The Educational Forum, 82: 1–18, 2018

ISSN: 0013-1725 print/1938-8098 online
DOI: 10.1080/00131725.2018.1474985

THE EDUCATIONAL FORUM

Evaluation of video-based instructional coaching for Middle school teachers: Evidence from a multiple baseline study

David S. Knight*[a], *Mike Hock*[b], *Thomas M. Skrtic*[b], *Barbara A. Bradley*[b] *and Jim Knight
[a]*University of Texas at El Paso, El Paso, Texas, USA;*
[b]*University of Kansas, Lawrence, Kansas, USA*

Q1
Q2

Abstract

We designed a multiple baseline study to evaluate an instructional coaching model in which coaches use video recordings of collaborating teachers' classrooms to inform their coaching practices. In this model, teachers and coaches use video evidence to co-construct a student-based goal and identify best practices to reach their goal. We found that the instructional coaching model is associated with greater use of effective pedagogical strategies among teachers and increased student engagement in the classroom.

Key Words: *classroom observation techniques, instructional coaching, instructional improvement, middle school teachers, professional development, student engagement*

Teacher professional development is increasingly seen as ineffective for improving teacher practice or student learning (e.g., The New Teacher Project, 2015). Traditional professional development is often disconnected from teachers' classrooms and involves little active participation. Instructional coaching represents a nontraditional form of professional development that encourages active participation among teachers and is embedded within the daily realities of teachers (Darling-Hammond, 2010; Darling-Hammond, Wei, Andree, Richardson, & Orphanos, 2009). Evaluations of coaching and mentoring of teachers

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/utef. *Address correspondence to David S. Knight, University of Texas at El Paso, Education Building, 105C, 500 W. University Avenue, El Paso, TX 79968, USA. E-mail dsknight@utep.edu*

demonstrate that intensive professional development on effective pedagogical practices shows promise for raising student achievement (Biancarosa, Bryk, & Dexter, 2010; Rockoff, 2008). Coaching and mentoring of middle school teachers is particularly important because teachers in middle schools play a critical role in preparing students for success in high school and beyond (George, 2004; Wentzel, 1997).

In this study, we evaluated a recently established video-based instructional coaching model (Bradley et al., 2013; Knight, Hock, & Knight, 2016). The model was developed over a 3-year period in which researchers used Design Based research (see Reinking & Bradley, 2008) to refine and improve on an instructional coaching model. The result was a video-based instructional coaching cycle in which teachers record their own classroom, review tapes with instructional coaches, set specific measurable goals, track progress toward those goals, and evaluate their own teaching over the course of a semester. This article reports the findings of a pilot study of this recently developed coaching model. Using classroom observations, we found that teachers who collaborated with an instructional coach using this coaching method increased their use of evidence-based instructional practices, compared to teachers who had not yet begun collaborating with a coach. Students in those classrooms showed increased engagement as assessed through measures of student "time-on-task" Q3 (Reinke, Lewis-Palmer, & Merrell, 2008).

Our article is organized as follows. We begin by synthesizing literature on the evaluation of teacher professional development and instructional coaching. Next, we describe the components of the instructional coaching model evaluated for this study. We then present the data, design, and methods. Finally, we report findings and discuss potential threats to internal and external validity and conclude with implications for practice and future research.

Evaluating Professional Development and Instructional Coaching

Prior to the 1990s, evaluation of professional development generally consisted of teacher reports of satisfaction and focused on small scale programs (Darling-Hammond & Baratz-Snowden, 2005; Frechtling, Sharp, Carey, & Vaden-Kiernan, 1995). More recent evaluations identify specific elements of professional development associated with instructional change (Cohen & Hill, 2001; Desimone, Porter, Garet, Yoon, & Birman, 2002; Garet, Porter, Desimone, Birman, & Yoon, 2001) and link professional development experiences to measures of student outcomes (Angrist & Lavy, 2001; Biancarosa et al., 2010; Harris & Sass, 2011; Jacob & Lefgren, 2004; Kraft & Blazar, 2017; Marsh, McCombs, & Martorell, 2010; Strunk, McEachin, & Q4 Westover, 2014). We discuss findings from each of these approaches to evaluation of profes- Q5 sional development.

