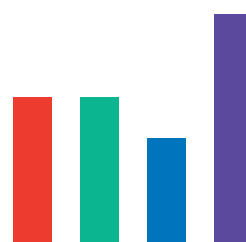




OUR THEORY OF CHANGE



Evidence Based
Education



CONTENTS

Why are we holding out for more professional development time
(even though school leaders say they can't manage it)?04

Why aren't we doing 'instructional coaching'
(even though everyone else seems to be)?07

Teacher collaboration: Why we're for it
(even when the evidence is weak).09

Why do we see teacher expertise as needing knowledge, skill and judgement
(even though most people seem to focus on just one of these)?10

Why are we betting on giving teachers more effective feedback
(even though no one else is really doing this)?13

References16



Professor Rob Coe

**Director of Research & Development,
Evidence Based Education**

Improving education at scale is hard. However, we know that teacher expertise is the key driver of student learning. There is abundant evidence that, of all the things schools can influence, “what teachers know, do, and care about” (Hattie, 2003) has the biggest impact on student outcomes, by some margin (e.g., Chetty et al. 2014; Rivkin et al., 2005; Rockoff, 2004), and that high-quality teaching narrows the attainment gap (Burgess et al., 2022; Slater et al. 2012). High-quality teaching is not a fixed or given quantity: it varies across classrooms and can be learnt, supported and nurtured (Coe et al., 2020). For all these reasons, we established the ‘Great Teaching’ project—to help improve outcomes by building teacher expertise.

We started with the Great Teaching Toolkit: Evidence Review where we set out to identify, review and then summarise the best available international evidence about what teachers’ practices, skills knowledge and behaviours are important for student learning. This gave us the Model for Great Teaching, our curriculum for teacher learning—a credible summary of the elements of great teaching practice, the kind that impacts most on learning.

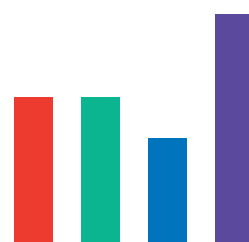
With this curriculum in place, we went on to develop the Great Teaching Toolkit—a package of courses and resources to enhance practice, tools to provide feedback for professional development over time, and a collaborative development model to help teachers work together.

This document sets out our thinking about key parts of our Theory of Change—why we designed the Great Teaching Toolkit as we have, what choices we made, what evidence underpins them, and the change we believe it can deliver. It presents a rationale for prioritising professional learning in schools. It explains what we mean by teacher expertise and where the focus should lie. It describes a coaching model that is based on reciprocal coaching in collaborative groups, supported by a range of resources. It explains why we think the popular “instructional coaching” model is unlikely to work at scale in real schools. It highlights the role of quality feedback as the key to helping complex systems (like classrooms and schools) improve themselves.



An important aspect of this Theory of Change is that it is more of a hypothesis than a guarantee. Yes, it is based on the best available research evidence and theory about how to improve education. But many aspects of that evidence are currently lacking, particularly in relation to how it applies in practice. So, we start with the best knowledge and assumptions we can—current best bets—but build in the ability to test and learn. We know that things will change as we learn how to make it work in practice in a wide range of contexts, and systematically test our hypotheses. This approach has been characterised as ‘improvement science’ (Lewis, 2015).

For schools and teachers who use the Great Teaching Toolkit (some of which is available for free), you are getting a set of tools that should help you to be even better than you already are. But you are also part of something much bigger: a systematic attempt to solve the intractable problem of improving education at scale, in a way that is authentic, sustainable and replicable. We can only do this with your help.



Why are we holding out for more professional development time (even though school leaders say they can't manage it)?

According to the Working Lives of Teachers and Leaders survey, the median reported time spent on continuing professional development by teachers in England is 21-30 hours per year—approximately half an hour a week (Adams et al., 2023, pp 36-46, p 124). That compares with their average working week of about 50 hours in total, about half of which is spent actually teaching. So, if we believe these surveys, roughly 1% of an average teacher's working time is spent in professional development (PD).

On the one hand, any call to get teachers to spend more time doing anything has to be made very cautiously. Workload for many is far too high and contributes to challenges of wellbeing, recruitment and retention. On the other hand, half an hour a week on the thing that has the most potential to improve the experience and outcomes of young people does seem pathetically little. Our response is that we do need to increase the time teachers spend on professional development, but also reduce their overall workload.

THE RATIONALE FOR PROFESSIONAL DEVELOPMENT AND ITS STRATEGIC IMPORTANCE

There is plenty of good evidence that teachers vary in the amount of learning that happens in their classrooms, that this effectiveness grows with experience, especially in the first few years of teaching, and that the growth can be accelerated by high-quality professional development (for example, Chetty et al., 2014; Fletcher-Wood & Zuccollo, 2020; Kraft & Papay, 2014; Sims et al., 2022). However, talking about the impact of 'professional development' as a thing is probably not very helpful. The reason for this is threefold: because PD covers such a range of different things; because most educational interventions will include some component of PD; and because the quality of different programmes varies so much.



Nevertheless, taken as a whole, research evidence supports the view that high-quality professional development for teachers gives a substantial boost to student attainment—so much so that it compares well with other interventions that schools might choose. Crucially, this boost is not a one-off impact that then washes out when the intervention ends; skills, knowledge and habits acquired through professional development keep giving benefits for every student that teacher subsequently teaches. And there is some evidence that PD can have wider benefits, including on teachers' job satisfaction and retention (Fletcher-Wood & Zuccollo, 2020; Sims et al., 2022).

