

# Understanding teacher effectiveness to raise pupil attainment

Teacher effectiveness has a dramatic effect on student outcomes—how can it be increased?

Keywords: teacher effectiveness, teacher selection, teacher training

## ELEVATOR PITCH

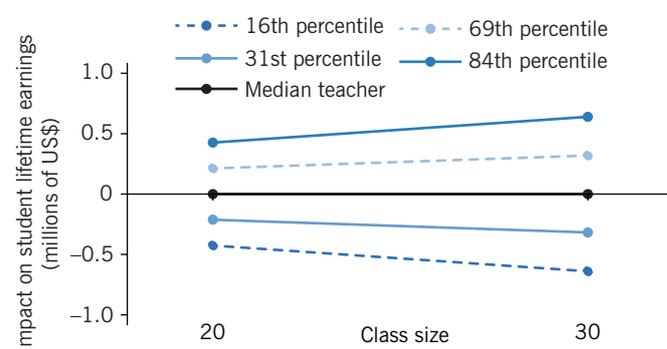
Teacher effectiveness is the most important component of the education process within schools for pupil attainment. One estimate suggests that, in the US, replacing the least effective 8% of teachers with average teachers has a present value of \$100 trillion. Researchers have a reasonable understanding of how to measure teacher effectiveness; but the next step, understanding the best ways to raise it, is where the research frontier now lies. Two areas in particular appear to hold the greatest promise: reforming hiring practices and contracts, and reforming teacher training and development.

## KEY FINDINGS

### Pros

- ⊕ Pupils taught by highly effective teachers get significantly higher grades; the effect is substantial and enduring.
- ⊕ Teacher effectiveness improves long-term outcomes such as earnings.
- ⊕ There are robust and persistent measures of teacher effectiveness that are supported by expert observation and pupil feedback.
- ⊕ Standard estimates of teacher effectiveness seem reliable and not to suffer from bias related to pupil selection.

### Teacher effectiveness has a huge impact on student earnings



Source: Based on data in [1].

IZA  
World of Labor

### Cons

- ⊖ Research shows that teacher effectiveness is largely uncorrelated with the teacher's own educational qualifications.
- ⊖ Teacher selection and hiring can be problematic because there is little useful information available pre-hire.
- ⊖ Some studies suggest that schools would benefit from a high optimal level of turnover among junior teachers, though recent research is re-opening the debate on the role of experience.
- ⊖ Tracking the persistent effects of training, mentoring, and development is difficult due to a general lack of longitudinal data.

## AUTHOR'S MAIN MESSAGE

A number of studies from different countries have produced similar estimates of the impact of teacher effectiveness. These estimates have been shown to be robust and are supported by studies using experimental assignment of teachers to classes. The results show that variations in teacher effectiveness are extremely important in understanding pupils' attainment. Studies of optimal contract structure for teachers show that probationary periods should be much longer than is common in the US and the UK. Additionally, informal learning and mentoring represent potentially very useful alternative routes for improving average teacher effectiveness.

## MOTIVATION

Recent research on the economics of teachers and teaching has shown that this is an area of great policy promise for raising student achievement. The author of an influential 2011 study states the case rather directly: “No other attribute of schools comes close to having this much influence on student achievement” [1], p. 467. Understanding the meaning and role of teacher effectiveness offers policymakers new opportunities to realize their education objectives. Such policies might include reforming the ways in which teachers are hired, paid, retained, and promoted—in other words, reforming the whole nature of the teacher contract. Policies might also include changing the ways in which teachers are trained and how they continue to learn throughout their professional careers.

The importance of teacher effectiveness has been vividly illustrated by several important studies [1], [2]. One of the most striking results is that replacing the lowest performing 5–10% of teachers with average teachers would deliver extremely large net present value calculations. Each effective teacher raises the lifetime earnings of a huge number of pupils over their career; the above-mentioned 2011 study suggests that using standard estimates “replacing the bottom 5–8 percent of teachers with average teachers [would have] ... a present value of \$100 trillion” [1]. The second study, from 2014, similarly finds that replacing the 5% least effective teachers with average teachers would yield around \$9,000 per classroom per year in future pupil earnings due to better education [2].

