

---

## 2. Teacher expertise: Why it matters, and how to get more of it

---

*Dylan William*

### **The need to improve education**

Higher educational achievement has a large number of benefits for education and society. With higher educational achievement, individuals earn more money (Crawford and Cribb, 2013), are healthier (OECD, 2010 Table A9.4, p. 164), are less likely to be disabled towards the end of their lives (Jagger et al., 2007) and live longer (Lleras-Muney, 2005). For society, the benefits are, if anything, even greater. More education is associated with reduced criminal justice costs (Levin, Belfield, Muennig, and Rouse, 2007), reduced healthcare costs (Levin et al., 2007), and increased economic growth (Hanushek and Wößmann, 2010).

It is therefore hardly surprising that the last 40 years have seen substantial investment in attempts to improve educational outcomes for young people. Initial attempts focused on increasing inputs into education, such as expenditure on buildings, teacher qualifications, or the number of years students spend in school, not least because these are relatively easy outcomes to measure. However, it has become clear in recent years that educational inputs are only weakly related to student outcomes. Increased expenditure on education can result in improved outcomes (Jenkins, Levacic, and Vignoles, 2006), but more often, does not (Hanushek and Rivkin, 1997). The relationship between teacher qualifications and student progress is at best weak, and often non-existent (Harris and Sass, 2007). It is possible for students to spend extra time in school without learning very much (Pritchett, 2013).

However, over the last 20 years, the quality of data on the performance of educational systems has improved dramatically. International comparisons such as PISA and TIMSS allow countries to compare themselves with others. In England, the adoption of a national curriculum, with standardised forms of assessment at the ages of 7, 11, 16, 17 and 18 together with the availability of unique pupil identifiers provides more data, of better quality, about what students are learning in school than ever before.

One of the most surprising findings generated by analysis of the national pupil database in England is that, as long as you go to school, it doesn't matter very much which school you go to. While average Key Stage 2 test scores and GCSE grades vary markedly from school to school,

most of these differences disappear when differences in the prior attainment of students is taken into account. Whether the outcome measure is PISA scores, or GCSE grades, only around 7 percent of the variation in student achievement is attributable to the school (William, 2010).

This explains why most of the attempts to improve schools in England have met with limited success. The kind of school a student attends – faith school or county school, private school or state school, specialist school or academy – seems to have relatively little impact on how much progress a student makes. As with many other social science phenomena, the difference within each category is far greater than the difference between the categories.

In particular, recent work has shown that the quality of the individual teacher is one of the most significant variables influencing how much progress students make in school (Hanushek and Rivkin, 2006). Now of course, it is important to note that teacher quality is not the same as teaching quality. The actual quality of teaching will depend on a number of factors, such as the time teachers have to plan teaching, the size of classes, and the human and material resources supporting the teacher. However, it does seem that a substantial proportion of the variation in how much students learn in different classrooms is due to the individual qualities of the teacher. Teachers who are more effective in any given year are more effective in subsequent years (McCaffrey, Sass, and Lockwood, 2008), and remain more effective when they move to other schools, even when the schools differ greatly in their socio-economic status (Sanders, Wright, and Langevin, 2008; Xu, Özek, and Corritore, 2012).

This is why, over recent years, in many countries, there has been a shift from treating teachers as a commodity (ie regarding all teachers as equally good, so that what matters is getting enough teachers at a reasonable cost) to regarding teacher quality as a key element in educational policy.

Teacher quality can be increased in two ways: we can replace existing teachers with better ones; or we can improve the teachers we already have. In many countries, policy efforts have focused on the former. Schemes such as Teach for America and Teach First have sought to attract those with higher college achievement into the profession. Such policies are often supported by appeals to international comparisons such as PISA, where entry into teaching in many high performing countries is highly competitive. While one cannot make causal inferences from such correlations, more direct studies have found that teachers with first class degrees are more effective than others (Slater, Davies, and Burgess, 2008) and that when students are taught by teachers with stronger academic backgrounds, they remember what they are taught longer (Master, Loeb, and Wyckoff, 2014). However, the differences found in these studies are small, and dwarfed by the variations in overall quality. Moreover, since the average amount of time Teach First teachers spend actually teaching is about three years, combined with the fact that it is, by definition, an elite route into the profession, the direct impact of Teach First is likely to be small (although the impact in terms of the status of teaching as a profession may well be substantial).