It follows that for a school leader who cares about student outcomes and has even a passing interest in evidence, investing in professional development of staff should be a top priority. Strategy is about finding ways to deliver an organisation's goals by matching its available resources to activities with high leverage. If the long-term goals include student learning and teacher wellbeing, then PD should be prominent in the strategy.

All this is hard to reconcile with conversations amongst teachers and school leaders who say, "We have no time for PD." Allocating 30 minutes from a 50-hour week does not scream 'top priority'. For a strategic leader, time is a resource, not an excuse; you make time for the things that matter.

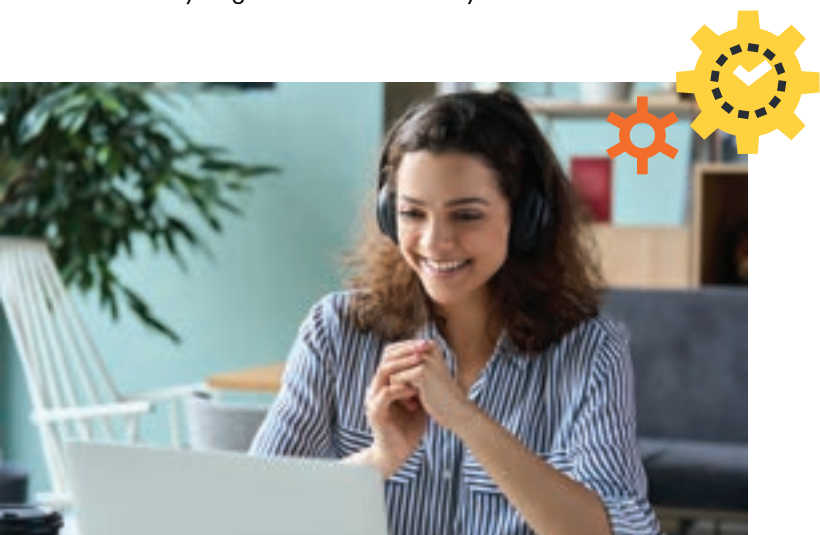


EXPLAINING THE ANOMALY

How do we explain the mismatch between the evidence that professional development should be a top priority, and the priority (as measured by time) that it typically gets? One explanation is that teachers and school leaders do not believe that PD will be effective (and perhaps do not know how to make it so). Actually, that might be quite a rational position to take; common experience (and a fair bit of evidence) is that PD often isn't great. So, to make the case for more time on PD, we have to ensure that what we use that time for has the most chance of leading to teachers becoming more effective. That is a core design focus of the Great Teaching Toolkit—and the subject of a future blog.

Another reason is that leaders concentrate more on short-term rather than long-term goals. Accountability pressures and a focus on current students could lead us to think that this year's results matter more than those in three (or five, or ten) years' time. For sure, some tactical quick fixes could have more impact on short-term outcomes than the slower burn that is PD—and for a school in acute crisis, a quick fix is needed. But as a strategy for most schools, rotating quick fixes is a toxic, inefficient way to go.

A third explanation is that there are indeed a lot of things teachers do that have more immediate drivers than PD. If teachers don't spend time on PD, no one is obviously going to suffer—or even notice. By contrast, if teachers have not prepared their next lesson or marked the work from the last one, the consequences can be immediate, obvious and painful. As human beings, we are all motivated more by pain or gain that is immediate. By choice, teachers will tend to fill up all their available time with these necessary tasks, and find it hard to replace the immediate gratification they provide with the promise of delayed gratification offered by PD.



HOW CAN LEADERS FIND MORE TIME FOR PROFESSIONAL DEVELOPMENT?

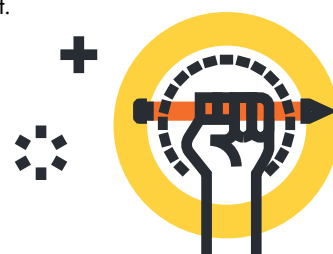
A key principle is that schools can't just add professional development on as an extra thing; something must be taken away. Fortunately, most schools should be able to find plenty of time-consuming things that add less value than professional learning. Exactly what these things are and how to stop doing so much of them is a key focus in the support for school leaders in the Great Teaching Toolkit. The three top things that teachers report spending time on are marking, lesson preparation and 'admin' (Adams et al., 2023).

All of these can be important and necessary. All are hard to reduce. But all can be done more efficiently than they often are, freeing time for other things that have more impact (Hamilton, Hattie, & Wiliam, 2023).

To give just one example, where a group of teachers teach the same content, particularly if they are likely to teach it more than once, it makes no sense for each of them to develop resources and schemes of work independently. In many cases, there will be high-quality resources available externally that they could all use. If not, collaboration—provided it is done well—is both more efficient in time and produces higher quality outputs. That does not mean that individual teachers cannot then adapt their lessons to their context; they should not have to, and certainly should not start from scratch. Many other suggestions and resources to help schools reduce workload can be found in the DfE's Workload Reduction Toolkit.

SUMMARY

Professional development is the strongest lever school leaders hold for increasing long-term student outcomes. Great leaders find ways to implement PD to maximise its impact, to prioritise long-term benefits over shiny quick fixes, and to make time for PD by reducing the time teachers spend on less effective things. Because we know that all these things are really hard to do, we have designed the Great Teaching Toolkit to offer PD structures and content, feedback tools for teachers and leaders, and a focus on effective classroom teaching and learning interactions that help all leaders to be Great.



Why aren't we doing 'instructional coaching' (even though everyone else seems to be)?

Instructional coaching seems to be the next big thing in education; it is the latest craze to ride a wave of enthusiasm across our schools. Faced with more and more schools 'nailing their colours to this mast', it feels almost brave to be holding back. But is it courageous and freethinking, or foolish and pig-headed, to go against the crowd on this?