Effects of Professional Development on Instructional Change

Two large-scale surveys identified the programmatic elements of professional development most likely to lead to instructional change. Garet et al.'s (2001) study, conducted with a nationally representative survey of 358 school districts, examined two forms of professional development, traditional and "reform type." Traditional professional development typically involves short-term workshops removed from the classroom setting with little follow-up training (Desimone, 2009). Earlier research on traditional professional development suggests this approach is unlikely to change instructional practice (e.g., Bush, 1984; Cohen & Hill,

2001; Darling-Hammond et al., 2009). Garet and colleagues (2001) define "reform type" professional development as professional learning communities, instructional coaching, or individual teacher research projects. The authors identified core and structural features associated with instructional change. Effective core features included a focus on content knowledge, opportunities for active learning, and coherence with teachers' other learning activities within the school. The authors used path modeling to show how structural features moderated the effect of core features on instructional change. Structural features that facilitated the efficacy of core features included longer time spans, for instance activities that took place over several months, larger amounts of contact hours (i.e., above 35 hours), and collective participation among teachers. Many of these structural features were associated with "reform type" professional development whereas traditional professional development activities were less likely to include these features.

Building on this research, Desimone and her colleagues (2002) surveyed teachers in 30 schools over a 3-year period. Longitudinal survey methods provided a more plausibly causal estimate of the effects of professional development on instructional change. The first and third surveys (year one and three, respectively) asked teachers to report on their instructional practices and teachers reported their experiences with professional development in the second survey (year two). The authors' findings were consistent with their earlier work (Garet et al., 2001): effective professional development was time-intensive, ongoing, focused on teachers' content area, included teachers as active learners, and facilitated collaboration among teachers at their schools.

Effects of Professional Development on Student Outcomes

Research on professional development that demonstrates instructional change is important; however, an additional important component of evaluation involves linking instructional change to positive student outcomes (Guskey, 2002). Obtaining causal estimates of the effect of teacher professional development on student outcomes is challenging because teachers typically self-select into professional development (Darling-Hammond et al., 2009) and data systems linking teachers to professional development experiences and their students' outcomes are rare (Harris & Sass, 2011). In 1996, Chicago Public Schools (CPS) reimbursed schools for teacher professional development up to $90,000 for one year for any school in which fewer than 15% of students scored at or above national norms on state standardized reading tests. The design of this policy provided a unique opportunity to estimate the causal impact of professional development because schools that scored near the 15% proficiency rate were essentially randomly assigned to professional development. A study using regression discontinuity design compared student outcomes for schools near either side of this proficiency cut point (Jacob & Lefgren, 2004). The program had no measurable effects on standardized test scores, which likely resulted from the low-intensity, short-term nature of the professional development intervention. Other studies of the CPS professional development intervention suggested that this program represented only a modest investment relative to the existing opportunities for professional development in the district (Finnigan, O'Day, & Wakelyn, 2001).

Particular university courses for in-service teachers may be more effective than others. Drawing on statewide data from Florida, Harris and Sass (2011) found that content-focused

courses directly related to the teachers' instructional area were associated with greater learning gains for students. However, despite the rich dataset, the authors could only measure university course-taking and the type of courses, and were unable to observe any other forms of professional development such as the reform type described above. During the time span of the study, the state had implemented a literacy policy ("Just Read, Florida!") that involved placing more than 2,000 literacy coaches in Florida middle schools, and one study found a positive association between coaching and student achievement in these schools (Marsh et al., 2010).

When researchers actually go into schools or talk to school and district personnel to understand the forms of professional development taking place, findings are more likely to provide useful insights into best practices. For instance, Strunk and colleagues (2012) examined the impacts of intensive professional development provided to persistently low-performing districts in California through the state's District Assistance and Intervention Teams. As part of No Child Left Behind, states were required to provide technical assistance in the form of capacity-building and professional development to schools in low-performing districts (Gottfried, Stecher, Hoover, & Cross, 2011). Persistently low-performing districts in California, like many other states, receive intensive intervention in the form of District Assistance and Intervention Teams (Strunk et al., 2014). Intensive professional development and capacity-building was associated with greater math achievement compared to a less intensive intervention. Qualitative findings showed that the most effective interventions focused on setting high expectations, using data to inform instruction, and providing instructional coaches to work with teachers on implementing these reforms.

