standard deviation increase in teacher skills increases student achievement with 0.1–0.15 standard deviations. One limitation with the study is that teacher relative skills across subjects are only measured at the country level. Although they show that results are robust to the inclusion of a range of observable country specific variables and placebo tests, the results might to some extent reflect unobserved country differences in subject specific skills rather than teacher skills.

Teacher labor market variables

If teacher wages were determined in a well-informed competitive market, teacher quality and teacher wages would be jointly determined. High-quality teachers would have higher wage than low-quality teachers. Moreover, exogenous negative shifts in teacher supply, for instance due to improved outside options, would increase teacher wages to equilibrate the market.

However, actual teacher labor markets are very far from this stylized free market model as emphasized in [Lankford and Wyckoff \(2010\)](#). Instead, teacher wages are to a large extent determined within rigid wage scales, often within collectively bargained contracts with teacher trade unions. Since teacher quality is important for student performance, understanding how structural changes in the teacher labor market affect teacher quality is important in order to design policies to retain and recruit effective individuals in the teaching profession. This section reviews some recent empirical evidence of the effects of teacher wages, wage setting systems, and other labor market characteristics on teacher quality.

Teacher wages and teacher wage systems

While earlier research from the US, as reviewed in [Hanushek and Rivkin \(2006\)](#), suggests that teacher quality is not systematically associated with teacher wages, recent evidence from other countries and other institutional contexts paints a more nuanced picture. While several recent studies find that teacher wages and wages in alternative jobs have causal effects on teacher supply, [Britton and Propper \(2016\)](#) takes this line of research a step further by assessing the effect of relative teacher wages on student outcomes in English schools.⁷ They exploit the fact that the centralized and regulated pay system in English schools implies large regional variation in the pay gap between teacher jobs and alternative jobs in the private sector. Such a gap may affect the recruitment of teachers to the school as well as the productivity of current teachers as suggested by efficiency wage theories. They find that a ten percent negative shock to the wage gap implies a 2% loss in

⁷ [Falch \(2011\)](#) find that an exogenous pay increase in hard to staff schools in Northern Norway increase supply of teachers to these schools. [Leigh \(2012\)](#) find that increased teacher pay boosts the average aptitude of students entering teacher education courses.

average school performance on key exams taken at the end of compulsory education.

A recent paper by [Willén \(2018\)](#) provides a more direct study of the effect of different pay setting systems. He exploits that the pay determination system of teachers in Sweden changed dramatically in 1996 from a uniform national pay schedule to a system with pay determined at the individual level. Exploiting pre-reform variation in regional non-teacher pay, he finds a long run elasticity of teacher pay with respect to pre-reform non-teacher pay of 0.2. The reform implied an increase in entry wages and a flattening of the age-wage relationship. However, he finds no effect on this change on teacher composition, nor on student performance.

[Hanushek et al. \(2019\)](#) study whether their estimated effects of cognitive teacher skills on student performance across countries is associated with cross country differences in factors related to teacher pay. Their results suggest that observed cognitive teacher skill differences across countries are positively associated with differences in women's access to high skilled non-teacher occupations and to salary premiums for teachers.⁸

Teacher unions

In most countries, teacher unions affect teacher wage setting and, in addition, the operation of schools through, e.g., recruitment and layoff policies. While there is agreement that teacher unions are important actors in the teacher labor market, the theoretical effects of unions on teacher productivity is ambiguous due to the conflicting effects of pure rent-seeking activities and the potential efficiency gain through collective voice. Credible empirical

evidence on the actual effect of teacher unions on student outcomes is limited. Earlier studies look at the effects of teacher collective bargaining on relatively short-run measures of student outcomes. [Hoxby \(1996\)](#) finds that unionization increases high school dropout, while [Lovenheim \(2009\)](#) finds no effect. A recent paper by [Loevenheim and Willén \(2019\)](#) estimates the effects of teacher collective bargaining laws on long run student outcomes as labor market performance. Their identification strategy exploits the staggered introduction of duty-to-bargain laws across cohorts within states. They find that exposure to collective bargaining laws reduces future labor market success in terms of earnings and labor market participation, especially among black and Hispanic men. The effects are not evident for women. Moreover, they also find that collective bargaining laws are associated with lower non-cognitive skills among young men.