In this article, we'll outline the thinking and rationale behind the flavour of coaching we have adopted within the Great Teaching Toolkit (GTT). But before addressing this in more depth, let us first try to define what we mean by instructional coaching. We may first note that instruction has quite different meanings in the UK and elsewhere. In the UK it means telling someone else what to do, and often has connotations of control and interference; in the US it simply means what in Britain would be teaching or pedagogy.

However, the confusion does not stop there. The same phrase, instructional coaching is used by different groups to refer to very different things. At one extreme, we have coaching guru Jim Knight (2019), whose version of instructional coaching emphasises teachers' self-determination, autonomy, reflection and empowerment. At the other, Paul Bambrick-Santoyo (2018) uses the same words for an approach where a relatively directive coach identifies granular 'action steps' and prescribes a set of 'coaching moves'. Others may use the same words to mean something between the two, or quite different from both—confusing, to say the least!

Nevertheless, there is good evidence to support the use of coaching as an approach to teacher development that leads to improved student outcomes. A meta-analysis by Kraft et al. (2018) found an overall positive effect of teacher coaching programmes—an effect that potentially out-performs other forms of teacher development or school-based interventions. This is a good study that presents solid evidence of effectiveness. An evidence-based practitioner would certainly want to look at coaching seriously. So, why are we not riding this wave?



RESEARCH-PRACTICE GAP

Kraft et al. (2018) found that larger programmes, and those with non-volunteer recruits, have smaller effects, and pointed to the challenges for recruiting effective coaches. Indeed, they cautioned against seeing coaching as a general solution, "It may be that coaching is best utilized as a targeted program with a small corps of expert coaches working with willing participants and committed schools rather than as a district-wide PD program" (p. 574).

The researchers were unable to say which approaches to coaching are most effective, though there is some evidence that the expertise of the coach is a key driver (Blazar & Kraft, 2015). There is also some evidence that more effective coaches may also be more effective teachers (Blazar et al., 2022; Goldhaber et al., 2020).

The coaching models that have been evaluated generally depend on a supply of expert coaches, with experience and training, each working with a relatively small number of teachers on a one-to-one basis, over an extended period. One study (D. S. Knight, 2012) estimated the cost per coached teacher as up to \$5,000—there are not many schools whose CPD budget can run to that!

The models that are being widely adopted in schools are quite different. In these ‘school-led coaching’ approaches, the coaches are teachers within the school, often senior staff. Leaving aside the question of whether those school leaders are also the most effective teachers or coaches, there are two obvious problems with this. One is the astronomical cost. It may be common in schools not to see people’s time as a direct cost, but of course it is. The other is the displacement. If we take all our best teachers out of the classroom to become coaches, any impact they have on the effectiveness of the coached teachers has to be offset by the loss to their own students’ learning. When we consider all that, instructional coaching does not seem like such a strong bet.

A SCALABLE, COST-EFFECTIVE APPROACH TO ‘COACHING’

One of the design principles behind the Great Teaching Toolkit is that it has to be scalable: something that any school can do, that generates maximum impact for minimal cost and time. There is good evidence that one-to-one coaching by an expert is one of the most effective ways to improve teaching—but this ‘Rolls-Royce’ model is not a scalable approach. If we try to unpick the mechanisms by which coaching supports improvement, we may be able to adapt it to create something that can be of wider benefit to every teacher across a school.

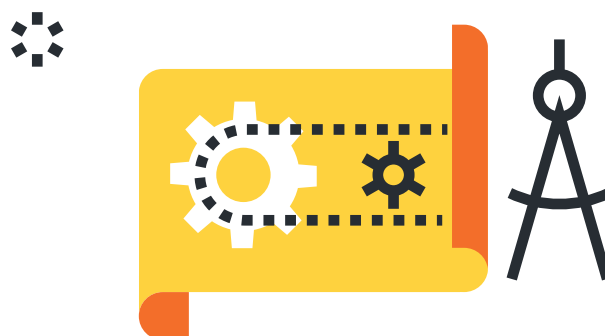
Part of the attraction of coaching is its stark contrast with the ‘inspire and forget’ training sessions (generic topic, same input for all, one-off event, conceptual rather than practice focus) that are often the mainstay of school-based CPD. Obvious differences are that coaching focuses on practice and on solving problems that are salient in that teacher’s classroom. Coaching is sensitive to the individual context. The personal coaching relationship is both empowering and motivating. An expert coach can diagnose, structure activities and respond—just as an expert teacher does. When done well, coaching also mobilises four components that we know are important for all learning (but are hard to fit with the standard CPD approach): feedback, modelling, reflection, and deliberate practice.

Our challenge was to come up with an approach that retains all those elements but is feasible and efficient. We can still call it coaching because it contains all the same active ingredients—and because the definition of coaching is so loose and contested. Our Great Teaching Teams model uses reciprocal and group-based: teams of four to six teachers who coach each other, focusing explicitly on their classroom practice and their individual challenges.

The social aspect of the team motivates and encourages. We deliberately build expertise within the team with tailored courses and structured reflection. We provide feedback that identifies excellence, highlights progress, and grounds everything in a reality-check. We use video to make observation efficient, practical and powerful, as well as to share models of excellence. We invoke the principles of deliberate practice (Deans for Impact, 2016), including challenging and specific goals, ‘decompositions’ and ‘approximations’ of practice, and feedback and developing mental models.