Economists define teacher effectiveness (or teacher “quality”) precisely but narrowly. It is based on the progress in academic achievement that a pupil makes over their time with the teacher, typically measured by standardized tests at the end of their time with a given teacher (and ideally at the beginning too). A teacher’s effectiveness is the average of their pupils’ progress measured across all the pupils they teach. Early studies on this topic were interested primarily in the overall distribution of teacher effectiveness, for example, the number of high- or low-effectiveness teachers, but later work has become focused on the effectiveness of individual teachers, the source of differences in effectiveness, and the role of performance management and reward. More recent work has broadened to include their impact on non-cognitive outcomes as well as test scores.

## DISCUSSION OF PROS AND CONS

### The value of measuring teacher effectiveness

Modern analysis of teacher effectiveness only began about 15 years ago, facilitated by the availability of administrative data. Given the definition of teacher effectiveness, the key to measuring it is having access to a class list—that is, which pupils were taught by which teacher—and this is typically only available in administrative data.

The metric that economists typically use to gauge effectiveness is based on pupil test scores: as teacher effectiveness explains a greater proportion of the variation in test scores, clearly the more important metric is teacher effectiveness. This is expressed as fractions of the standard deviation (a measure of the variation) in pupil test scores that effectiveness accounts for. Early studies found that a one standard deviation increase in teacher quality results in around 10% of a standard deviation increase in reading and writing test results per year. An alternative way of expressing the difference due to increased teacher effectiveness is in terms of years of achievement gain; for instance,

“some teachers [produce] 1.5 years of gain in achievement in an academic year while others with equivalent students produce only 1/2 year of gain” [1], p. 467.

Many studies have followed these early breakthroughs. Again referencing the 2011 study, the author states: “Literally hundreds of research studies have focused on the importance of teachers for student achievement” [1], p. 467. The typical result is remarkably consistent: a one standard deviation change in teacher effectiveness yields a 10–20% standard deviation change in pupil attainment, larger than most educational interventions and comparable with the impact of high-performing charter schools. This appears to be true across different stages of school, different subjects (though typically greater in mathematics), and (to the extent that evidence exists) across different countries. A 2013 study shows that teacher impacts on non-test behaviors such as absences and grade progression also predict later educational outcomes [3]. As researchers’ interest in non-cognitive attributes has risen, studies have further shown that teachers also influence self-efficacy in mathematics, as well as happiness and behavior in class.

Pupil test scores have been shown to be correlated with human capital growth and thus influence multiple outcomes of interest. The 2014 study finds that pupils taught by highly effective teachers earn more, are more likely to go to university, and to live in richer neighborhoods [2]; other research shows that teacher effectiveness predicts high school dropout rates and college plans.

### **Teacher effectiveness measures are reliable and informative**

One key aspect of teacher effectiveness measures that policymakers find quite appealing is that the technique has been shown to be robust to strong critiques of bias. The main concern in this regard has traditionally been that teachers might be assigned pupils with particular characteristics that lead to different test scores. Suppose, for instance, that for some reason an averagely effective teacher was assigned a group of ambitious, highly motivated pupils; if that is not measured in the data set, then the higher test scores achieved by those pupils will be incorrectly assumed to derive from a highly effective teacher. This is the line taken by the author of a 2009 study, who argues that there is strong non-random sorting within schools [4]. This has generated a number of responses within the literature, some statistical and some quasi-experimental, and it is the experimental response that is most convincing.

Experimental evidence from two studies reinforces the view that conventional estimates do not suffer from strong bias [5], [6]. The respective authors estimated teacher effectiveness from past classes and then randomly assigned high- and low-effectiveness teachers to new classes. The historical estimates predicted student progress well—that is, the teachers estimated to be more effective influenced pupils more positively than those estimated to be low-effective teachers, even after random class assignment. This suggests that when students’ prior ability is sufficiently accounted for, standard techniques for estimating teacher effectiveness are reasonably free from bias. Yet another study implements a quasi-experimental test for bias, comparing cross-cohort variation in estimated teacher effects and cross-cohort variation in attainment; the authors find no evidence for bias [7].

These results provide ample support for the notion that estimates of the distribution of teacher effectiveness are both plausible and valuable for policy making.