A recent paper by [Baron \(2018\)](#) reaches quite different conclusions. He exploits that the state of Wisconsin limited teacher bargaining power in 2011 by imposing legal restrictions on unions funding and bargaining ability in order to reduce the state budget deficit in the aftermath of the Great Recession. To identify effects on student outcomes he exploits variation in the timing of the decrease in teacher bargaining power due to variation in the expiration dates of existing union contracts across school districts in the state. He finds that the reduction in teacher union power reduced student achievement, in particular among students in the lower part of the achievement distribution. He argues that the reduction in achievement was due to a reduction in teacher salaries combined with an increase in teacher turnover. His results may

⁸ This is consistent with the evidence in [Bacolod \(2007\)](#) indicating a marked decline in the quality of young women going into teaching between 1960 and 1990 in the US as measured by standardized test scores and undergraduate institution selectivity together with an increase in the non-teacher alternatives available for females. [Dolton and Marcenaro-Gutierrez \(2011\)](#) also conclude that teacher pay differences contributes significantly to explain differences in cross-country student performance.

indicate that the effect of changes in teacher union power on student outcomes is not symmetric. Further analysis is needed to conclude whether this conclusion extends to student outcomes in the longer run.

As a result of the reform in Wisconsin, some school districts replaced the old seniority-pay system with a flexible-pay system. [Biasi \(2018\)](#) finds that high-quality teachers, measured using a value-added approach, moves away from school districts with a seniority-pay system to districts with a flexible-pay system, increasing average teacher quality in the latter school districts. To gauge the general effects of such moves she performs a simulation exercise based on a structural model of the teacher labor market. The simulations show that the implementation of individual salary models in all school district generates smaller improvement in teacher quality entirely connected to the exit of low quality teachers.

Teacher mobility

Since the evidence clearly shows that teacher quality is important for student outcomes, knowledge about the link between teacher mobility and teacher quality is crucial in order to understand the determination of overall teacher quality as well as quality distribution across schools. Teachers' decisions about job changes is a choice among different job alternatives that maximizes individual utility. This choice will depend on relative wages, as well as nonpecuniary job conditions and teacher characteristics. Previous studies have shown that teachers are more likely to quit schools with large shares of minority and disadvantaged students.⁹

However, an important and difficult issue is to gain empirical knowledge of the change in overall teacher quality and the quality distribution across schools and students as teachers

exit the public school sector or move across schools. [Feng and Sass \(2017\)](#) tries to answer this question using matched school, teacher and student data from North Carolina and Florida. Using the two-step approach, they estimate the effect of both own quality as well as the quality of their peers on the teachers' decisions to move to another school within or across school districts or leave the public school sector. They do not find strong relationships between teacher quality and intra or interdistrict teacher mobility. Regarding exit from the public school sector, they find that both the most effective teachers and the least effective teachers are most likely to exit while teachers in the middle of the quality distribution are most likely to stay. Moreover, the quality of ones' peer teachers are important for the mobility decision. Top quality teachers are more likely to stay if surrounded by high quality peers and teachers tend to move to schools where the average teacher quality is roughly similar to their own. The broad conclusion is that teacher mobility tend to exacerbate quality differences across schools.

Use of teacher quality measures in pay and evaluation systems

The evidence suggest that teacher quality are only weakly related to specific teacher characteristics traditionally included in pay schedules (formal education, experience and teacher credentials). Thus, it would be tempting to replace traditional pay schedules with individual pay setting based on teacher quality measures obtained from value added models. There is also some evidence that performance pay systems contribute positively to student outcomes ([Lavy \(2009\)](#)), although the literature does not provide definitive evidence in support of performance pay.

⁹ See [Hanushek et al. \(2004\)](#) for evidence from the US and [Falch and Strøm \(2005\)](#) for evidence from Norway.

First, practical problems arise when trying to implement value added based evaluation systems. Evaluation systems must have legitimacy inside the institution, i.e., among the teachers. Research have demonstrated the importance of teachers, but even unbiased value added estimates are measured with error. For example, the precision depends on the number of students assigned to the teacher. Ignoring measurement error may imply that two teachers with the same true quality will be treated differently just because they have different number of students. How to handle measurement error and other practical implementation problems is not trivial in this context as emphasized by [Ballou and Springer \(2015\)](#). Implementation issues are also discussed in [Staiger and Kane \(2014\)](#).