Another key design principle (perhaps our secret weapon and a unique feature of the Great Teaching Toolkit) is that we systematically test and learn. We know that the way we configure, structure and advise gets interpreted and adapted when schools take it on, often turning it into something unrecognisable. We are not affronted or surprised by these adaptations; instead we work with teachers to understand them, to pave the paths they choose to walk, or, if necessary, to refine the choice architecture to better align practice and theory. Crucially, we build evaluation in at every stage, so that over time, with the collaboration of teachers and school leaders, we can learn which inputs give the best output.

We are confident the coaching model we ultimately converge towards will be scalable, efficient and effective. But we know it will not be the same as our current, evidence-based, best bet. If you would like to help us learn how to make coaching work best in classrooms and staff rooms in every context, then join us on that journey.





Teacher collaboration: Why we're for it (even when the evidence is weak)

A collaborative environment, on its surface, seems like an obviously good thing for any school. Who doesn't like the idea of colleagues working together to increase student learning?

But as evidence-informed practitioners, we prefer a rationale grounded in research that our efforts are indeed "best bets" for teachers' time. Here, the evidence on teacher collaboration (a term which lacks a single, clear definition) is somewhat mixed. While some studies (e.g., Vescio et al., 2008) are often cited in support, the research landscape overall is weak and unhelpful about effective forms of collaboration. There's a lack of validation of measures, limited replication, absence of strong causal evidence, and an over-reliance on retrospective assumptions or rationalisations of "success."

So why, then, does the Great Teaching Toolkit CPD package (rather than just the Evidence Review) encourage teacher collaboration?

We have tried to think harder about collaboration, and understand what exactly it is about teacher collaboration that can lead to improved outcomes. From there, it is more apparent why we decided teacher collaboration has a place in the GTT.

A key ingredient for teacher collaboration is a focus for the group. The group's time needs to be structured with a clear goal, and one that has practical relevance to teaching (and student learning). Unstructured chat with colleagues may feel helpful, but is less likely to bring about improved practice. For this reason, the GTT guides teams of teachers through specific processes like identifying, exploring, and integrating elements of great teaching.

We know that students learn best when teachers create a supportive environment, with a climate of high challenge and also high trust. Teachers' learning benefits in the same way; teacher collaboration that is simultaneously supportive and

challenging helps increase the effectiveness of collaboration.

By virtue of comprising multiple people, collaborative learning benefits from the collective expertise of the teachers in the group. Group members may arrive with their own prior knowledge, or they develop expertise through exploration of an element of great teaching. Collaboration encourages this expertise to be shared for the benefit of the whole group; all teachers can be brought up to the level of the best. This is the ultimate aim of the GTT. The courses underpin this, by clarifying what changes in practice are most likely to impact student outcomes.

To round out our rationale of teacher collaboration, let us return to our initial reaction. It sounds like a nice thing and feels like a nice thing simply because it is a nice thing. Collaboration builds social bonds between colleagues. We feel attached to groups when we share a focus; we feel motivation for our goal, and our group's success; collaborative groups promote and sustain buy-in. Aside from the learning gains that come from strong group affinities and motivation, it may support staff wellbeing and retention. These are ends in their own right as part of a positive school environment, but they also can lead to further gains in staff development.

Ultimately, there is probably not a single model of teacher collaboration that works best—but where there is quality collaboration, teachers seem to be more effective and improve faster than others (Ronfeldt, 2017).

Teacher collaboration is not a panacea. Like everything else in education, it can work; it depends how you do it. We have created a particular model to leverage collaboration, in line with the best available evidence. Almost certainly, it can be improved—and it will be, as we collaborate with teachers to learn more about how they use it. But for now, we are optimistic and excited about this feature of the Great Teaching Toolkit.

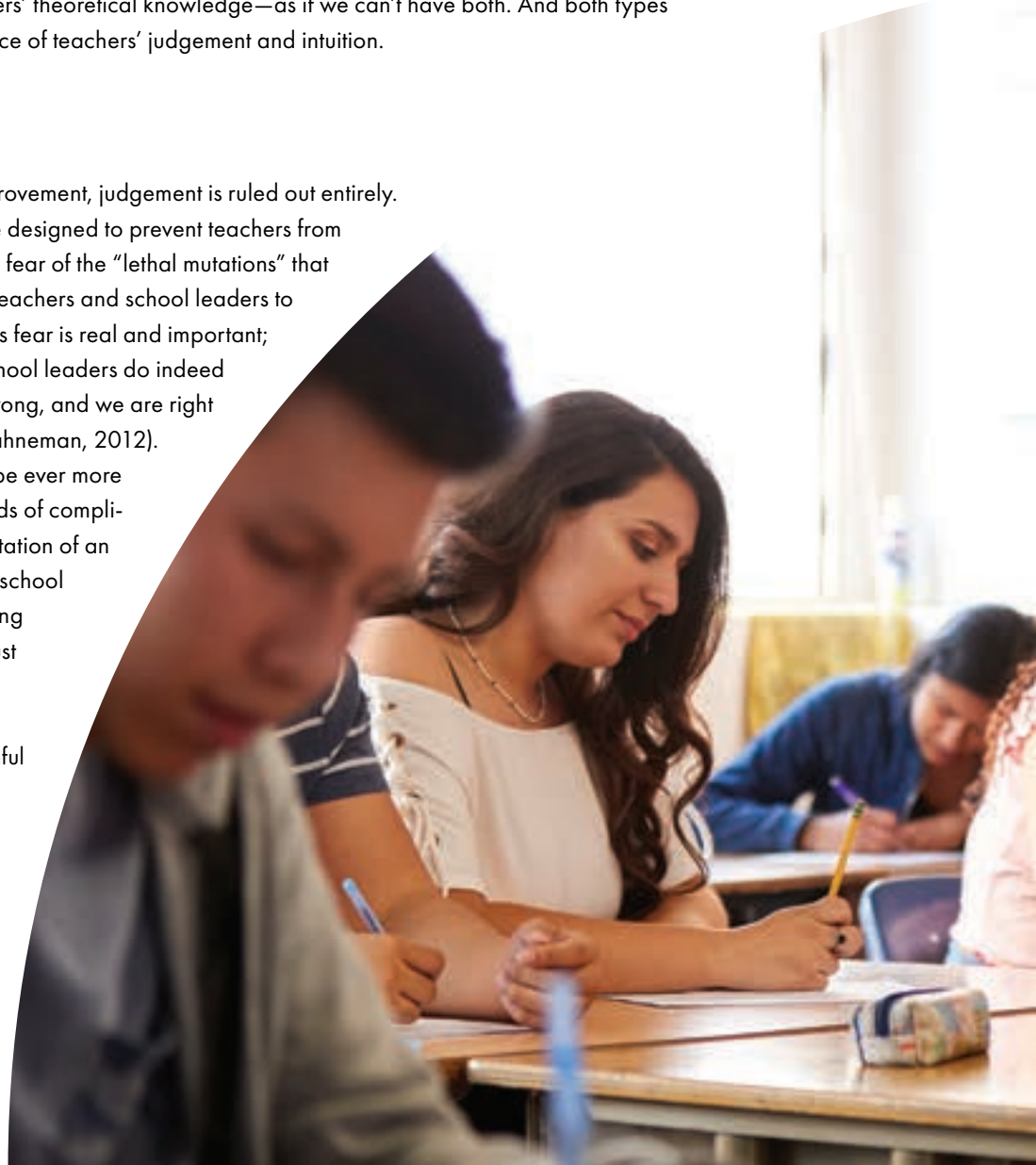
Why do we see teacher expertise as needing knowledge, skill and judgement (even though most people seem to focus on just one of these)?