An alternative strategy for improving teacher quality is to identify and remove the weakest teachers. Even if they are only replaced with average teachers (thus avoiding the need to find especially good teachers), over

time, the impact on teacher quality would be substantial. While removing teachers who have been identified as ineffective has become much more straightforward over the last 10 years, deciding whether a teacher is ineffective turns out to be rather difficult. As John Mason has noted, “Teaching takes place in time, but learning takes place over time” (Griffin, 1989). What may appear to be effective practice when observed may not lead to longer-term retention. The best currently available observation systems, such as Charlotte Danielson’s (1996) Framework for Teaching, do predict student progress – if you are taught by a teacher rated as “distinguished” you will learn 30 percent more than if you are taught by a teacher rated as “unsatisfactory” (Sartain et al., 2011). But the best teachers are 400 percent more effective than the least effective (Hanushek and Rivkin, 2006), which suggests that the Danielson framework captures only around one-tenth of the variation in teacher quality. The idea that some teachers are 400 percent more productive than others may seem to be at variance with the fact that only 7 percent of the variation in student achievement is attributable to the school, but the distribution of teachers in the system is fairly random, so that all schools have a broad mix of more effective and less effective teachers.

More observations would, of course, probably improve the relationship between observations and student progress, but Hill, Charalambous, and Kraft (2012) estimated that using observations of practice to produce ratings of teacher quality with a reliability of 0.9 would require seeing a teacher teaching five different classes and having each lesson observed by six independent observers, which would probably be unmanageable across the system.

This is why some policies have focused on determining the progress made by students from year to year by estimating the value-added by the teacher. The problem with such approaches is that because of the complexity of assumptions involved, alternative, but equally reasonable, models yield different results. For example, Goldhaber, Goldschmidt, and Tseng (2013) found that 9 percent of the teachers placed in the top quintile of teacher quality with one model were placed in the bottom quintile with a different, but equally plausible model.

Moreover, value-added models do not capture all of what teachers contribute to student learning. The best teachers benefit their students for at least three years after they stop teaching them (Rothstein, 2010). In other words, the best teachers appear to develop capabilities in their students that are not captured in measures of achievement at the end of that year, but appear to be important for long-term success.

Finally, the political cost of removing ineffective teachers should not be underestimated. Teacher observations and value-added measures of teaching performance have large margins of error. If the burden of proof is set too low, then effective teachers are dismissed, but if it is set too high, then few teachers will be identified as ineffective. For example, in a study of reading teachers in Florida, Winters and Cowen (2013) found that if the criterion for removal of teachers was set as being in the lowest 5 percent of value-added for two consecutive years then only one teacher in every 500 would be identified for removal. Relaxing the burden of proof would lead to more teachers being removed, but at the expense of the removal of a number of effective teachers.

## **So what's to be done?**

To sum up the argument so far, increased achievement is a necessity for young people to be able to function effectively, and find meaningful work, in today's increasingly complex society. Increased achievement requires increased teacher quality, which, in turn, requires improving the quality of those teachers already working in our schools.

In this context, it is worth noting that continuing to improve practice post qualification is not a requirement in all professions. In law and accounting, for example, there is an expectation that professionals will keep up with new developments, but there is no explicit requirement that they improve their performance. It is the moral imperative – that when teachers do their jobs better, their students are healthier, live longer, and contribute more to society – that should drive teachers to improve.

There are those who believe that teachers cannot improve, beyond the sharp improvements in the first few years of practice (Rivkin, Hanushek, and Kain, 2005). It is certainly the case that left to their own devices, teachers do not improve very much. For example, while Leigh (2010) found that teachers did continue to improve over the course of a 20 year career, the improvement of teachers over this time was only equivalent to about an extra two weeks' learning per year for students. In other words, the difference between a good teacher and an average teacher on their first day is many times larger than the improvement of a teacher over a 20 year career. This would appear to be in contrast to trends in other professions, where professional learning appears to be more rapid. For example, Norcini (2009) suggests that one year's training in cardiac surgery improves performance by 0.3 standard deviations, which suggests that an outstanding novice is on a par with an average surgeon with six years' experience.

However, these average trends mask important differences between teachers. Atteberry, Loeb, and Wyckoff (2013) found that lower performing novices improved much faster than other teachers – so much so that initial performance explained less than 5 percent of the variation in performance after five years in practice – and the best teachers (ie top quintile initially) did not improve at all.

