

Abstract Title Page
Not included in page count.

Title:

Improving Teacher Practice:
Experimental Evidence on Individualized Teacher Coaching

Authors and Affiliations:

Matthew A. Kraft
David L. Blazar
Harvard Graduate School of Education

Abstract Body

The term ‘professional develop’ includes activities that . . . (iv) improve classroom management skills; (v)(I) are high quality, sustained, intensive, and classroom-focused in order to have a positive and lasting impact on classroom instruction and the teacher’s performance in the classroom; and (II) are not 1-day or short-term workshops or conferences.

No Child Left Behind 2001 – Title IX – General Provisions – SEC 9101 Definitions

Background / Context:

For over a century, school systems in the U.S. have attempted to improve instructional quality by investing in the education and training of their teachers. Today, over 90% of teachers report participating in some form of PD, with districts spending between \$2,000-\$8,000 annually *per teacher* (Darling-Hammond, Wei, Andree, Richardson, & Orphanos, 2009; Miles, Odden, Fermanich, Archibald, & Gallagher, 2011). Theory suggests that professional development (PD) can increase student achievement by improving teacher knowledge and skills (Scher & O’Reilly, 2010; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). However, research on specific PD programs indicates that most fail to produce systematic improvements in teacher effectiveness (Garet, Cronen, Eaton, et al., 2008; Garet, Porter, Desimone, et al., 2011; Harris & Sass, 2011; Yoon et al., 2007). These results, combined with studies concluding that teacher effectiveness does not meaningfully improve after the first few years on the job (Rockoff, 2004; Rivkin, Hanushek, & Kain, 2005), have raised doubts about the potential for PD to improve instructional quality at scale.

Practitioners have responded to critiques of PD by re-envisioning it in the form of individualized and sustained teacher coaching programs that provide tailored feedback to teachers about their classroom practices. This approach has been most widely adopted in PD for literacy instructors, such as Reading First, a large federal program. However, the current literature on coaching is largely descriptive and documents inconsistent evidence of the relationship between literacy coaching, teachers’ instructional practices, and student achievement (Darling-Hammond et al., 2009). To our knowledge, there exists only one evaluation of a teacher-coaching program that uses a rigorous experimental design. Allen, Pianta, Gregory, Mikami, and Lun (2011) find that the My Teacher Partner-Secondary program increased student achievement by 0.22 standard deviations on state standardized tests in the post-intervention year.

Purpose / Objective / Research Question / Focus of Study:

In this paper, we build on the work of Allen and his colleagues by evaluating the effect of a time-intensive, individualized coaching program. MATCH Teacher Coaching (MTC) focuses on improving teachers’ classroom management and instructional practices, both of which are cited as important mechanisms behind improvements in student outcomes (Freiberg, Huzinec, & Templeton, 2009; Kane, Taylor, Tyler, & Wooten, 2010). In contrast to the My Teacher Partner-Secondary program, MTC coaches observe live instruction and work in person with teachers to help them improve their practice.

In May of 2011, we recruited 59 teachers, working in Recovery School District charter schools across New Orleans, to participate in a randomized trial of the year-long coaching program. Detailed coaching logs allow us to describe the key focus areas and coaching techniques covered during coaching sessions. We examine whether the MTC program improved

teachers' instructional practices as captured by a range of measures including classroom observations, principal evaluations, and student surveys.

Setting, Population:

MTC coaches worked in partnership with New Schools for New Orleans to recruit teachers of all subjects and grade levels employed at charter schools across the Recovery School District. Among the final sample, teachers taught at 21 different schools operated by 13 charter management organizations. Participating teachers were early to mid-career teachers who taught kindergarten through 12th grade across all core academic subjects. Seventy-five percent were female, 76% were white, and 17% were African-American. Over three-fourths of the teachers entered the profession through alternative licensure programs such as Teach for American or TeachNOLA. Twenty-two percent of teachers held a master's degree.