In their chapter in the second edition of the Cambridge Handbook of Expertise and Expert Performance, Stigler and Miller claim that “most teacher education and professional development programs for teachers have focused on making teachers more knowledgeable; few actually give teachers opportunity to practice the skills of teaching” (2018, p. 441). Recently, with the growth of interest in coaching and teaching techniques, some approaches to teachers’ development have moved in the opposite direction to focus on classroom skills more explicitly. But these often then seem to place less emphasis on teachers’ theoretical knowledge—as if we can’t have both. And both types generally say little about the importance of teachers’ judgement and intuition.

THE NEED FOR JUDGEMENT

In some approaches to education improvement, judgement is ruled out entirely. Interventions require “fidelity” and are designed to prevent teachers from deviating from the script, driven by the fear of the “lethal mutations” that often arise when we allow individual teachers and school leaders to make their own choices. Of course, this fear is real and important; given a choice, many teachers and school leaders do indeed make bad choices. Intuition is often wrong, and we are right to be sceptical of where it leads us (Kahneman, 2012). So, a plausible approach is to prescribe ever more tightly, specifying and policing the kinds of compliance required to get faithful implementation of an “evidence-based” programme. Some school leaders also follow this route, mandating “non-negotiables” that all teachers must comply with.

Unfortunately, such approaches to faithful programme implementation have been found to have small effects at best (Lortie-Forgues & Inglis, 2019). Aside from their impoverished view of the complexity of teaching and undermining of teachers’ autonomy—with consequent negative effects on motivation and joy in work—attempts to improve practice by tightening compliance are, on average, simply not very effective.



(It should be noted that there may be some groups of teachers or contexts for which these prescriptive approaches are more effective. In particular, the expertise reversal effect [Kalyuga et al., 2003], whereby novices generally need more structure than experts, may be relevant here.)

In the Great Teaching Toolkit (GTT), we see teachers' judgement and intuition as crucial. So much of the complexity of classroom interaction depends on subtle choices teachers make, mostly below the level of conscious awareness. No one can observe classroom practice without being struck by the importance and intricacy of context; the very same practice can be right in one situation and wrong in another. It follows that the way teachers adapt techniques to their context is a feature, not a bug.

Adaptation is the lifeblood of effective teaching (Hatano & Inagaki, 1986). Intuition becomes something we have to work with, to develop and celebrate, not to squash.

And, it turns out, this is a feature of most other kinds of expertise too (Ericsson, 2018). Experts see things differently: they focus on the underlying patterns and structure in a situation, not its surface features; they pay attention to what really matters; they internalise the causal mechanisms that determine outcomes and act on them directly; they predict consequences and evaluate their own impact against these sophisticated counterfactuals; in doing this, they draw on a depth of experience, recognising patterns, constantly formulating, testing and developing explanatory theory (Ericsson, 2018; Stigler and Miller, 2018). All these thought processes can be characterised as "intuition" (Hogarth, 2001).

In short, developing teachers' judgement and intuition is crucial to learning to be more effective.



A BALANCE BETWEEN KNOWLEDGE AND SKILLS

Underpinning this intuition is an integrated balance of theoretical knowledge and practical skills.

In the GTT, what teachers need to know is grounded in research evidence. Specifically, the model for Great Teaching that we set out in our evidence review (Coe et al., 2020) provides the structure (see editor's note, below). Great Teachers understand the evidence and formal theory that explain the importance of:

- 1. Understanding the content they are teaching and how it is learnt;**
- 2. Creating a supportive environment for learning;**
- 3. Managing the classroom to maximise opportunity to learn;**
- 4. Presenting content, activities and interactions that activate their students' thinking.**

But just understanding the formal theory is not enough: teachers have to be able to apply it in practice in their context and connect their own experiences to it. This more personalised, applied theory corresponds to the notion of a "mental model" that features in most versions of "deliberate practice" (Deans for Impact, 2016; Ericsson & Pool, 2016).