Instructional coaching models are not all the same and scholars have highlighted the importance of implementing research-based coaching models with well-trained coaches (Costa & Garmston, 2015; Desimone, Smith, & Phillips, 2013; Duessen, Coskie, Robinson, & Autio, 2007; Kraft, Blazar, & Hogan, 2018). Two rigorously executed randomized trials of intensive, evidence-based professional development including instructional coaching both found changes in teacher knowledge and instructional practice, but no significant effects on student achievement (Garet et al., 2008, 2011). Coaches in these studies were only trained for one week prior to the intervention and there was no formal coaching model to frame their work. Other studies evaluate coaching models that include substantial training for coaches prior to beginning their work. For example, the Literacy Collaborative is a professional development program in which instructional coaches receive significant training over several years and provide professional development within a well-established literacy framework (Atteberry & Bryk, 2011). The program showed positive effects on kindergarten through grade 2 student achievement over a 4-year time period (Biancarosa et al., 2010).

Current Status of Professional Development

Although research highlights the importance of intensive investment in professional development, school districts rarely implement such policies (Desimone et al., 2002; Garet et al., 2001). In their review of teacher professional development in the United States and abroad, Darling-Hammond et al. (2009) found that most learning opportunities for teachers resembled traditional workshops with a duration less than two days. Only 23% of teachers surveyed in the Garet et al. (2001) study reported taking part in reform-type professional

development. Part of the reluctance to implement coaching may be due to inadequate funding. In our prior work, we found that instructional coaching was six to 12 times more expensive for school districts compared to traditional professional development (Knight, 2012). Resource tradeoffs among salaries, class sizes, and professional development may limit investments in intensive professional development such as instructional coaching (Cortina, 2011). Principals and school district leaders need to consider these tradeoffs when deciding whether to implement instructional coaching (Picus, 2017).

The Instructional Coaching Model

Coaches in the current study implemented the Partnership Instructional Coaching model. The model includes five steps. First, once a teacher decides she or he would like to collaborate with a coach, the coach videotapes one or more of the teacher's classroom lessons. Second, the coach uploads the video to the teacher's computer and each person watches the video separately. During the next one-on-one session, under the direction of their coach, teachers develop a goal related to student outcomes and select an instructional practice that would allow them to achieve that goal. For example, one teacher's goal in the current study was that 80% of her students would be proficient in the unit being taught, as measured by the end of unit test. In the fourth step, the instructional coach may choose to model the instructional practice while the teacher observes, before the teacher attempts to implement the new instructional practice. Finally, the coach video records the teacher's use of the new practice and the teacher and coach meet again one-on-one to reflect on the lesson, and possibly construct a new goal. This model is further described in Knight et al. (2016).

Design and Methodology

We employed a single-case experimental design known as Multiple Baseline Design (Hersen & Barlow, 1976). The defining characteristic of Multiple Baseline Design is that Q7
intervention's implementation is staggered over time, and observation takes place before and after the intervention (Gay & Airasian, 2000). Because the researcher controls when the intervention takes place for each participant, Multiple Baseline Design provides strong evidence that any changes in the dependent variables are the result of intervention, and not due to history, maturation, or other confounding factors (Campbell & Stanley, 1963; Hersen & Barlow, 1976). In the section below, we describe the study participants and how data were Q8
collected and analyzed.

Study Participants and Data Sources

The study included four instructional coaches who collaborated with two teachers each, for a total of eight teachers, over the course of one semester. Coaches had between 5–12 years of teaching experience, while teachers' experience ranged from 1–8 years. Teacher–coach dyads were separated into three groups, and collaboration began at three different points in time during the semester. Teachers and coaches were spread across three middle schools in a U.S. Pacific Northwest, suburban school district with two instructional coaches working in one school and one instructional coach working in each of the other two schools. The school district included approximately 40,000 students; 2,500 full-time teachers; and 50 elementary, middle, and high schools when the study took place during the 2012–13 school year. At the time of the study, the school district was experiencing some changing demographics. These included increases in students eligible for free and reduced lunch (17–33%),

students with limited English proficiency (9–15%) and students of color (24–43%) over the previous 10 years. These changes were reflected in the three schools involved with the study.

At least five classroom observations of an entire class period were administered both before and after each teacher collaborated with an instructional coach. Observations took place over a 4-month span. To track changes in instruction, research assistants used the Classroom Observation Form (as shown in the Appendix). The form includes three focus areas: (a) *Planning Content*, (b) *Formative Assessment*, and (c) *Community Building*. The dependent variable for each teacher was based on her or his planned goal; all goals fell into one of the three focus areas. For example, one teacher–coach dyad selected the goal of increasing the number of students engaged in the lesson to 90% of students. The coach recommended to the teacher that she increase the amount of wait time after asking open-ended questions to the class. We therefore categorized this teachers' focus area as *Community Building* and tracked changes in this focus area of the Classroom Observation Form. Each focus area was scored by percentage of behaviors observed, thus the scale was 0–100.