Intervention / Program / Practice:

The MTC coaching program is centered on a skills-based approach to teaching. Coaches work with teachers to identify and set instructional practice improvement goals in several key areas including behavior management, lesson planning and execution, productive use of class time, student engagement, and classroom climate. Participating teachers attended a four-day training workshop during the summer and then worked individually with one of three experienced coaches for at least three weeks during the school year. Coaches set rigorous expectations for teacher growth and evaluated teachers' progress through formative assessments on the classroom observation rubric developed by the coaching program.

Research Design, Data Collection and Analysis:

Among the 59 teachers who expressed high levels of interest in coaching and obtained the support of their principals to participate, we randomly assigned half to receive an offer of coaching using a block-randomized design. Specifically, we randomly assigned teachers to either the treatment or control condition within blocks determined by which school they taught at in the 2010/11 school year. We observe the effect of treatment on three main outcomes: (1) the MATCH observational instrument developed by the coaching program, (2) a principal survey including ten items, and (3) student responses on the TRIPOD survey.

The MATCH rubric is comprised of two overall codes, *Overall Achievement of Lesson Aim* and *Overall Behavioral Climate*. Each code is scored holistically on a scale of 1-10 based on key indicators throughout the lesson. Indicators for *Behavioral Climate* include time on task, transitions, and student response to teacher corrections. Indicators for *Achievement of Lesson Aim* include clarity and rigor of the aim, alignment of student practice, and assessment and feedback. Raters also look for general practices in classroom management, student engagement, and instructional quality. Observations were conducted by two trained and experienced educators who were blind to treatment status and who achieved high levels of one-off agreement with master raters (88% for *Overall Achievement of Lesson Aim* and 100% for *Overall Behavioral Climate*) on the 10-point rubric scale.

We utilize a principal survey adapted from surveys developed by Jacob and Lefgren (2008) and Harris and Sass (2009). We ask principals to rate teachers' performance on ten elements of teaching practice from 1-9 where 1 is inadequate, 5 is average, and 9 is exceptional. Performance areas include: *Overall Effectiveness, Dedication and Work Ethic, Organization, Classroom Management, Time Management in Class, Time on Task in Class, Relationships with*

Students, Communication with Parents, Collaboration with Colleagues, and Relationships with Administrators (Cronbach's alpha between 0.92 and 0.95, depending on inclusion of missing values).

Finally, we measured teachers' performance on the TRIPOD survey, which is comprised of items designed to capture students' opinions about their teacher's instructional practices. Measures of teacher effectiveness are categorized into seven overall domains or "C's": *Care, Clarify, Control, Challenge, Captivate, Confer, and Consolidate*. In addition, the survey asks students about their level of academic and social engagement in class. We focus our confirmatory analysis on two specific measures, *Control* and *Challenge*, which ask students about the behavioral climate and the level of academic rigor in their class.

We examine the effect of MTC on these outcomes of interest using ordinary least squares regression. Using Y to represent a given teacher-level outcome of interest, we estimate the following regression:

$$Y_j = Y_{j,t-1} + \beta MTC_j + \alpha_{s,t-1} + \varepsilon_j \quad (1)$$

for teacher j in school s at time t . We were able to collect baseline measures for the observational instrument and for principal surveys, which we include in the model, $Y_{j,t-1}$, to increase the precision of our estimates. The vector, $\alpha_{s,t-1}$, represents fixed effects for the school blocks in which teachers were randomly assigned. The coefficient β on the indicator for the offer of treatment is our parameter of interest. We interpret these estimates as Average Treatment Effects (ATE) given that every teacher who was offered coaching, except two who withdrew prior to the 2011/12 school year, fully participated in the program. We estimate standard errors clustered by teachers' 2011/12 school in order to account for the likely non-independence of teacher's error terms within schools.

Findings / Results:

We find that that MTC was implemented with a high degree of fidelity and had a profound effect on teachers' instructional practices compared to control teachers who had access to standard professional development opportunities. We analyze teachers' emails and coaches' logs to assess the content and methods used during coaching sessions. Over the course of the year, teachers and coaches focused predominantly on *behavior management* and *instruction* (Figure 1). Coaches provided direct feedback to teachers in almost 80% of the sessions and engaged in lesson planning and review of videotaped lessons in approximately half the sessions (Figure 2).