Part of the justification for the importance of knowledge is the need for adaptation, outlined above. If teachers just learn a technique or skill, they may have the kind of "routine expertise" (Hatano & Inagaki, 1986) that allows them to perform this technique effectively under standard conditions. But to be able to adapt and apply it in a different situation they need "adaptive expertise": an understanding of why, when, how and with what it should be used or modified.

The development of techniques is also a specific focus of the GTT. Sometimes people worry that focusing on techniques may be atomistic and oversimplistic, reducing the complex art of teaching into a set of decontextualised skills. Actually, we learn complex skills most effectively by breaking them down, using "decompositions of practice" (Grossman et al., 2009) to identify, isolate and practise specific elements of classroom teaching. The learning from such practising may be accelerated by using "approximations of practice"—simulations or rehearsal opportunities that are simpler and lower-stakes than the real thing, to scaffold the learning (Deans for Impact, 2016).

Both decompositions and approximations are key elements of deliberate practice (Deans for Impact, 2016). Deliberate practice also requires the development of skills to be supported by challenging and specific goals for improvement, as well as feedback to inform learning. Both goals and feedback are a core part of the GTT environment.

Although teaching skills can be developed and practised in this artificial way, those skills then need to be applied, incorporated and embedded in the classroom. Again, the GTT provides for this, scheduling practice in context to ensure skills become integrated, fluent and automatic.

Further support for this balanced view of expertise is found in the review by Sims et al. (2022) of the impact of CPD. They find that programs that instil insight, motivate goals, teach techniques and embed practice have slightly bigger effects than those that do not do all four. (There are, however, caveats to this support, as the small difference may not be clear, given the number and differences of the studies.)

CONCLUSION

If teacher expertise is the strongest determinant of student achievement, and the main function of professional development is to develop expertise, then how we think about expertise really matters. Expertise does not just grow naturally with experience, but it can be developed—given the right conditions. Expertise requires a balance of knowledge, skills, and judgement. To be effective, professional development needs to address all three.

Designing a CPD package that systematically develops the most powerful knowledge, skills and judgement for all teachers—and then helps them to embed that into everyday practice—is a challenging task. For a school to create something this complex and of high quality is pretty much impossible. Fortunately, the Great Teaching Toolkit does it all, in a way that is flexible and easy to use. Find out more here.

(Editor's note: Sometimes people think the GTT is the Evidence Review, perhaps because it was the first part we published. But the GTT is a genuine toolkit of resources, including evidence summaries, courses, feedback instruments, structures for collaboration, and more. In other words, a comprehensive CPD package.)



Why are we betting on giving teachers more effective feedback (even though no one else is really doing this)?

Great teachers know about the power of feedback, and research supports this. In the Great Teaching Toolkit: Evidence Review (Coe et al., 2020), we summarised the evidence about feedback in both directions: giving students feedback to guide their learning and getting feedback from students to make teaching responsive.

But feedback doesn't just help school students. A classic review by Kluger and de Nisi (1996) of the impact of feedback on performance found positive effects on average in a wide range of contexts. However, this and later reviews also demonstrated a wide variation of effect sizes, including many cases where feedback harms performance. The research on when feedback is most helpful is complex and hard to interpret easily (e.g., Shute, 2008; Wisniewski et al., 2020). On its own, feedback does not necessarily enhance performance, unless it is used to promote learning or motivation, and effects are greater if recipients are supported to implement changes.

Part of the power of feedback is that it provides a "reality check". Despite our tendency as human beings to believe we can judge how well we are doing things, we are generally wrong. "The correlation between self-ratings of skill and actual performance in many domains is moderate to meagre" (Dunning et al., 2004, p. 69). In situations where we do not have good feedback, self-assessments of performance are inaccurate and mostly over-optimistic. To put it bluntly, we're all probably worse at things than we think we are.

EXISTING FEEDBACK IS A POOR GUIDE

Classroom teachers do get some feedback. Teachers constantly evaluate how well a lesson is going, looking for signs of confusion or for flagging interest among students, for example. But, as I wrote a decade ago (Coe, 2013), the things that are visible in classrooms are mostly "poor proxies" for student learning. Perceptions of our own role are also hugely distorted, as shown by the surprise—even shock—most teachers experience on seeing themselves teaching on video.

Many teachers do also get feedback from observations by colleagues. However, as I wrote in 2014, most of the judgements made by observers without specialist training are wrong (Coe, 2014). Even without scoring or rating the lesson, if the judgement that underpins any feedback is wrong, the feedback is unlikely to be helpful.

WHY ISN'T FEEDBACK CENTRAL TO ALL PROFESSIONAL LEARNING?

Given the power of feedback, its particular importance in learning and improvement in complex tasks, and the poverty and paucity of easily generated feedback in classrooms, it may seem surprising that more attempts at education improvement have not featured feedback more prominently. Most teachers would struggle to imagine having to support their students' learning without giving and receiving feedback; yet in most models of professional learning, teachers receive very limited feedback about their performance and their planned learning is relatively unresponsive to their current status or progress. Coaching offers the potential to incorporate both kinds of feedback, but, as I have argued, the impact of coaching depends heavily on the scarce expertise of the coach. So this is unlikely to be an efficient, scalable approach on its own.