To increase interrater reliability, five research assistants were trained to use the Classroom Observation Form in an actual classroom setting, prior to the beginning of the study. In order to measure interrater reliability, a portion of the observations were completed with two research assistants in the classroom making independent observations. Table 1 shows the number of pre- and post-coaching observations for each group, the number of interrater reliability checks, the percent agreement across raters within a particular classroom observation, and Cohen's Kappa for that group of observations. At least 20% of all observations involved two research assistants, including 20% of each groups' pre-intervention observations and 20% of each groups' post-intervention observations. All five research assistants were involved in interrater reliability checks and an overall score of 88.1% agreement was reached as well as Cohen's Kappa = .688 (Cohen, 1960). Kratochwill and colleagues (2010) suggested average percent agreement on individual observations of between 80–90% and Hartmann, Barrios, and Wood (2004) suggested Kappa ≥0.60 as an appropriate level of interrater reliability, thus the Classroom Observation Form was found to be a reliable tool to measure instructional practice. Instructional coaches were considered expert coaches by district administrators. Our research team provided a total of nine days of training during the two years leading up to the study. All coaching sessions were video recorded and closely

Table 1. Interrater Reliability Measures for the Classroom Observation Form

Group	*Pre-coaching observations*	*Reliability assessments (% of total)*	*Post-coaching observations*	*Reliability assessments (% of total)*	*Percent agreement*	*Cohen's Kappa*
1	15	3 (20%)	24	5 (21%)	81.9%	0.606
2	33	7 (21%)	20	4 (20%)	91.5%	0.669
3	24	6 (25%)	10	4 (40%)	90.0%	0.748
Total	72	16 (22%)	54	13 (24%)	88.1%	0.668

monitored by the researchers to ensure coaches were implementing the Partnership Coaching model with fidelity.

Classroom Observation Form

Each focus area of the Classroom Observation Form measures teachers' use of research-based instructional practices. Under *Planning Content*, the observation form measures teachers' proper use of learning maps (Crank & Bulgren, 1993), advance organizers (Lenz, Alley, & Schumaker, 1987), and guiding questions (Wilhelm, 2007). The *Formative Assessment* focus area tracks the quality of teachers' use of informal assessment tools (Black & William, 1998), including whether teachers provide feedback based on the assessment and when necessary adjusts instruction based on assessment. Finally, the third focus area, *Community Building*, tracks whether teachers make behavioral expectations explicit during various phases of the lesson (Sprick, Garrison, & Howard, 2009), and teachers' ratio of positive to negative interactions with students (Sprick, Knight, Reinke, & McKale, 2006). Teachers were scored in each of these three categories, and the dependent variable for each teacher was their focus area score. In addition, three measures of student time-on-task were taken at the beginning, middle, and end of each observation using the method suggested by Reinke et al. (2008). Students' time-on-task was the second dependent variable measured. Table 2 shows the focus area and number of observations for each teacher–coach dyad.

Analytic Approach

We analyze our data and report findings using two approaches. First, in the tradition of Multiple Baseline Design, we plot, over time, the average score in the teachers' focus area for each group of teachers. This visual analysis makes clear to the reader that the change in the dependent variable is the result of treatment (Kratochwill et al., 2010). We also test a series of regressions that estimate the impact of coaching on both teachers' average focus area scores as well as student classroom engagement.

Table 2. Number of classroom observations per teacher before and after collaborating with an instructional coach

Teacher	*Coach*	*School*	*Group*	*Pre-coaching observations*	*Post-coaching observations*	*Focus area*
1	1	A	2	10	7	3
2	1	A	1	5	8	3
3	2	B	1	5	9	1
4	3	B	1	5	7	2
5	2	B	2	13	6	3
6	3	B	3	10	5	3
7	4	C	2	11	7	2
8	4	C	3	13	5	3
Total: 8	4	3	3	72	55	3

Note. Group refers to the timing in which coaching began (see text and Figure 1 for more information).