One interpretation of these findings is that once teachers reach a particular level of performance, further improvement is difficult. However, given that other studies show that suitably-focused professional development support can have a significant impact on student achievement for all teachers – equivalent to an improvement of teacher quality of at least one standard deviation (Fennema et al., 1996; Wiliam, Lee, Harrison, and Black, 2004) – then it appears as if all teachers can make significant improvements in their practice. Such a finding is also consistent with the research on expertise, which suggests that at least 10 years of “deliberate practice” is required to attain high levels of performance in a domain (Ericsson, Charness, Feltovich, and Hoffman, 2006). The question is then how to provide such support.

## **Creating a framework for Continuing Professional Development**

As noted earlier, many professions regard practitioners as fully qualified after a certain amount of professional preparation undertaken at the

beginning of a career. Others require teachers to keep current with new developments, although improvements in practice are not required. Still others require professionals to undertake a certain amount of professional development in order to retain a licence to practice.

In some ways, licensing schemes that specify that teachers undertake a certain amount of professional development without specifying what form it should take are understandable. The diversity of contexts of application is so great that requiring professionals to undertake specific programmes of study in order to retain their professional accreditation would be highly bureaucratic, and would be likely to lead to charges that the professional development being imposed was not relevant to particular individuals. The problem, however, is that such schemes quickly become unfocused, with professionals being able to satisfy their professional development requirements by undertaking activities that have no impact on their professional competence. In other words, most licensing schemes quickly degenerate into a system where professionals have to prove that they have endured a certain number of hours of approved professional development. The crucial question is then no longer, “Will this make me a better practitioner?” but “How many PD hours will it give me?”

Also, if the range of performance of those with the same amount of experience ranges greatly (as it clearly does for teaching) then the amount of time someone has been doing a job is not a useful guide to the training needed and some form of assessment of current skill would be needed, adding further to the bureaucracy. In many medical specialisms, rigorous assessments are in fact undertaken, with some individuals progressing much faster than others to higher levels of recognition, but the lack of consensus about what does, in fact, constitute higher levels of skill in teaching suggests that such schemes would be difficult to implement fairly in education.

Additionally, the higher levels of recognition used in some professions are almost invariably limited in number. There are a certain number of posts as consultant cardiac surgeons, or advanced skills teachers, available, which makes the system inherently competitive. There is, almost always, a criterion-referenced component to the promotion, but when the number of posts is limited, to be successful, a candidate needs also to be better than the other candidates. In other words, competition between individuals would appear to be an inevitable aspect of such systems, if only to create a cap on salary costs.

The difficulty of specifying relevant professional development for teachers, the diversity of skill levels of those with the same duration of experience, the difficulty of measuring such skills validly, and the competitive nature of rewards, whether in terms of compensation or recognition, suggests that the best way forward, for teaching at least, is to create a structure in which all teachers are expected to improve their practice as long as they remain teaching.

### **Making teaching a learning profession**

There are many structures that could support teacher improvement, but in the remainder of this essay, I outline a model that I believe would be effective, and politically acceptable to all stakeholders in

education. The key change is that teachers' contracts should include an explicit expectation that they improve their practice every year, that the improvement should be focused on aspects of practice that are likely to improve outcomes for their students, that teachers should be responsible for deciding themselves which aspects of practice would be most appropriate to develop, and for providing the evidence of improvement. Each of these is discussed briefly below.

### **Each teacher should improve**

There are both moral and pragmatic rationales for requiring all teachers, rather than just those that need help, to improve. At the heart of the moral imperative is the demonstrable empirical fact that when teachers do their job better, their students are healthier, live longer, and contribute more to society. With such a moral imperative, even the best teachers have a moral duty to improve. The pragmatic case is perhaps just as important, and related to the changing nature of the world of work. While preparation for the world of work is just one of the aims of education, it is perhaps the one where the demands are changing most rapidly. Education should of course pass on the great things that have been thought and said (Arnold, 1869/1932, p. 6), though what is considered great work does not, by definition, change that rapidly. Education also has a role in preparing young people to take greater control of their lives, and to prepare them for active citizenship, but while society is changing, the impacts for educational systems are less than the changing demands of work. Put simply, if we focus only on the teachers who need help the improvement in teacher quality will be too small to win “the race between education and technology” (Tinbergen, 1975).

### **The focus should be on things that benefit students**

It may seem obvious that teachers should receive professional development that is focused on aspects of practice that benefit their students, but much, perhaps most, of the professional development that teachers have been subjected to since the 1988 Education Reform Act appropriated five days of teachers' vacation for inservice training and has been focused on fads with little research evidence in their support. Educational research is unlikely ever to be able to tell teachers what to do, but it can indicate which directions are likely to be the most productive for the development of practice.