A number of studies have evaluated the impact of interventions that feature feedback to teachers to raise attainment of students (e.g, Kraft and Christian, 2021; van Geel et al., 2016). Although some do find positive effects, overall, the picture is mixed. There are three particular reasons identified why feedback may not lead to improvement that are salient:

- **Feedback can be brutal.** We are not as good as we think we are and ignorance is bliss. Even if we know that the natural feedback we currently receive is actually uninformative, it is still comfortable and reassuring. Plus, if feedback tells us we're not that good, we will have to do something about it (more work!). Therefore most people do not naturally seek out (or may avoid or disregard) helpful feedback.
- **Practical measures are hard to create.** Even when we want it, effective feedback is hard to get. The requirements of practical measures are demanding and the expertise to create them is thinly spread. In practice, we often depend on weak proxies instead—when we even bother.
- **Acting on feedback is hard.** Even when we receive high-quality feedback, the challenge of the hard work to implement and sustain a significant change remains.

IMPROVEMENT SCIENCE AND FEEDBACK

The prioritisation of feedback within the Great Teaching Toolkit (GTT) has also been influenced by an area of research that foregrounds the power of feedback: improvement science (Lewis, 2015). Improvement solutions must fit the context, hence the need to be developed and adapted by local actors—improving quality is the job of those who do the job. For it to be successful, plans are treated as hypotheses to be tested; feedback is collected about implementation and impact using “practical measurement”.

Practical measurement has four distinctive requirements (Yeager et al., 2013). First, it focuses on intermediate “leading” indicators and direct measures of the underpinning mechanisms, not just the final outcome. Second, it provides granular and specific information, not just the high-level constructs that are often the target of research measures. Third, it is designed to have meaning and salience for the people who own the change (e.g., classroom teachers). Fourth, it has to be manageable to collect and interpret in the context of everyday work. All told, these measurements drive effective feedback—so long as they are meaningful and practical. After all, “We cannot improve at scale what we cannot measure” (Bryk et al., 2015, Chapter 4).

MORE EFFECTIVE FEEDBACK IN THE GREAT TEACHING TOOLKIT

Conscious of the limitations of readily available feedback for teachers and school leaders, in the GTT we have prioritised the development of feedback tools that aim to make better feedback more easily available. As part of this process, we have identified four specific mechanisms by which feedback can help people to improve their performance:

1) Holding up a mirror

A key aim of our feedback tools is to help teachers to see their own classroom in a way that is broader, clearer, and more accurate than their raw experience can provide. They should be like holding a mirror up to allow teachers to see themselves. This gives them insights into their classroom that may already be readily actionable, especially if they have a sound mental model of what great teaching looks like. The feedback itself may not always provide all the “answers” on what to do next. These additional insights may require further support (for example, through collaboration) to draw out diagnostic interpretations and convert them into action.



2) Motivating improvement

If feedback is received on a repeated basis, it allows teachers to see the progress they are making. Being able to see that you are improving some aspect of your practice is hugely motivating. Given the investment required in professional learning, it is important that these feelings of self-efficacy, competence and improvement are supported.

But feedback can also motivate by drawing attention to a gap between actual and desired performance, particularly where individuals have self-efficacy: the perception of their competence to reduce the gap by improving performance. The availability of feedback (and the social pressure of its uptake by others) also drives a shift from teachers thinking about their teaching. It is no longer something that is “just what they do,” “for the children,” or “good enough.” Instead, teachers’ thinking shifts to a focus on their effectiveness as a thing that can be improved (no matter how good they already are); they can feel their agency and ownership of it.



3) Focusing attention on what matters

Feedback directs attention to key goals and outcomes (Kluger & deNisi, 1996; Locke & Latham, 2002). Knowing that an aspect of teaching is being captured, measured and fed back increases its salience. If the feedback tools focus on the right things, they increase the alignment between the aspects of practice that really matter for student outcomes—and also what matters to teachers and school leaders.

4) Clarifying what good looks like

A requirement for deliberate practice, and for developing expertise in general, is the development of mental models (Deans for Impact, 2016). Operationalising an element of great teaching into a well-specified and transparent measurement process helps to build a clearer shared understanding: a mental model of that aspect of great teaching. Without that, it is conceivable for colleagues to have a conversation about an aspect of practice—say, “great questioning”—using the same words but actually meaning quite different things. Understanding is further supported by having clear and rich descriptions, along with examples (including a wide range of examples, boundary cases, and non-examples with different characteristics) across the spectrum between exemplary and routine practice.

CONCLUSION

Feedback can be one of the most powerful ways to improve goal-directed performance, supported by a vast body of research and theory. However, it often fails to live up to its potential, and needs the right supports in place to work best. Feedback tools are a key component of the Great Teaching Toolkit; their design and implementation are guided by the best available evidence.

Moreover, there is a recursive twist to the way we have built these feedback tools: every time a teacher or school uses the GTT, we are getting feedback about how effective it is. In the same way we tell teachers how feedback can help them be even better, we are using that feedback to help the GTT be even better.

Feedback really is the key.