## Key challenges to making the most of teacher effectiveness metrics

While teacher effectiveness has been shown to be quite important for student outcomes, the literature has been unable to identify a consensus list of teacher characteristics that are correlated with their effectiveness. This is key, because a strong predictor of effectiveness would be very helpful for schools in decisions to hire, retain, and reward teachers. Arguably the most important characteristics for policymakers and school administrators to gain more understanding of are whether teacher effectiveness is correlated with teaching experience, the teacher's own educational achievement, and with teacher training or certification—that is, their route into teaching.

Until recently, the relationship between effectiveness and experience had been considered settled: effectiveness was markedly lower in the initial few years of work, but improved substantially during a teacher's first three years; thereafter, however, teacher effectiveness was believed to remain essentially constant. This is very surprising—in most other professions and vocations, it would be expected that a worker improves with practice. To find that a teacher with three years' experience is generally as effective as one with 13, 23, or 33 years' experience is quite interesting.

However, more recent studies have used new data to show teacher effectiveness increasing much later into the job. The argument is a simple and plausible one: the new data suggest that teachers do continue to slowly improve over their careers, but that this is offset by highly effective teachers being more likely to leave the profession. Both of these aspects are of great policy interest: increasing effectiveness over time matters for teacher retention decisions (see below), and the presence of a higher loss rate for more talented teachers is relevant for pay decisions, among other factors. Of course, it remains true that a teacher with a lot of experience chosen at random will have about the same level of effectiveness as a teacher with a much shorter amount of experience.

The second key characteristic of interest is the teacher's own academic background. There seems to be consensus on this point that it is largely uncorrelated with effectiveness. A 2015 study, for instance, notes that most of the papers in this field find no effect of the teacher holding a master's degree [8].

The third key characteristic, the link between effectiveness and teacher training or the “route” into teaching, is clearly important for policymaking: do some methods of selecting and training teachers produce more effective teachers? This is becoming increasingly important, as recent times have seen an increase in the availability of new pathways into teaching (e.g. school-based routes such as Teach for America or Teach First in the UK). Research has so far found mixed evidence on teacher effectiveness from different teacher certification programs. This is surely a key area for further research: whether teacher training programs can be evaluated and modified to raise teacher effectiveness. There are obviously a number of complexities, not least the issue, largely unconsidered in research so far, of the non-random selection of trainees into programs and then the non-random recruitment from particular programs into particular schools.

Overall, the fact that teacher effectiveness does not appear to be reliably correlated with observable teacher characteristics is a substantial problem for policymaking in this area.

## The knowns and unknowns of teacher policy

While there has been a great deal of research in this field in the last decade, there is still much that remains unknown. Researchers have a clear sense of the huge importance of teacher effectiveness in relation to public policy. But the necessary next step, understanding the best ways to raise it, is where the research frontier currently lies.

The key question in this realm is: are effective teachers born or made? In other words, is the ability innate (“born”), or are there reliable ways to improve the effectiveness of current and nascent teachers (“made”)? Undoubtedly, the answer will be somewhere between the two extremes, but the distinction illustrates the two main tracks taken within the literature. If effective teachers are “born” then the salient issues revolve around selection, identifying that ability, hiring, and retention. On the other hand, if effective teachers are “made” then researchers should focus on the best training, mentoring, and feedback mechanisms. It should be noted that there is a third dimension to consider, namely teacher effort; in this case the policy interest is about enhancing such effort, typically via performance pay. However, this topic is not addressed in the present article. Having said that, it is worth noting that measures of teacher effectiveness would be required for a well-founded performance pay scheme. The suitability of such a scheme would vary from case to case, depending on the reliability and accuracy of the measurement framework, and the degree to which the measure was “game-able” by teachers and schools.

## Teacher selection: Identification, hiring, and retention

As discussed, reliable measures of teacher effectiveness exist; however, these are only applicable for people who teach. They are not available when making hiring decisions, and this is one of the central problems in teacher selection: how should teachers be hired when there is little information at hand to predict whether they will be good at the job? Research has looked at commonly accessible indicators, such as psychological traits, that would be available when interviewing for non-teacher jobs and found very little that is predictive of teacher effectiveness.

