



ILLUSTRATION / NEIL WEBB

# TAKING TEACHER COACHING TO SCALE

## CAN PERSONALIZED TRAINING BECOME STANDARD PRACTICE?

**THE IMPORTANCE OF INDIVIDUAL TEACHERS** has emerged in sharp focus over the past decade, with compelling evidence that teachers have large effects on a range of student outcomes. Wide variability in teacher effectiveness, both across and within schools, highlights the persistent challenge of providing students with access to high-quality teachers. However, traditional efforts to increase teacher quality through professional development (PD) have been largely ineffective. That may be changing, as a new form of PD, teacher coaching, has emerged to disrupt the PD industry.

Historically, PD has been dominated by daylong seminars that took teachers out of the classroom and delivered the same tips and tricks to an entire department, grade level, or school. But as research has found, these programs to have little or no effect on teacher quality. Some training has shifted to a customized, smaller-scale approach: instructional coaching, whereby an expert mentor works one-to-one with teachers to provide a steady stream of feedback and suggest new techniques based on

frequent classroom observations. By the 2015-16 school year, 27 percent of public K-12 schools reported having a reading coach on staff, 18 percent had a math coach, and 24 percent had a general instructional coach, according to the National Teacher and Principal Survey.

Researchers have studied individualized coaching programs for decades, but only began to evaluate their effects using randomized control trials in the last dozen years. We set out to examine what this growing literature now says about the efficacy of teacher coaching as a development tool. Does one-to-one coaching help teachers get better? If so, how powerful a strategy might this be to improve teacher practice and student outcomes?

Our analysis of results from across 60 studies found that coaching works. With coaching, the quality of teachers' instruction improves by as much as—or more than—the difference in effectiveness between a novice and a teacher with five to 10 years of experience, a more positive estimated effect than traditional PD and most other school-based interventions.

by MATTHEW A. KRAFT and DAVID BLAZAR

However, larger coaching programs are less effective than smaller ones, raising questions about whether coaching can be brought to scale in a way that preserves its impact.

### Teacher Development Gets Personal

Public school systems in the United States spend billions of dollars annually on PD to help teachers meet the diverse needs of their students—with limited results. Most PD remains of the “sit and get” variety: one-off workshops delivered to large groups, with little obvious connection to the needs of individual teachers or classrooms. Rigorous studies find that PD programs more often than not fail to produce systematic changes in teachers’ instructional practice, much less improvements in student achievement, especially when implemented at scale.

Yet expectations for teachers have grown in recent years, as states have adopted new college- and career-ready standards and as education agencies increasingly emphasize the importance of balancing expert content delivery with nurturing the social-emotional skills that are also important for students’ lifelong success. Taken together, teachers’ expected roles range from content expert, curriculum developer, and pedagogue, to social worker,

to incorporate these practices into their classrooms. In contrast to traditional PD, coaching is intended to be individualized, time-intensive, sustained over the course of a semester or year, context-specific, and focused on discrete skills. Coaches engage in a sustained professional dialogue with teachers focused on developing skills to enhance their classroom practice; ideally, the specific skills targeted for development differ based on individual teacher needs.

### Examining the Teacher Coaching Literature

As researchers, we have worked to develop and evaluate several coaching programs, including the MATCH Teacher Coaching program operated by the eponymous Boston charter-management organization and the Mathematical Quality of Instruction Coaching program developed by Heather Hill and colleagues at the Harvard Graduate School of Education. The results of these studies were encouraging, particularly with respect to the degree to which the programs generated noticeable changes in teachers’ practice. Yet studies of discrete programs cannot, on their own, speak to the efficacy of coaching as a new model for teacher professional development. To address

---

**With coaching, the quality of teachers’ instruction improves by as much as—  
or more than—the difference in effectiveness between a novice  
and a teacher with five to 10 years of experience.**

---

psychologist, mentor, and motivator. Every teacher has dimensions of this interrelated skill set on which they can improve—a complex and dynamic reality reflected in the one-to-one coaching model, which seeks to align the support provided to individual teachers to their unique challenges and needs.

Most teacher-coaching programs share several key features, but no one set of features defines all coaching models. In our review of the literature, we encountered multiple, sometimes conflicting, definitions of teacher coaching. Some envision coaching as a form of implementation support to ensure that new teaching practices or teaching materials—often introduced in an initial group training session—are executed with fidelity. Others see coaching as a tool that enables teachers to learn and apply new pedagogical practices to support student learning. The role of the coach may be performed by a range of personnel, including administrators, master teachers, curriculum designers, external experts, and other classroom teachers.

Synthesizing this body of theoretical work, we characterize coaching as an observation and feedback cycle in which coaches model research-based practices and work with teachers

that broader question, we sought to synthesize results across the full body of research on instructional coaching programs.