We regress each dependent variable on a dummy variable labeled "*post*," in Equation (1). This treatment indicator takes the value of one if observation *i* took place after the teacher and coach began the instructional coaching process, and zero if the observation took place before coaching. In the second regression, we control for teacher fixed effects in the sample. Given the sampling design, controlling for teacher effects inherently accounted for school, coach, and timing effects.

$$Y_{ij} = b_0 + b_1 post_i + e_{ij} \tag{1}$$

$$Y_{ij} = b_0 + b_1 post_{ij} + \mathrm{b}_{2j} T_j + \mathrm{b}_{3j} T_j * post_i + \mu_{ij} \tag{2}$$

T_j represents a vector of seven dummy variables for teachers, with the eighth teacher as the reference and μ_{ij} represents unobserved factors at the teacher and classroom observation level (standard error at clustered at the teacher level). We use the same equations to estimate the effect of instructional coaching on student engagement. We also experimented with using a regression discontinuity approach in which the pre- and post-trends in outcomes are explicitly modeled. Our results did not change considerably across these models so we present the more straightforward fixed effects models.

Results

Teachers' Instructional Change

Figure 1 shows each teacher group's average score on the Classroom Observation Form over time (scores range from 0–100%). Blue lines indicate scores from pre-coaching observations, the vertical black line indicates when each group of teacher–coach dyads began meeting, and the red lines indicate scores from post-coaching observations. Scores on the first classroom observation after coaching began are displayed as grey dots. Because observations took place roughly once per week, and coaches met with teachers over several weeks, the first classroom observation typically took place after only one or two meetings between the teacher and coach. We therefore coded these observations as "neutral," while including them as post-coaching observations in the statistical analysis described below.[1]

Our second outcome of interest was students' classroom engagement, as measured by the percent of students on task at three points in time during a lesson. In Figure 2, each dot represents one classroom observation and the y-axis represents the percent of students engaged during the lesson. A separate line of best fit is also plotted to show the trends in student engagement before and after coaching. While student classroom engagement appeared to be increasing throughout the semester, on average engagement increased significantly after the teacher began collaborating with an instructional coach. We describe our tests for statistical significance below.

Table 3 shows the results of models estimated by Equations (1) and (2). After collaborating with an instructional coach, teachers in the sample made an average 44.3% gain in their focus area on the Classroom Observation Form (column 1 of Table 3). Our second model shows that the impact of coaching is robust to teacher, coach, school, and timing effects for participants in the sample. In addition, although two teachers began the intervention with