While research on informed teacher hiring is still at an early stage, there is some work on teacher retention or layoff which can directly use teacher performance information. An influential and controversial contribution comes from a 2010 study, which uses the known facts about teacher effectiveness to simulate an optimal teacher retention policy [9]. The policy is based on value-added information about a given teacher’s performance; the core tradeoff is essentially between laying off low value-added teachers and hiring inexperienced teachers as replacements, thereby accepting the early-career effectiveness penalty noted above. The result is very strong: within the assumptions of the study, schools should retain less than half of their teachers after their first year of teaching. The fact that differences in effectiveness are large, persistent, and unknowable before hiring, means that even on the basis of one year’s measure, the one-time cost of replacement is worth it to avoid a career-long tenure of an ineffective teacher. Needless to say, this finding has not been popular with teacher unions. A subsequent study notes that the 2010 study’s model assumes a perfectly elastic supply of teachers [10]. The later study adapts the earlier model to allow for wage rises that would accompany such a retention policy. It finds that optimal layoff rates are lower than originally found, but still considerably higher than current practice [10].

A few studies have examined actual layoff programs and their relationship to teacher effectiveness. The results show that using teacher effectiveness as the layoff criterion can have substantial effects. For example, the controversial US IMPACT system of teacher evaluation in Washington DC entailed unusually high and credible threats of teacher dismissal for poor performance. The strong dismissal threat increased quits by low-performing teachers by 50% and raised performance among the remaining low performers by 27% of a (teacher-level) standard deviation.

### **Teacher improvement: Training, mentoring, and feedback**

At the other end of the spectrum, policies could endeavor to improve the effectiveness of existing teachers. Here too, researchers have made progress, but have not yet found a consensus approach. There are several layers to improvement. First, there is the formal initial training, preparation, or certification to become a teacher. The work on this to date does not suggest any dramatic differences in effectiveness from different training routes, although there are concerns about the importance of endogeneities in route selection, school selection, and trainee selection. Once a teacher is hired there are a number of potential channels for improving effectiveness. These channels are reviewed below according to increasing degrees of formality: informal learning, peer mentoring, peer coaching, and evaluation.

The most informal method of improvement is simply learning on-the-job from other teachers. Based on some observational data and field experiments, research shows that working with effective peer groups can raise a teacher's own effectiveness. For a school trying to assemble a group of effective teachers, there would appear to be positive side-effects related to mutual learning.

Taking one step up the formality rung, formal programs of peer mentoring do not appear to generate strong or substantial impacts. However, a slightly different and more promising approach involves personalized teacher coaching; initial results suggest substantial, persistent, and significant impacts on teacher classroom practices at least. Studies of these two ideas again struggle to nail down any medium- to long-term effects due to the difficulties in generating data that follow teachers over time and across schools. It is thus challenging to calculate robust measures of their effectiveness.

Lastly, teacher evaluation programs can work to raise individual teacher effectiveness and help to identify low-performing teachers. The potential improvements come via enhancing teachers' skills or raising effort, or both. This seems to be a promising approach. One research team ran a field experiment with a set of interventions focused on ongoing, daily interactions of teachers with their students. The intervention included initial workshop-based training, access to an annotated video library, and a year of personalized coaching followed by a brief booster workshop. This program delivered substantial impacts on student attainment, over 10% of a standard deviation, in the year after the intervention.

A 2012 US study looked at the impact of an evaluation program in Cincinnati on teacher effectiveness [11]. Teachers were quasi-randomly assigned to evaluation during a year-long program of classroom observation. The observation used a particular framework

or “rubric” for describing and evaluating teaching practices. Crucially, the analysis was able to track teachers and their pupils for some years after the evaluation. The study shows that teachers involved in the program were more effective to a substantial and enduring degree, and that the biggest gains were seen in initially low-performing teachers. The findings are even more impressive in that they relate to mid-career teachers who might be expected to have hard-to-shift capabilities. Moreover, the analysis suggests the effect largely derives from improvements in effectiveness and classroom skills, rather than a one-off increase in effort due to the presence of the evaluators. Further research on an evaluation and discussion intervention also shows raised attainment during the evaluation period which persisted afterwards in the treatment schools.

## LIMITATIONS AND GAPS

As mentioned above, there are many potential avenues that require further research within this field, and thus gaps in its current understanding. For example, research into modes of improving teacher effectiveness suffer from a key empirical problem: tracking the persistent effects of training, mentoring, and development is hampered by a lack of longitudinal data. This data gap makes it very difficult to determine longer-term outcomes from many innovative programs and impedes the robustness of potential policy evaluations.