We conducted a meta-analysis of the literature on coaching by collecting, coding, and analyzing the findings across all rigorous evaluations of teacher coaching in developed countries published through 2017. This first enabled us to estimate the average effect of all coaching programs—or at least all those that have been subjected to rigorous evaluation—on teacher practice and student achievement. We also used the same information to determine whether coaching programs with certain characteristics produce stronger results.

A meta-analysis is only as good as the underlying studies it aggregates. Ours includes only randomized controlled trials and quasi-experimental research designs that could credibly isolate the effect of coaching. We further restricted our review to studies that focus on two key outcome measures that we see as critical components in the theory of action linking coaching to increased student skill: measures of teachers’ instructional practice as rated by outside observers and direct measures of student achievement on standardized assessments.

In total, we identified 60 studies on teacher coaching that met these requirements. It is remarkable that such a rich set of empirical research has emerged over the last decade given that a landmark review in 2007 looking at *all* research on teacher PD found only nine studies that supported causal inferences.

In order to draw comparisons and synthesize the studies' findings, we rescaled their results to effect size units that measure the change in outcomes due to the coaching program in standard deviations—that is, relative to how much the relevant outcome varies across the teachers or students in the study sample. We also coded studies to track unique elements of the coaching models such as their size, their focus on content or teaching skill, whether they are paired with workshops or curriculum materials, and whether they were delivered in person or via videoconference platforms.

### Does Teacher Coaching Work?

Teacher coaching has large positive effects on both instructional practice and student achievement (see Figure 1). On average, coaching improves the quality of teachers' instruction and its effects on student achievement by 0.49 standard deviations and 0.18 standard deviations, respectively. For both outcomes, the magnitude of the effect of coaching is comparable to or exceeds the largest published estimates of the difference in performance between a novice teacher and an experienced veteran. Our estimates of the effectiveness of teacher coaching as assessed on these two outcome measures also compare favorably when contrasted with the larger body of literature on teacher PD, as well as most other school-based interventions.

These findings may come as a surprise given researchers' general inability to identify characteristics that differentiate highly effective from ineffective teachers. However, one exception to the disappointingly weak relationships between teachers' skill and their observable characteristics like certification, licensure, or even content knowledge is the quality of teachers' classroom practice. Teachers with strong behavior-management skills and the ability to deliver cognitively demanding, error-free content produce substantively and substantially larger student-achievement gains than other teachers without these skills. It should perhaps not be a surprise, then, that teacher coaching is able to

improve student outcomes because of the interventions' specific attention to teachers' core classroom practices.

Even so, our analyses suggest that noticeably improving student achievement likely requires large improvements in teachers' instructional practice; the observed improvement in instructional practice due to coaching is significantly larger than the resulting impact on student outcomes (see Figure 2). This may explain why other PD programs such as generalized workshops, which may produce more moderate improvements on intermediate outcomes such as teacher knowledge or classroom practice, do not have similar effects on student outcomes.

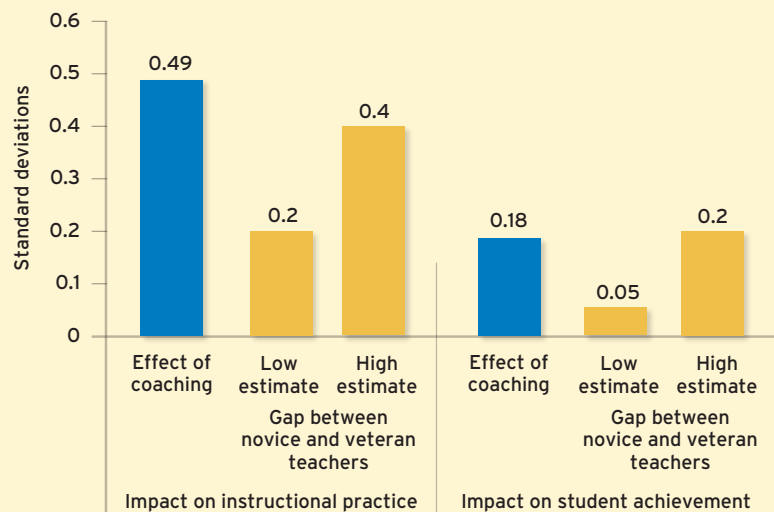
Teacher coaching is a rare model of PD that has been shown to improve teacher practice to the degree required to impact student-achievement outcomes. However, even here, relatively large improvements for teachers turn into much more moderate gains for students.

### Taking Coaching to Scale

Although these findings demonstrate the potential of coaching as a development tool, questions remain about the features

## Teacher Coaching Improves Instructional Practice, Student Achievement (Figure 1)

*On average, teacher coaching raises the quality of teachers' instructional practice and their impact on student achievement by 0.49 standard deviations and 0.18 standard deviations, respectively—as much as or more than the differences observed between a novice teacher and an experienced veteran.*



SOURCE: Authors' calculations

of effective coaching programs and the feasibility of providing coaching more broadly. Do schools have enough expert teachers who can serve as coaches across content areas? If not, where might schools find coaches? Will PD budgets support the relatively high costs of implementing coaching with fidelity?