Second, when value added based teacher quality estimates become high stake, the incentives for teachers to teach to the test, manipulate assignment of students to teachers and outright cheat or manipulate tests becomes apparent. Earlier research has shown that test based school accountability systems are vulnerable to manipulation or cheating by the actors involved, see [Figlio and Winicki \(2005\)](#) and [Jacob \(2005\)](#) for evidence and examples. Arguably, such behavior will be exaggerated when implemented at the individual teacher level, although [Lavy \(2009\)](#) finds that introduction of a teacher performance pay system in Israel did not lead to teacher manipulation of test scores.

Third, the quality of a single teacher is likely to depend on the quality of his/her peers as shown in [Jackson and Bruegmann \(2009\)](#). Based on the value added approach, they show that a single teachers' students has larger achievement gains in math and reading if the teacher have effective colleagues. This makes it difficult to isolate the single teacher's contribution to student achievement and is rather an argument for group based performance pay systems,

although such systems face serious problems with free-riding. [Goodman and Turner \(2013\)](#) present evidence that a group based bonus program to teachers in New York City schools had no effects on student achievement in general, although small positive effects on math achievement were detected when the program contained small groups of teachers.

A related problem with regard to individual teacher pay systems based on teacher performance is that large pay differences themselves may reduce effort and productivity as suggested by the fair-wage mechanism in efficiency wage models introduced by [Akerlof and Yellen \(1990\)](#). Recent experimental evidence indicates that pay differences which are not regarded as fair may reduce effort, see [Cohn, Fehr, and Goette \(2015\)](#) and [Breza, Kaur, and Shamdasani \(2018\)](#)¹⁰.

Summary and the way forward

There are large differences in teacher quality. This is demonstrated in a large number of studies, using impressive data-sets and sophisticated empirical methods. The value-added approach is most common, but the between-subject approach is a fruitful alternative because it relies on different assumptions on the underlying education production function and is less demanding with regard to data.

One important question is how this insight can be used in policy. In a recruitment situation, estimates of teacher quality will typically be unobservable and the principal must rely on other types of information. However, teacher quality seems only weakly related to observable teacher characteristics. There is firm evidence that teacher quality is positively associated with the experience of the teacher, the cognitive skill of the teacher and some teacher credentials.

¹⁰ [Ashraf and Bandiera \(2018\)](#) contains a broader discussion of the role of social incentives in organizations.

However, the findings in this literature seems to depend on the specific context of the data.

One might think that estimated teacher quality can be used in an incentive pay system. However, the design of pay and teacher evaluation systems based on estimated teacher quality face important practical implementation problems and a risk of behavior that is gaming the system. Careful empirical research taking implementation and institutional details into account is needed to conclude whether such systems will enhance teacher productivity and student long run outcomes.

A fruitful future avenue of research is to explore what we may denote the teacher “production function”, which can be loosely described as

$$T_j = f(C_j, M_j, \varepsilon_j)$$

where T is teacher quality, C is teacher characteristics, M is teaching methods and ε is the error term. Some experimental studies identify effective teaching methods. Such findings have the potential to increase teacher quality along the whole distribution of teacher quality. Arguably, the potential is largest at the low end of teacher quality distribution, because high-quality teachers presumably use effective teaching methods already. An early study on teaching methods is [Machin and McNally \(2008\)](#). They find that the introduction of a “literacy hour” in English schools increased student achievement. Another more recent example is [Jerrim and Vignoles \(2016\)](#) who study the effect of introducing Singaporean inspired ‘mastery’ approach to teaching mathematics into schools in England. They find a small positive effect of the change in teaching method. A third example is [Machin, McNally, and Viarengo \(2018\)](#) studying the effect of an intervention that changes how teachers teach young children to read represented by the staggered introduction of “synthetic phonics” in English primary schools. They provide evidence that the intervention had a positive causal effect

on young childrens’ literacy and especially among children with disadvantaged background and non-native speaking English background.