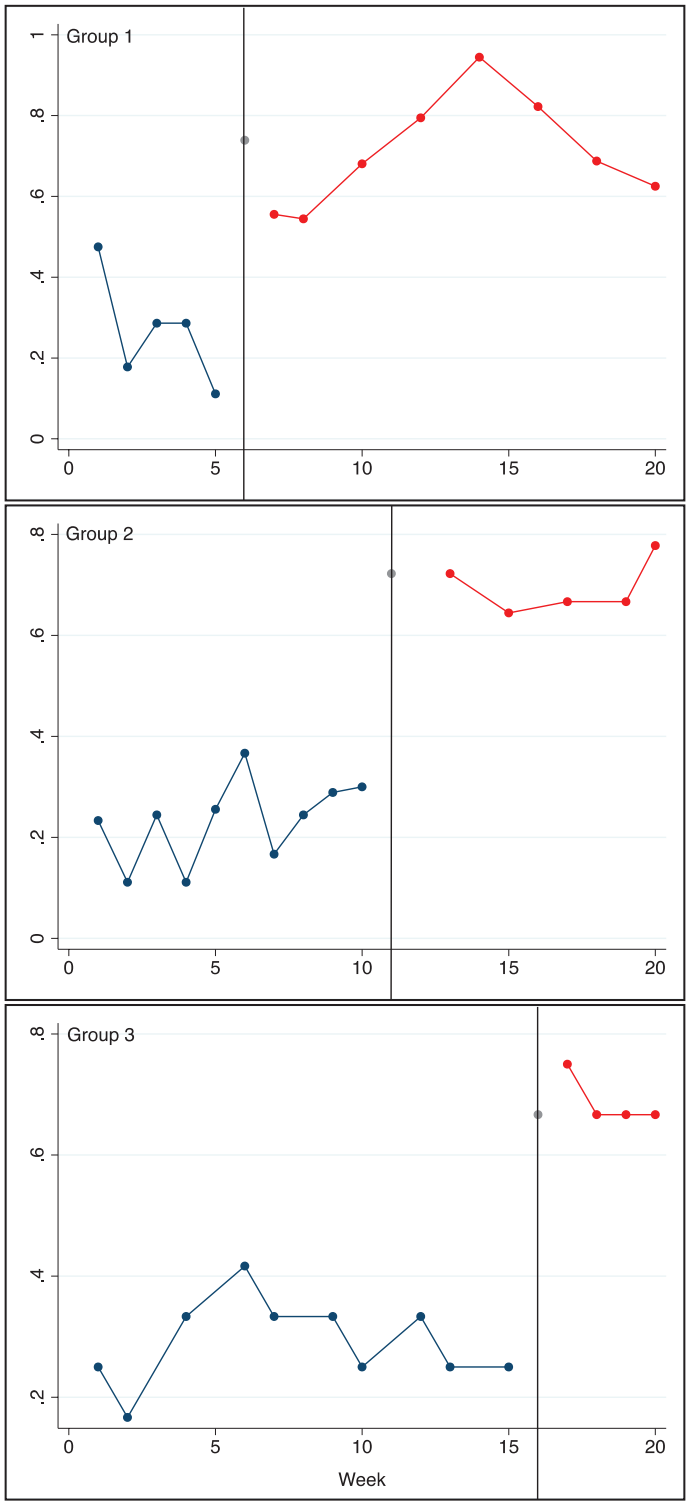

Figure 1. Focus area scores on the classroom observation form for Groups 1, 2, and 3.

Note: The black vertical line indicates when each group of teacher–coach dyads began coaching sessions. The grey dot (on the vertical line) represents the initial classroom observation after coaching began.

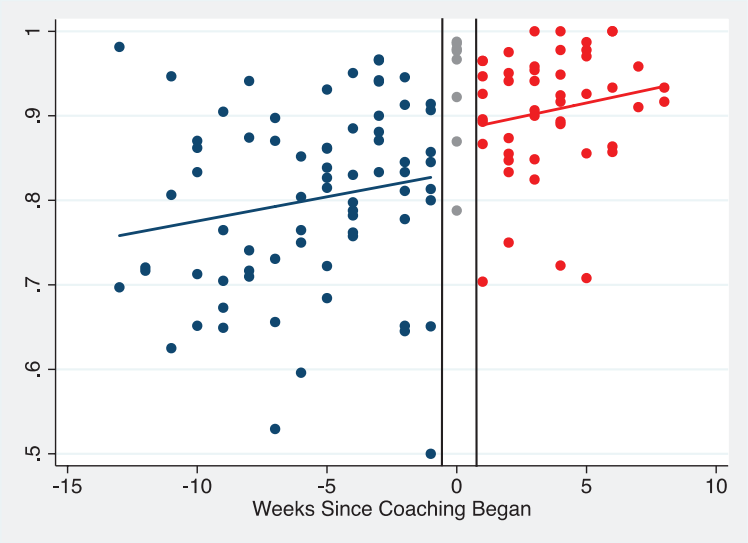

Figure 2. Average percent of students on task before and after instructional coaching.

Note: Blue dots indicate pre-coaching classroom observations, grey dots represent classroom observation while coaching was taking place, and red dots represent post-coaching classroom observations. A total of 127 classroom observations were conducted during the school year.

significantly lower pre-scores, the gain scores for all but one teacher did not differ significantly from one another, implying that the effect of coaching was not significantly different between teachers. These results provide strong evidence that teachers adjusted their instruction in measureable ways following their experience with the coaching model.

Students' Classroom Engagement

Table 6 provides results for measures of student engagement. The classrooms of teachers that collaborated with instructional coaches showed significant gains in the percent of students engaged during a class period. Column 1 of Table 6 shows that the average baseline score was 80.1% (the constant), while the average post-coaching score was 91.3%, a gain of 11.2 percentage points. The estimated gain score is very similar (an 11.1 percentage point gain) when controlling for the individual effect of teachers within the sample (and implicitly the coach, group, school, and timing effects, see Column 2 of Table 6). Given the standard deviation in pre-scores of 0.108, we found effect sizes of 1.03. This effect size is not as easily comparable to other measures of student achievement because of our small sample size and the unique measure of classroom engagement.