Our analysis of the relationship between various program characteristics and their impacts is able to address some of these questions. Surprisingly, we find little evidence that coaching “dosage”—that is, the number of times teachers and coaches meet—is associated with the effectiveness of a given coaching program. We interpret this descriptive finding to mean that, when comparing *across* coaching programs, quality matters more than quantity. Coaching models that build in frequent observation and feedback cycles are not uniformly better; other program elements such as coach quality matter, too. We speculate, however, that for an *individual* coaching program of fixed quality, it is likely better to have more coaching cycles, not fewer.

Further, we find little difference in the effectiveness of coaching programs delivered online versus face to face. This suggests that

schools that lack in-house coaches are still able to implement coaching programs through the use of digital video recorders to capture instruction and online videoconferencing to interact with coaches. Although this technology is not cheap, the cost of these tools has dropped rapidly in recent years, and the technology could support both teacher PD and evaluation efforts.

These findings show the potential feasibility of expanding teacher coaching across schools and districts, but other results show how difficult maintaining program fidelity may be. Looking at the size of coaching programs, we find that the average effectiveness of the coaching program declines as the number of teachers involved increases, suggesting the difficulty of successfully taking such programs to scale. Our analyses of both instruction and achievement depict a clear negative relationship between program size and program effects, consistent with a theory of diminishing effects as programs are scaled up.

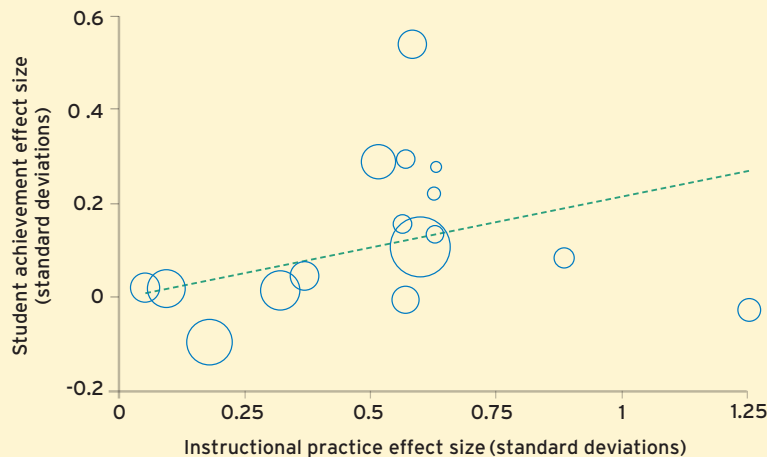
We see similar patterns when we test more formally for evidence of potential scale-up implementation challenges by comparing effect sizes between two types of studies: those with

fewer than 100 teachers and those with 100 teachers or more (see Figure 3). The average effects in larger studies are only one-third to one-half as large as those found in smaller studies. Additional analyses confirm that these differential results are not driven by a pattern in which studies of smaller coaching programs with small or no effects are less likely to be published because of their limited precision.

## Achievement Gains Require Major Instructional Improvement

(Figure 2)

*Coaching programs that produced larger improvements in teacher practice tend to show larger effects on student achievement. Even so, relatively large improvements in instructional practice translate into more moderate gains in achievement.*



NOTE: Data points are calculated by averaging across all effect sizes for a given outcome from the same research project. Dashed line is a regression line weighted by the average sample size of each study.

SOURCE: Authors' calculations

## Key Considerations for Scaling Up

In our view, the growing body of research on teacher coaching provides strong evidence of its effectiveness as a development tool. However, our meta-analysis also raises difficult questions about whether and how to implement coaching programs at scale. Several factors likely contribute to the diminishing returns to coaching as the size of programs increases, including coach quality, financial constraints, standardization, and teacher engagement and school climate.

**Coach quality:** A fundamental challenge to scaling up coaching programs is finding enough expert coaches able to deliver these services. After all, coaches *are* the intervention. Most of the studies we examine had only a handful of coaches, many of whom were key program staff or even program developers. Scaling up from

**Teacher coaching is a rare model of PD that has been shown to improve teacher practice to the degree required to impact student achievement outcomes.**

a small corps of coaches to a large staff requires new systems for recruiting, selecting, and training coaches. These systems are still largely underdeveloped in most contexts. Research that seeks to understand the characteristics and skills of effective coaches (such as teaching/coaching experience, content knowledge, and rapport with teachers) can aid in the development of these systems.