References

- Akerlof, G. A., & Yellen, J. L. (1990). The fair wage-effort hypothesis and unemployment. *Quarterly Journal of Economics*, 105, 255–283.
- Ashraf, N., & Bandiera, O. (2018). Social incentives in organizations. *Annual Review of Economics*, 10, 439–463.
- Bacher-Hicks, A., Kane, T. J., & Staiger, D. O. (2014). *Validating teacher effect estimates using changes in teacher assignments in Los Angeles*. NBER Working Paper No. 20657.
- Bacolod, M. P. (2007). Do alternative opportunities matter? The role of female labor markets in the decline of teacher quality. *Review of Economics and Statistics*, 89, 737–751.
- Ballou, D., & Springer, M. G. (2015). Using student test scores to measure teacher performance: Some problems in the design and implementation of evaluation systems. *Educational Researcher*, 44, 77–86.
- Baron, E. J. (2018). The effect of teachers’ unions on student achievement: Evidence from Wisconsin’s act 10. *Economics of Education Review*, 67, 40–57.
- Bastian, K. C. (2019). A degree above? The value-added estimates and evaluation ratings of teachers with a graduate degree. *Education Finance and Policy*, 14, 652–678.
- Ben-Porath, Y. (1967). The production of human capital and the life-cycle of earnings. *Journal of Political Economy*, 75, 352–365.
- Biasi, B. (2018). *The labor market for teachers under different pay schemes*. NBER Working Paper 24813.
- Bonesrønning, H., Falch, T., & Strøm, B. (2005). Teacher sorting, teacher quality, and student composition. *European Economic Review*, 49, 457–483.
- Breza, E., Kaur, S., & Shamdasani, Y. (2018). The morale effects of pay inequality. *Quarterly Journal of Economics*, 133, 611–663.
- Britton, J., & Propper, C. (2016). Teacher pay and school productivity. *Journal of Public Economics*, 133, 75–89.
- Chetty, R., Friedman, J. N., & Rockoff, J. N. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104, 2593–2632.
- Chetty, R., Friedman, J. N., & Rockoff, J. N. (2014b). Measuring the impacts of teachers II: Teacher value added and student outcomes in adulthood. *American Economic Review*, 104, 2633–2679.
- Chetty, R., Friedman, J. N., & Rockoff, J. N. (2017). Measuring the impacts of teachers: Reply. *American Economic Review*, 107, 1685–1717.

- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, 41, 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, 655–681.
- Cohn, A., Fehr, E., & Goette, L. (2015). Fair wages and effort provision: Combining evidence from a choice experiment and a field experiment. *Management Science*, 61, 1777–1794.
- Dieterle, S. G., Guarino, C. M., Reckase, M. M., & Wooldridge, J. M. (2014). How do principals assign students to teachers? Finding evidence in administrative data and the implications for value-added. *Journal of Policy Analysis and Management*, 34, 32–58.
- Dolton, P., & Marcenaro-Gutierrez, O. D. (2011). If you pay peanuts do you get monkeys? A cross-country analysis of teacher pay and performance. *Economic Policy*, 26, 5–55.
- Falch, T. (2011). Teacher mobility responses to wage changes: Evidence from a quasi-natural experiment. *American Economic Review*, 101, 460–465.
- Falch, T., & Strøm, B. (2005). Teacher turnover and non-pecuniary factors. *Economics of Education Review*, 24, 611–631.
- Feng, L., & Sass, T. R. (2017). Teacher quality and teacher mobility. *Education Finance and Policy*, 12, 396–418.
- Figlio, D., & Winicki, J. (2005). Food for thought? The effects of school accountability plans on school nutrition. *Journal of Public Economics*, 89, 381–394.
- Goldhaber, D., & Chaplin, D. D. (2015). Assessing the "Rothstein falsification test": Does it really show teacher value-added models are biased? *Journal of Research on Educational Effectiveness*, 8, 8–34.
- Goodman, S. F., & Turner, L. J. (2013). The design of teacher incentive pay and educational outcomes: Evidence from the New York city bonus program. *Journal of Labor Economics*, 31, 409–420.
- Gronqvist, E., & Vlachos, J. (2016). One size fits all? The effects of teachers' cognitive and social abilities on student achievement. *Labour Economics*, 42, 138–150.
- Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *American Economic Review*, 61, 280–288.
- Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources*, 14, 351–388.
- Hanushek, E. A., Kane, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of Human Resources*, 39, 326–354.
- Hanushek, E. A., Piopiunik, M., & Widerhold, S. (2019). The value of smarter teachers: International evidence on teacher skills and student performance. *Journal of Human Resources*, 54, 857–899.
- Hanushek, E. A., & Rivkin, S. G. (2006). Teacher quality. In E. A. Hanushek, & F. Welch (Eds.), *Handbook of the Economics of Education* (volume 2, pp. 1051–1078). North Holland: Elsevier.
- Hanushek, E. A., & Rivkin, S. G. (2012). The distribution of teacher quality and implications for policy. *Annual Review of Economics*, 4, 131–157.
- Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, 95, 798–812.
- Hoxby, C. M. (1996). How teachers' unions affect education production. *Quarterly Journal of Economics*, 111, 671–718.
- Jackson, C. K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, 1, 85–108.
- Jackson, C. K., Rockoff, J. E., & Staiger, D. A. (2014). Teacher effects and teacher-related policies. *Annual Review of Economics*, 6, 801–825.
- Jacob, B. (2005). Testing, accountability, and incentives: The impact of high-stakes testing in Chicago public schools. *Journal of Public Economics*, 89, 761–796.
- Jerrim, J., & Vignoles, A. (2016). The link between East Asian 'mastery' teaching methods and English children's mathematics skills. *Economics of Education Review*, 50, 29–44.
- Kane, J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment*. Seattle, WA: Bill & Melinda Gates Foundation.
- Kane, T. J., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation*. NBER Working Paper No.14607.
- Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review*, 47, 180–195.
- Lankford, H., & Wyckoff, J. (2010). Teacher labor markets: An overview. In D. J. Brewer, & P. J. McEwan (Eds.), *Economics of education*. Elsevier. Academic Press.
- Lavy, (2009). Performance pay and teachers' effort, productivity, and grading ethics. *American Economic Review*, 99, 1979–201.
- Leigh, A. (2012). Teacher pay and teacher aptitude. *Economics of Education Review*, 31, 41–53.
- Loeb, S., & Page, M. E. (2000). Examining the link between teacher wages and student outcomes: The importance of alternative labor market opportunities and non-pecuniary variation. *Review of Economics and Statistics*, 82, 393–408.
- Lovenheim, M. F. (2009). The effect of teachers' unions on education production: Evidence from union election certifications in three midwestern states. *Journal of Labor Economics*, 27, 525–587.

III. Production, costs and financing

- Loevenheim, M. F., & Willén, A. (2019). The long-run effects of teacher collective bargaining. *American Economic Journal: Economic Policy*, 11, 292–324.
- Machin, S., & McNally, S. (2008). The literacy hour. *Journal of Public Economics*, 92, 1441–1462.
- Machin, S., McNally, S., & Viarengo, M. (2018). Changing how literacy is taught: Evidence on synthetic phonics. *American Economic Journal: Economic Policy*, 10, 217–241.
- Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market. Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, 130, 105–119.
- Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can we recognize an effective teacher when we recruit one? *Education Finance and Policy*, 6, 43–74.
- Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics*, 125, 175–214.
- Rothstein, J. (2017). Measuring the impacts of teachers: Comment. *American Economic Review*, 107, 1656–1684.
- Slater, H., Davies, N. M., & Burgess, S. (2012). Do teachers matter? Measuring the variation in teacher effectiveness in England. *Oxford Bulletin of Economics and Statistics*, 74, 629–645.
- Staiger, D. O., & Kane, T. J. (2014). The relationship between annual student achievement gains and a teacher's career value added. In T. J. Kane, K. A. Kerr, & R. C. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the measures of effective teaching project*. San Francisco, CA: John Wiley & Sons Inc.
- Willén, A. (2018). *From a fixed national pay scale to individual wage bargaining: The labor market effects of wage decentralization*. Working Paper.
- Wiswall, M. (2013). The dynamics of teacher quality. *Journal of Public Economics*, 100, 61–78.