Discussion

This study makes two main contributions to research on professional development and instructional coaching. We estimated that over the course of one semester, teachers significantly changed their instructional practices after collaborating with an instructional coach. Additionally, we found that student classroom engagement, as measured by students' time-on-task during classroom observations, increased dramatically following the coaching intervention. Our second contribution to the field is our novel

Table 3. Change in focus area scores following instructional coaching

	(1)	*(2)*
Instructional coaching	0.443*** (0.037)	0.305*** (0.092)
Teacher effects		
Teacher 1		−0.028 (0.074)
Teacher 2		0.066 (0.084)
Teacher 3		−0.220* (0.092)
Teacher 4		0.066 (0.092)
Teacher 5		−0.141* (0.069)
Teacher 6		0.005 (0.074)
Teacher 7		0.014 (0.072)
Interactions		
Teacher 1 × Instructional coaching		−0.142 (0.126)
Teacher 2 × Instructional coaching		0.042 (0.132)
Teacher 3 × Instructional coaching		0.217 (0.134)
Teacher 4 × Instructional coaching		0.220 (0.138)
Teacher 5 × Instructional coaching		−0.043 (0.126)
Teacher 6 × Instructional coaching		0.161 (0.133)
Teacher 7 × Instructional coaching		0.328** (0.125)
Intercept	0.264*** (.024)	0.295*** (0.049)
Adjusted *R*-squared	0.537	0.658

Note. N = 127. The main effect of coaching in Model (2) is the gain score of Teacher 8. Robust standard errors are in parenthesis.

$^{*}p < .050$

$^{**}p < .010$

$^{***}p < .001.$

approach to studying newly developed professional development models. We discuss each of these in turn below.

Demonstrating Potential Impact

Our main purpose in this study was to estimate the impacts of the Partnership Instructional Coaching model under favorable conditions. We found an average effect size for instructional change 1.16 and average effect size of student engagement of 1.03 (both of which were statistically significant). Under typical circumstances, in which the coaching intervention is scaled up to a district-wide intervention, we would not necessarily expect to find as large effect sizes. Instructional coaches selected for participation in the study received considerable professional development on the coaching model, and teachers volunteered to participate in the study. Although the coaching model requires that teachers volunteer to participate, there may be unobserved characteristics in a larger sample of teachers that mitigate program effects. For instance, a principal may recommend that a teacher work with an instructional coach. In this case, the teacher may be less receptive to instructional change.

Table 4. Changes in student classroom engagement following instructional coaching

	(1)	*(2)*
Instructional coaching	0.112*** (0.017)	0.111* (0.044)
Teacher effects		
Teacher 1		−0.090* (0.035)
Teacher 2		−0.069 ~ (0.042)
Teacher 3		−0.027 (0.044)
Teacher 4		0.078 ~ (0.044)
Teacher 5		−0.091** (0.033)
Teacher 6		−0.025 (0.035)
Teacher 7		0.051 (0.034)
Interactions		
Teacher 1 × Instructional coaching		−0.004 (0.061)
Teacher 2 × Instructional coaching		0.011 (0.063)
Teacher 3 × Instructional coaching		−0.011 (0.065)
Teacher 4 × Instructional coaching		−0.060 (0.066)
Teacher 5 × Instructional coaching		0.053 (0.061)
Teacher 6 × Instructional coaching		−0.017 (0.064)
Teacher 7 × Instructional coaching		−0.036 (0.060)
Intercept	0.801*** (.011)	0.823*** (0.023)
Adjusted R-squared	0.252	0.409

Note. N = 127. The main effect of coaching in Model (2) is the gain score of the students of Teacher 8. Robust standard errors are in parenthesis. ~ indicates statistical significance at p < .100.

$^{*}p < .050$

$^{**}p < .010$

$^{***}p < .001.$

Another potential impediment to scaling up this coaching intervention may be cost (Cortina, 2011). One estimate of the cost of instructional coaching found that coaching can be six to 12 times as expensive as traditional approaches to professional development (Authors, 2012). We recommend that larger-scale studies of instructional coaching include analysis of costs so that practitioners and policy makers can gain a sense of the cost-effectiveness of coaching models.

Evaluation of Instructional Coaching

Central to the coaching model evaluated here is the teacher's ability to determine overall goals of the coaching sessions. We therefore introduced a novel approach to professional development evaluation in which the measured outcome is based on the intended goal of the coach–teacher relationship. While classroom observations measured teachers' use of instructional practices along three focus areas (content planning, formative assessment, and community building), each teachers' pre- and post-coaching observations scores were based on the particular focus area that best matched their chosen goal.