### **Each teacher should be responsible for deciding what to work on**

Perhaps the strangest feature of the professional development landscape in the UK over the past 30 years is the apparently widespread belief that all teachers in a school should receive, and would benefit from, the same professional development. The idea that the same intervention will improve mathematics teaching and physical education is rather odd. Moreover, teachers are different, so that what would help one art teacher become a better teacher may not help another art teacher with a different teaching style. Novice teachers will, of course, need clear direction from their supervisors about which aspects of their practice should be priorities for development, but once teachers are established in their classrooms,

it seems reasonable to assume that each teacher has a better idea of what will improve the learning of their students, in their classroom, in the context of what they are teaching them, than anyone else. Also, even if this assumption is occasionally not correct, it is a far better starting point for a conversation with professionals than the idea that they are doing things wrong, and that they need to be fixed. Moreover, such an assumption seems to me to be far more likely to encourage teachers to take some risks in developing their practice. Each teacher would need to produce research evidence that makes at least a *prima facie* case that what they are working on is likely to benefit their students. Teachers would therefore need to engage with research, but in a very grounded way, and one that is directly related to their practice.

### **Each teacher is responsible for providing evidence of improvement**

As well as deciding what to improve, each teacher would be responsible for deciding what evidence they needed to collect to demonstrate that their practice had improved. The evidence could take the form of student achievement data, videos of classroom practice or even questionnaire responses from students. The important point is that the teacher would be free to collect the form of data they felt most strongly supported their claim to have improved their practice.

As well as getting teachers into the habit of routinely evaluating their practice, it would prevent charges that the evidence being collected was inappropriate. After all it would be the teacher's own evidence, related to their own claims about which aspects of practice they had improved. Teachers would meet at least annually with a supervisor to discuss the evidence, and the supervisor would have to determine whether there was evidence of improved practice. Obviously there would need to be appeal mechanisms to ensure fairness, but such procedures need not be unduly burdensome because the task would be simply to determine whether the available evidence supported the claim of improvement by the teacher. These annual meetings would also provide time for the teacher to plan with the supervisor the next steps for continued professional learning, including identifying what support might be needed, especially where the proposed improvements involve a degree of risk.

These basic principles form, in my view, the minimal core of a professional development structure for teachers. There are other features that could be added, but they would weaken the case that the structure should apply to *all* teachers. For example, in my own work with teachers, I have emphasised the importance of collaboration with other teachers (William, 2012). However, currently, the available evidence does not support the idea that collaboration with other teachers will *always* be the best way for *every* teacher to improve her or his practice. Given the diversity of contexts in which teacher professional development will take place, we should be wary of adding features that will typically, but not always, improve outcomes for teachers and the learners they serve.

### **Conclusion**

The main argument of this essay is that improving the quality of serving teachers is essential, both for the individual and society, and that it is

best achieved simply by requiring each teacher to improve their practice year-on-year, in ways that are likely to benefit students. Such an approach side-steps many issues such as what should teachers improve, and how to measure improvement, because these are matters for the individual teacher. In terms of political economy, the interests of key stakeholders are addressed. The idea of all teachers improving is at the heart of teacher unions' discourse of professionalism, but there is a hard edge to this – the need to provide evidence of improvement – so that pay rises are earned rather than received simply for ageing. Perhaps most importantly, when all teachers in a school are committed to improving their practice, collegiality is more likely than competition, which creates a virtuous circle of continuing, substantial improvement.

## References

- Arnold, M. (1869/1932). *Culture and anarchy: An essay in political and social criticism*. London, UK: Cambridge University Press.
- Atteberry, A., Loeb, S. and Wyckoff, J. (2013). *Do first impressions matter? Improvement in early career teacher effectiveness*. Washington, DC: Center for Analysis of Longitudinal Data in Educational Research.
- Crawford, C. and Cribb, J. (2013). *Reading and maths skills at age 10 and earnings in later life: a brief analysis using the British Cohort Study* (Vol. REPO3). London, UK: Institute for Fiscal Studies.
- Danielson, C. (1996). *Enhancing professional practice: a framework for teaching*. Alexandria, VA: ASCD.
- Ericsson, K. A., Charness, N., Feltovich, P. J. and Hoffman, R. R. (2006). *The Cambridge handbook of expertise and expert performance*. Cambridge: UK.
- Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R. and Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education*, 27(4), 403–434.
- Goldhaber, D. D., Goldschmidt, P. and Tseng, F. (2013). Teacher value-added at the high-school level: Different models, different answers? *Educational Evaluation and Policy Analysis*, 35(2), 220–236. doi: 10.3102/0162373712466938
- Griffin, P. (1989). Teaching takes place in time, learning takes place over time. *Mathematics Teaching*, 12–13.
- Hanushek, E. A. and Rivkin, S. G. (1997). Understanding the twentieth-century growth in U.S. school spending. *Journal of Human Resources*, XXXII(1), 35–68.
- Hanushek, E. A. and Rivkin, S. G. (2006). Teacher quality. In E. A. Hanushek and F. Welsh (Eds.), *Handbook of the Economics of Education* (Vol. 2, pp. 1051–1078). Amsterdam, Netherlands: Elsevier.
- Hanushek, E. A. and Wößmann, L. (2010). *The high cost of low educational performance: the long-run impact of improving PISA outcomes*. Paris, France: Organisation for Economic Co-operation and Development.
- Harris, D. N. and Sass, T. R. (2007). *Teacher training, teacher quality and student achievement* (Vol. 3). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