We examine the validity of our random assignment by comparing average demographic characteristics across the treatment and control groups in Table 1. Differences in mean values across the treatment and control groups are small and insignificant for each measure, suggesting that teachers assigned to treatment and control conditions were equal in expectation on both observed and unobserved characteristics.

Our findings suggest that teachers who received coaching are more effective than those who participated in the standard PD activities provided by their schools. On the observational instrument, outside observers rated treatment teachers 0.64 ($p=.03$) and 0.55 ($p=.06$) standard deviations (sd) higher on the classroom behavioral climate and student achievement of lesson aim (Table 2). In addition, principals rated treatment teachers somewhat higher on a composite measure of teachers' overall effectiveness (0.11 sd, $p=.56$), although this small effect size cannot

be distinguished from zero. Finally, on the TRIPOD survey, students rated treatment teachers 0.28 sd higher ($p=.04$) than control teachers on their ability to challenge them with rigorous work. The estimated effect size of coaching on teacher's ability to control the classroom according to student's opinions was positive (0.14 sd, $p=.45$) but not statistically different from zero.

Exploratory analyses of the effect of MTC on specific instructional practices as rated by principals and as self-assessed by teachers further suggests that the program had a concentrated effect on several specific aspects of instructional and professional practice. Principals saw the biggest improvement in teachers' ability to use time effectively in class (0.45 sd, $p=.08$) followed by their classroom management skills (0.19 sd, $p=.52$). Estimated treatment effects using teacher self-reported survey data suggest that teachers in the treatment group experienced the largest gains in their organization (0.49 sd, $p=.11$) and their dedication and work ethic (0.46 sd, $p=.23$).

In order to examine the possible threat of sample attrition to our estimates, we conduct a variety of tests and robustness checks.* We find no evidence of differential attrition across any of the observed teacher characteristics in our dataset, suggesting that those teachers who left the study were not systematically different across treatment and control groups (Table 3). We also re-estimate treatment effects using our full sample by imputing missing data for teachers who left the study using multiple imputation (Rubin, 1987). Results are robust to the inclusion of our full sample of teachers across all outcomes (Table 2, column 2). Overall, we interpret results of these tests as strong evidence that our findings cannot be explained away by differential sample attrition.

Conclusions:

Improving the instructional practices of the 3.5 million teachers in classrooms across the United States will be central to any large-scale effort to increase teacher effectiveness in public schools. However, studies repeatedly find that PD, as it is practiced in most public schools, does little to change teachers' classroom practices or improve student achievement. Findings from this study, combined with those of Allen et al. (2011), suggest that teacher coaching can enhance teachers' classroom practices dramatically. We are currently analyzing student achievement data that were recently provided by the Louisiana Department of Education in order to evaluate the effect of MTC coaching on student achievement. Furthermore, we have expanded this experiment in order to increase the precision on our estimates by including a second cohort of 90 teachers, half of whom were randomly assigned to receive coaching.

Given that PD programming in public schools supports a multibillion-dollar industry, the choices policymakers and administrators make when allocating these funds are critical. Individualization makes coaching widely applicable to early and mid-career teachers across grades and subjects and suggests that coaching is a viable alternative to school-wide PD programming. If the effects of teacher coaching are sustained over time and coaching proves to be equally effective across a wide variety of contexts, then all teachers should have the opportunity to be coached. Building the necessary evidentiary base and shifting longstanding norms about PD practices will take time, but the evidence-to-date suggests doing so will be a valuable investment.

* Seven out of the 59 teachers in our study were censored from our analysis because they left teaching, withdrew participation, or moved to a new school that did not grant us access.