**Financial constraints:** Teacher coaching is a relatively expensive form of PD due to the large personnel costs of hiring coaches who meet with teachers on a regular basis. There are very few economies of scale available when the primary intervention is one-to-one interaction. Efforts to scale up coaching often lead to programmatic changes to cut costs, such as having coaches meet less frequently with each teacher or even coaching teachers in small groups. While we do not have definitive evidence on the effect of these adaptations, we suspect that they may decrease the efficacy of coaching as a PD tool.

**Standardization:** Scaling up coaching can require building more formal sets of systems and structures to ensure program fidelity, which may have the unintended consequence of constraining a coach's ability to tailor her approach to the individual needs of each teacher. Because coaching is by definition differentiated, we see a need for program developers to think critically about how they can implement organizational structures and systems that provide scaffolded supports to individual coaches without restricting their judgment and flexibility.

**Teacher engagement and school climate:** Bringing coaching to scale likely would include a prescriptive approach, requiring teachers who may be hesitant or resistant to engage in the coaching process to take part. This may be understandable given an expanded emphasis on linking scores from classroom observation rubrics to high-stakes job decisions. However, coaching is unlikely to be successful without teachers' openness to feedback and willingness to adapt their practice. Here, school

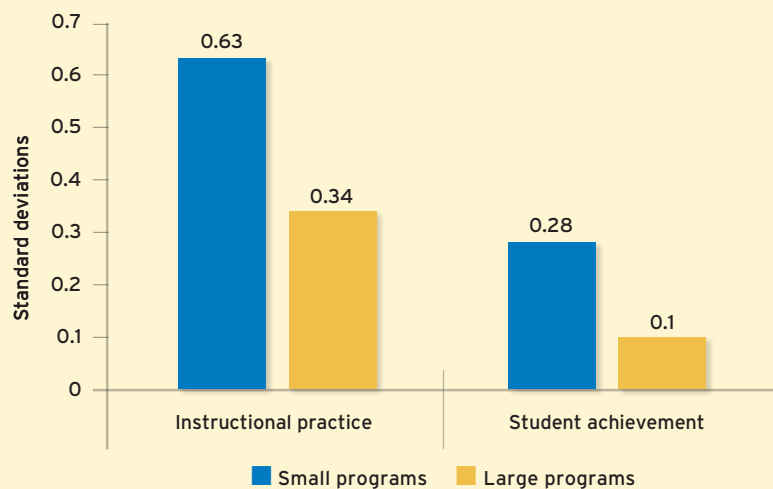
leaders have a key role to play in creating a culture of trust and respect among administrators and staff in order to ease teachers' concerns and increase their willingness to actively engage.

**Looking Ahead**

We see real potential for coaching programs to innovate and address many of these challenges. As an inherently customizable intervention, coaching may be well suited to meeting a variety of teacher-development needs. For example, new technologies are powering distance or virtual programs, which draw on coaches from afar to provide specialized development to teachers in small and rural districts who may not ordinarily be partnered with instructional experts in their specific grades and subject areas. Coaching also is being paired with

**Larger Programs Less Effective (Figure 3)**

*Studies of coaching programs with samples of 100 teachers or more generated average effects only one-third to one-half as large as studies with samples of fewer than 100 teachers, consistent with a theory of diminishing effects as programs are taken to scale.*



SOURCE: Authors' calculations

## A fundamental challenge to scaling up coaching programs is finding enough expert coaches able to deliver these services.

computer-simulation-based student teaching, which allows teachers to teach a lesson, receive feedback, and immediately try it again. Finally, emerging peer coaching models present a promising approach to creating observation and feedback cycles that leverage expertise within a school building, by pairing up teachers with different strengths and weaknesses to observe each other's practice and provide suggestions.

As researchers and practitioners continue to develop and refine coaching programs, we encourage them to consider the delicate balance between efficiency and efficacy. Coaching in all forms is a resource-intensive intervention that requires fairly sizable investments, both in terms of money and staff. Expanding coaching will require policymakers and administrators to engage in critical conversations about how current expenditures on PD could be used more effectively. For example, one approach may be to reallocate some PD spending to provide high-cost but effective PD programs like coaching to schools or teachers most in need of support, rather than uniformly providing less-effective and

less-expensive traditional PD for all schools and teachers.

Ultimately, strengthening the teacher workforce will require improving the classroom performance of individual teachers. Given the decades of investment in traditional PD for relatively small returns, policymakers and educators should support innovation in this sector. Coaching can provide a flexible blueprint for these efforts, but questions remain about the factors and local contexts that can influence its effectiveness. It remains to be seen whether coaching is best implemented as smaller-scale targeted programs tailored to local contexts, or if it can be taken to scale in a high-quality and cost-effective way.

*Matthew A. Kraft is an associate professor of education and economics at Brown University. David Blazar is an assistant professor of education policy and economics at the University of Maryland, College Park. The full meta-analysis on which this article is based is available at the Review of Educational Research.*

AD

AD