REFERENCES

- Adams, L., Coburn-Crane, S., Sanders-Earley, A., Keeble, R., Harris, H., Taylor, J., & Taylor, B. (2023). Working lives of teachers and leaders -- wave 1.
- Bambrick-Santoyo, P. (2018). *Leverage Leadership 2.0*. Wiley.
- Blazar, D., & Kraft, M. A. (2015). Exploring mechanisms of effective teacher coaching. *Educational Evaluation and Policy Analysis*, 37(4), 542–566. <https://doi.org/10.3102/0162373715579487>
- Blazar, D., McNamara, D., & Blue, G. (2022). Instructional coaching personnel and program scalability (499; 21).
- Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Harvard Education Press. <https://books.google.co.uk/books?id=CKZhDwAAQBAJ>
- Burgess, S., Hauberg, D. S., Rangvid, B. S., & Sievertsen, H. H. (2022). The importance of external assessments: High school math and gender gaps in STEM degrees. *Economics of Education Review*, 88(April), 102267. <https://doi.org/10.1016/j.econedurev.2022.102267>
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, 104(9), 2633–2679. <https://doi.org/10.1257/aer.104.9.2633>
- Coe, R. (2013). Improving education: A triumph of hope over experience. <https://profcoe.net/>
- Coe, R. (2014, January 9). Classroom observation: It's harder than you think. CEM Blog. <http://www.cem.org/blog/414/>
- Coe, R., Rauch, C. J., Kime, S., & Singleton, D. (2020). Great teaching toolkit: Evidence Review. Evidence Based Education. <https://evidencebased.education/great-teaching-toolkit-evidence-review/>
- Deans for Impact. (2016). Practice with purpose: The emerging science of teacher expertise.
- Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment. *Psychological Science in the Public Interest*, 5(3), 69–106. <https://doi.org/10.1111/j.1529-1006.2004.00018.x>
- Ericsson, K. A. (2018). An introduction to the second edition of *The Cambridge Handbook of Expertise and Expert Performance: Its development, organization, and content*. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 3–20). Cambridge University Press. <https://doi.org/10.1017/9781316480748.001>
- Ericsson, K. A., & Pool, R. (2016). *Peak: Secrets from the new science of expertise* (NV-). Houghton Mifflin Harcourt.
- Fletcher-Wood, H., & Zuccollo, J. (2020). The effects of high-quality professional development on teachers and students: A rapid review and meta-analysis.
- Goldhaber, D., Krieg, J., & Theobald, R. (2020). Effective like me? Does having a more productive mentor improve the productivity of mentees? *Labour Economics*, 63, 101792. <https://doi.org/10.1016/j.labeco.2019.101792>
- Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. W. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record*, 111(9), 2055–2100. <https://doi.org/10.1177/016146810911100905>
- Hamilton, A., Hattie, J., & Wiliam, D. (2023). *Making room for impact: A de-implementation guide for educators*. Corwin.
- Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Child development and education in Japan* (pp. 262–272). W.H. Freeman and Co.
- Hattie, J. (2003). Teachers make a difference: What is the research evidence? ACER Research Conference. https://research.acer.edu.au/cgi/viewcontent.cgi?article=1003&context=research_conference_2003
- Hogarth, R. M. (2001). *Educating intuition*. Chicago.
- Kahneman, D. (2012). *Thinking, fast and slow*. Penguin Books London.

- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254.
- Knight, D. S. (2012). Assessing the cost of instructional coaching. *Journal of Education Finance*, 38(1), 52–80. <http://www.jstor.org/stable/23259121>
- Knight, J. (2019). Why teacher autonomy is central to coaching success. *Educational Leadership*, 77(3). <https://www.ascd.org/el/articles/why-teacher-autonomy-is-central-to-coaching-success>
- Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, 88(4), 547–588. <https://doi.org/10.3102/0034654318759268>
- Kraft, M. A., & Christian, A. (2021). Can teacher evaluation systems produce high-quality feedback? An administrator training field experiment. In *EdWorkingPapers.com* (62; 19). <https://doi.org/10.26300/ydke-mt05>
- Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*, 36(4), 476–500. <https://doi.org/10.3102/0162373713519496>
- Lewis, C. (2015). What is improvement science? Do we need it in education? *Educational Researcher*, 44(1), 54–61. <https://doi.org/10.3102/0013189X15570388>
- Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9), 705–717. <https://doi.org/10.1037/0003-066X.57.9.705>
- Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? *Educational Researcher*, 48(3), 158–166. <https://doi.org/10.3102/0013189X19832850>
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458. <https://doi.org/10.1111/j.1468-0262.2005.00584.x>
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2), 247–252. <https://doi.org/10.1257/0002828041302244>
- Ronfeldt, M. (2017). Better collaboration, better teaching. In E. Quintero (Ed.), *Teaching in Context: The Social Side of Education Reform*. Harvard University Press.
- Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Sims, S., Fletcher-Wood, H., O'Mara-Eves, A., Cottingham, S., Stansfield, C., Goodrich, J., Herwegen, J. Van, & Anders, J. (2022). Effective teacher professional development: New theory and a meta-analytic test (507; 22). <https://doi.org/10.26300/rzet-bf74>
- 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(5), 629–645. <https://doi.org/10.1111/j.1468-0084.2011.00666.x>
- Stigler, J. W., & Miller, K. F. (2018). Expertise and Expert Performance in Teaching. In *The Cambridge Handbook of Expertise and Expert Performance* (pp. 431–452). Cambridge University Press. <https://doi.org/10.1017/9781316480748.024>
- van Geel, M., Keuning, T., Visscher, A. J., & Fox, J.-P. (2016). Assessing the effects of a school-wide data-based decision-making Intervention on student achievement growth in primary schools. *American Educational Research Journal*, 53(2), 360–394. <https://doi.org/10.3102/0002831216637346>
- Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education*, 24(1), 80–91. <https://doi.org/10.1016/j.tate.2007.01.004>
- Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10, 3087. <https://doi.org/10.3389/fpsyg.2019.03087>
- Yeager, D., Bryk, A., Muhich, J., & Morales, L. (2013). Practical measurement (78712).

A smiling male teacher with glasses, wearing a light blue shirt and a patterned tie, is seated in a classroom. He is looking towards the left, where the back of a student's head is visible. The background shows a classroom wall with colorful letters 'D', 'M', and 'S' on it.

GREAT TEACHING MAKES A BIG DIFFERENCE

Research tells us that what teachers know and do matters more to student outcomes than anything else we can influence. The **Great Teaching Toolkit** makes it easier for all teachers to make a difference, however good they are.

evidencebased.education