Appendix A. References

- Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, 333, 1034-1037.
- Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad*. Palo Alto, CA: National Staff Development Council and The School Redesign Network, Stanford University.
- Freiberg, H. J., Huzinec, C. A., & Templeton, S. M. (2009). Classroom management – a pathway to student achievement: A study of fourteen inner-city elementary schools. *Elementary School Journal*, 110(1), 63-80.
- Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., Uekawa, K., Falk, A., Bloom, H. S., Doolittle, F., Zhu, P., & Sztejnberg, L. (2008). *The impact of two professional development interventions on early reading instruction and achievement*. Washington, D. C.: U.S. Department of Education, National Center for Education Statistics.
- Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Eaton, M., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., & Doolittle, F. (2011). *Middle school mathematics professional development impact study: Findings after the second year of implementation*. Washington, D. C.: U.S. Department of Education, National Center for Education Statistics.
- Harris, D. N., & Sass, T.R. (2009). What makes for a good teacher and who can tell? CALDER Working Paper No. 30
- Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality, and student achievement. *Journal of Public Economics*, 95, 798-812.
- Jacob B.A., & Lefgren L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 20(1), 101-136.
- Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2010). *Identifying effective classroom practices using student achievement data*. Cambridge, MA: National Bureau of Economic Research.
- Miles, K. H., Odden, A., Fermanich, M., Archibald, S., & Gallagher, A. (2011). Inside the black box of professional development spending: Lessons from comparing five urban districts. *Journal of Education Finance*, 30(1), 1-26.
- Rivkin, G., Hanushek E., & Kain, J. (2005). Teachers, schools, and academic achievement.

Econometrica, 73(2), 417-458.

Rockoff, J.E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2), 247-252.

Rubin, Donald. 1987. *Multiple imputation for nonresponsive in surveys*. New York: Wiley & Sons Inc.

Scher, L., & O'Reilly, F. (2009). Professional development for K-12 math and science teachers: What do we really know? *Journal of Research on Educational Effectiveness*, 2(3), 209-249.

Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

Appendix B. Tables and Figures

Figure 1. Percent of sessions (n=101) where focus area was addressed (red) and percent of teachers (n=28) who ever worked on given focus area (blue).

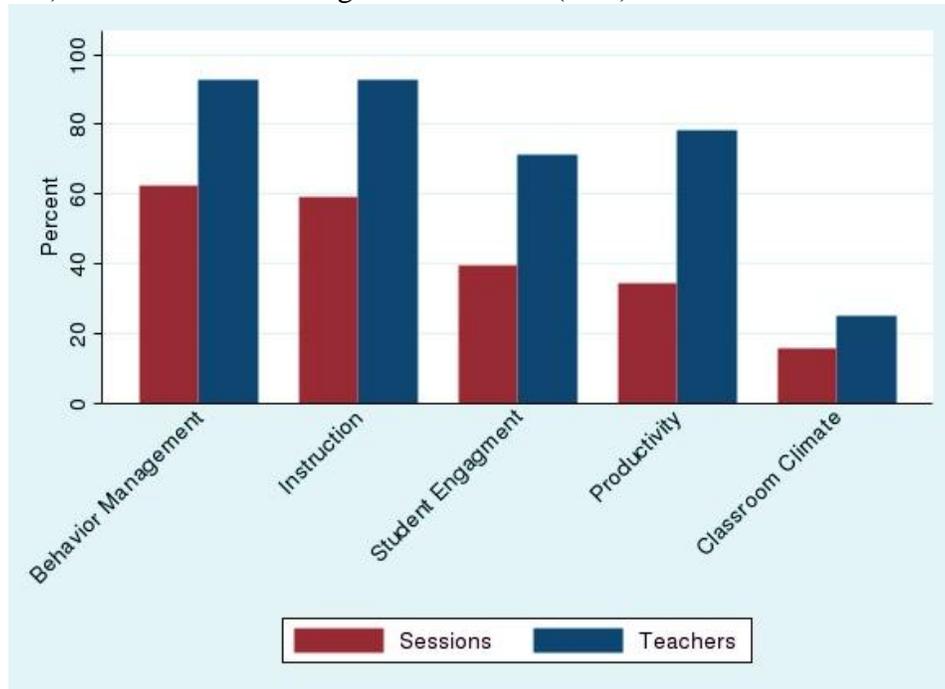
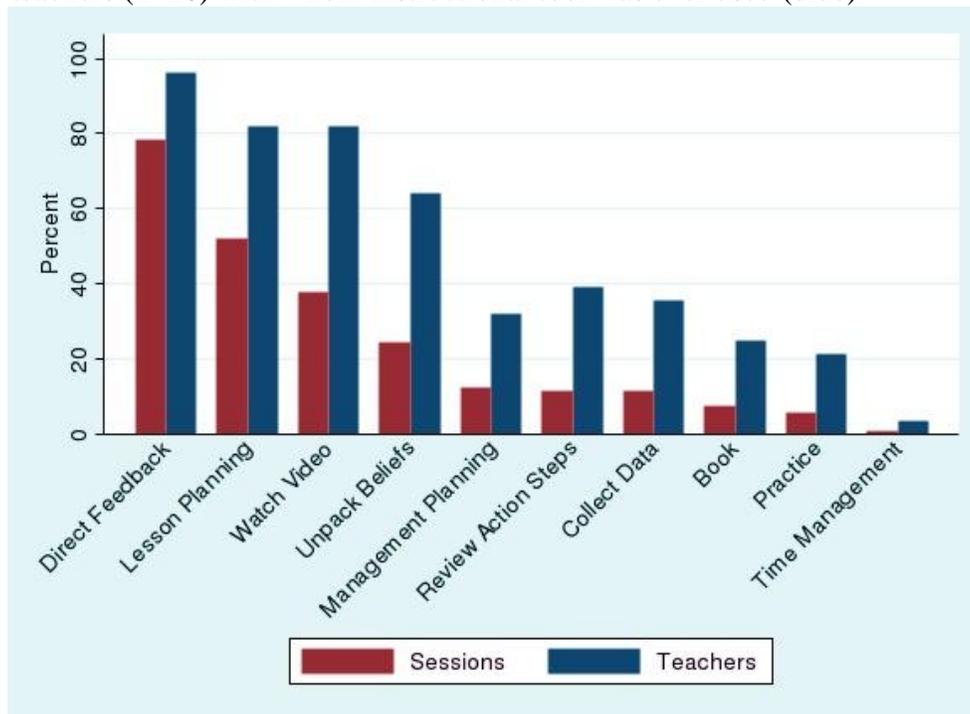