Ultimately, despite the many promising results seen in relation to isolated classroom interventions, more work is required to replicate and confirm these early findings before researchers can confidently recommend such policies.

## SUMMARY AND POLICY ADVICE

Given its very strong potential impact on pupil achievement, teacher effectiveness should be a central concern for education policymakers. The literature clearly demonstrates that teacher effectiveness can be robustly and reliably measured. Moreover, researchers have identified three main pathways by which policymakers might best make use of this reliable measure: (i) improving teacher selection and hiring procedures, (ii) reforming teacher contracts and the tenure/retention decision, and (iii) re-thinking teacher professional development.

Despite the presence of many promising program examples, it is important to reiterate the need for considerable further research to provide replication, confirmation, nuance, and detail with respect to all of these policy possibilities. The potential size of the impact of improving teacher effectiveness represents a truly grand prize for the countries, cities, and schools which manage to crack the code of how to raise teacher effectiveness.

## Acknowledgments

The author thanks an anonymous referee and the IZA World of Labor editors for many helpful suggestions on earlier drafts. Financial support from COEURE is gratefully acknowledged. No conflicts of interest arise. This article draws heavily on the author’s

review for COEURE: *The State of the Art in the Economics of Education*. Online at: <http://www.coeure.eu/wp-content/uploads/Human-Capital-and-education.pdf>

### **Competing interests**

The IZA World of Labor project is committed to the IZA Code of Conduct. The author declares to have observed the principles outlined in the code.

© Simon Burgess

## REFERENCES

### Further reading

Dee, T., and J. Wyckoff. *Incentives, Selection, and Teacher Performance: Evidence from IMPACT*. NBER Working Paper No. 19529, 2013.

Jackson, C. K., J. Rockoff, and D. Staiger. “Teacher effects and teacher-related policies.” *Annual Review of Economics* 6 (2014): 801–825.

Wiswall, M. “The dynamics of teacher quality.” *Journal of Public Economics* 100 (2013): 61–78.

### Key references

- [1] Hanushek E. “The economic value of higher teacher quality.” *Economics of Education Review* 30:3 (2011): 466–479.
- [2] Chetty, R., J. Friedman, and J. Rockoff. “Measuring the impacts of Teachers II: Teacher value-added and student outcomes in adulthood.” *American Economic Review* 104:9 (2014): 2633–2679.
- [3] Jackson, C. K. *Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina*. NBER Working Paper No. 18624, 2012.
- [4] Rothstein, J. “Student sorting and bias in value-added estimation: Selection on observables and unobservables.” *Education Finance and Policy* 4:4 (2009): 537–571.
- [5] Kane, T., and D. Staiger. *Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation*. NBER Working Paper No. 14607, 2008.
- [6] Kane, T., D. McCaffrey, T. Miller, and D. Staiger. *Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment*. Seattle: Bill & Melinda Gates Foundation, 2013.
- [7] Chetty, R., J. Friedman, and J. Rockoff. “Measuring the impacts of Teachers I: evaluating bias in teacher value-added estimates.” *American Economic Review* 104:9 (2014): 2593–2632.
- [8] Ladd, H., and L. Sorensen. *Do Masters Degrees Matter?: Advanced Degrees, Career Paths and the Effectiveness of Teachers*. CALDER Working Paper No. 136, 2015.
- [9] Staiger, D., and J. Rockoff. “Searching for effective teachers with imperfect information.” *Journal of Economic Perspectives* 24:3 (2010): 97–118.
- [10] Rothstein, J. “Teacher quality policy when supply matters.” *American Economic Review* 105:1 (2015): 100–130.
- [11] Taylor, E., and J. Tyler. “The effect of evaluation on teacher performance.” *American Economic Review* 102:7 (2012): 3628–3651.

### Online extras

The **full reference list** for this article is available from:

<https://wol.iza.org/articles/understanding-teacher-effectiveness-to-raise-pupil-attainment>

View the **evidence map** for this article:

<https://wol.iza.org/articles/understanding-teacher-effectiveness-to-raise-pupil-attainment/map>