- Hill, H. C., Charalambous, C. Y. and Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher*, 41(2), 56–84.
- Jagger, C., Matthews, R., Melzer, D., Matthews, F., Brayne, C. and Medical Research Council Cognitive Functioning Aging Study. (2007). Educational differences in the dynamics of disability incidence, recovery and mortality: Findings from the MRC Cognitive Function and Ageing Study (MRC CFAS). *International Journal of Epidemiology*, 36, 358–365.
- Jenkins, A., Levacic, R. and Vignoles, A. (2006). *Estimating the relationship between school resources and pupil attainment at GCSE* (Vol. RR727). London, UK: Department for Education and Skills.
- Leigh, A. (2010). Estimating teacher effectiveness from two-year changes in students' test scores. *Economics of Education Review*, 29(3), 480–488.
- Levin, H. M., Belfield, C., Muennig, P. and Rouse, C. (2007). *The costs and benefits of an excellent education for all of America's children*. New York, NY: Teachers College.
- Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. *Review of Economic Studies*, 72(1), 189–221.
- Master, B., Loeb, S. and Wyckoff, J. (2014). *Learning that lasts: Unpacking variation in teachers' effects on students' long-term knowledge* (Vol. 104). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.
- McCaffrey, D., Sass, T. R. and Lockwood, J. R. (2008). *The intertemporal stability of teacher effect estimates*. Nashville, TN.
- Norcini, J. J. (2009, November 16). Personal communication. *Impact of training on surgical performance*.
- Organisation for Economic Cooperation and Development (2010). *Education at a glance*. Paris, France: Organisation for Economic Co-operation and Development.
- Pritchett, L. (2013). *The rebirth of education: Schooling ain't learning*. Washington, DC: Brookings Institution Press.
- Rivkin, S. G., Hanushek, E. A. and Kain, J. F. (2005). Teachers, schools and academic achievement. *Econometrica*, 73(2), 417–458.
- Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics*, 125(1), 175–214.
- Sanders, W. L., Wright, S. P., and Langevin, W. E. (2008). Do teacher effect estimates persist when teachers move to schools with different socioeconomic environments? (Working Paper 2008–20.) Nashville TN: National Center on Performance Incentives, Vanderbilt University's Peabody College.
- Sartain, L., Stoelinga, S. R., Brown, E. R., Luppescu, S., Matsko, K. K., Miller, F. K., Durwood, C.E., Jiang, J.Y. and Glazer, D. (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation*. Chicago, IL: Consortium on Chicago School Research.
- Slater, H., Davies, N. and Burgess, S. (2008). *Do teachers matter? Measuring the variation in teacher effectiveness in England* (Vol. 09/212). Bristol, UK: University of Bristol Institute of Public Affairs.
- Tinbergen, J. (1975). *Income distribution: Analysis and policies*. Amsterdam, Netherlands: North-Holland.

- Wiliam, D. (2010). Standardized testing and school accountability. *Educational Psychologist*, 45(2), 107–122.
- Wiliam, D. (2012). *Sustaining formative assessment with teacher learning communities*. Seattle, WA: Kindle Direct Publishing.
- Wiliam, D., Lee, C., Harrison, C. and Black, P. J. (2004). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education: Principles Policy and Practice*, 11(1), 49–65.
- Winters, M. A. and Cowen, J. M. (2013). Who would stay, who would be dismissed? An empirical consideration of value-added teacher retention policies. *Educational Researcher*, 42(6), 330–337. doi: 10.3102/0013189X13496145
- Xu, Z., Özek, U. and Corritore, M. (2012). *Portability of teacher effectiveness across school settings* (Vol. 77).