Figure 2. Percent of coaching sessions (n=106) where instructional tool was used (red) and percent of teachers (n=28) with whom instructional tool was ever used (blue).



Tables

Table 1. Baseline characteristics across treatment and control groups

Variable	Full Sample	Treat	Control	p-value
Female	0.746	0.700	0.793	0.420
African American	0.169	0.200	0.138	0.533
White	0.763	0.767	0.759	0.943
Age	26.1	26.2	26.0	0.798
Experience	2.96	4.00	3.93	0.905
Alternative Certification	0.763	0.833	0.690	0.201
Masters' Degree	0.220	0.200	0.241	0.707
Level of Interest in Coaching	9.11	9.23	8.98	0.330
F-statistic from Joint Test				0.764
p-value				0.636

Notes: n=59

Table 2: Parameter estimates of the effect of Match Teacher Coaching on measures of teacher effectiveness

	Primary Findings	Multiple Imputation
<u>MATCH Classroom Observation Rubric</u>		
Lesson Aim	0.550+ (0.265)	0.635* (0.235)
Behavioral Climate	0.640* (0.278)	0.726** (0.226)
Observations	52	59
<u>Principal Evaluation</u>		
Overall Composite	0.105 (0.177)	0.125 (0.283)
Observations	52	59
<u>TRIPOD Student Survey</u>		
Control	0.140 (0.182)	0.145 (0.156)
Challenge	0.281* (0.130)	0.245* (0.114)
Observations	50	59

Notes: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Each cell contains results from a separate regression. All estimates, except the % who agree that they "learn a lot" in their class, are reported as effect sizes with corresponding standard errors clustered by school in parentheses. All regressions include fixed effects for schools (in the prior year). Parameters estimated with multiple imputation use all demographic variables in Table 2 to impute missing values across ten replication data sets.

Table 3. Parameter estimates of the difference in demographic characteristics of attritors across treatment and control groups

Variable	Coefficient	p-value
Female	0.019	0.961
African American	-0.048	0.888
White	-0.336	0.388
Age	-3.506	0.328
Experience	-1.501	0.443
Alternative Certification	0.045	0.906
Masters' Degree	-0.406	0.282
Level of Interest in Coaching	-0.807	0.358

Notes: n=59. Seven teachers left the study, two teachers in the treatment group and five teachers in the control group.