Limitations

We highlight three limitations that might each be addressed in future research. First, while the findings presented here clearly demonstrate systematic changes in instructional practice and student engagement, it is unclear what aspects of the coaching model were most effective. Future work might compare multiple coaching models to a business-as-usual condition. Second, our findings are not necessarily generalizable to other settings. Teachers and coaches in our sample volunteered to participate in the study. We therefore do not know whether the coaching model would have the same effects in a typical setting. In short, the experimental design provides evidence of internal validity, suggesting that instructional coaching increased both teachers' use of evidence-based practices and student classroom engagement *for the participants in our study*; however, the study offers less evidence of external validity. Finally, teachers' classrooms were observed over only a short period of time and teachers may ultimately revert back to their teaching practices prior to the coaching intervention. Further research on this coaching model will focus on particular components of the model, such as goal setting, modeling, or issues around the use of video, and follow coaches for a longer period of time.

Conclusions

This study examined the effects of the Partnership Instructional Coaching model on teachers' instructional practices and students' classroom engagement. As a new approach to instructional coaching, our purpose was to provide an initial evaluation of the program. We found that under favorable conditions, the model shows promise for changing instructional practices and increasing student engagement. Next steps in this work are to continue making improvements to the model based on our continuing use of Design Based research (Reinking & Bradley, 2008). At the same time, we note the importance of increasing the scale of the research design to examine the impacts of coaching in more typical school settings. Our findings suggest that such efforts may provide school leaders and policy makers with effective interventions for improving instruction and student outcomes.

Acknowledgments

The authors thank Argun Saatcioglu for helpful comments and suggestions and Jeff Levering, Barb Milliken, and Sarah Estes for excellent research assistance.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Note

1 Omitting these observations from the analysis did not change our quantitative findings significantly. Q10

References

Angrist, J. D., & Lavy, V. (2001). Does teacher training affect pupil learning? Evidence from matched comparisons in Jerusalem Public Schools. *Journal of Labor Economics*, *19*(2), 343–369. doi:10.1086/319564 Q11

Atteberry, A., & Bryk, A. S. (2011). Analyzing teacher participation in literacy coaching activities. *The Elementary School Journal*, *112*(2), 356–382. doi:10.1086/661994

Bradley, B. A., Harvey, S., Knight, J., Hock, M., Skrtic, T. M., Deshler, D., … Brasseur-Hock, I. (2013). Improving instructional coaching to support middle school teachers in the United States. In T. Plomp & N. Nieveen (Eds.), *Educational design research – Part B: Introduction and illustrative cases* (pp. 299–318). Enschede, The Netherlands: SLO.

Biancarosa, G., Bryk, A., & Dexter, E. (2010). Assessing the value-added effects of literacy collaborative professional development on student learning. *The Elementary School Journal*, *111*(1), 7–34. doi:10.1086/653468

Black, P., & William, D. (1998). Inside the black box raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 138–148. doi: 10.1177/003172171009200119

Bush, R. N. (1984). *Effective staff developments in making our schools more effective: Proceedings of three state conferences*. San Francisco, CA: Far West Laboratories.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental design for research*. Chicago, IL: Rand McNally.

Cohen, D. K., & Hill, H. (2001). *Learning policy: When state education reform works*. New Haven, CT: Yale University Press. doi:10.12987/yale/9780300089479.001.0001

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational Psychology Measurement*, *20*(1), 37–46. doi:10.1177/001316446002000104

Cortina, L. (2011). *School administrators and the professional learning of general education teachers related to gifted education: A Delphi study* (A Doctoral dissertation). Retrieved from Seton Hall University Dissertations and Theses (ETDs). 1421.

Costa, A. L., & Garmston, R. J. (2015). *Cognitive coaching: Developing self-directed leaders and learners*. Lanham, MD: Rowman & Littlefield.

Crank, J. N., & Bulgren, J. A. (1993). Visual depictions as information organizers for enhancing achievement of students with learning disabilities. *Learning Disabilities Research & Practice*, *8*(3), 140–147.

Darling-Hammond, L., & Baratz-Snowden, J. C. (2005). *A good teacher in every classroom: Preparing the highly qualified teachers our children deserve*. New York, NY: Jossey-Bass.

Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to equity will determine our future*. New York, NY: Teachers College Press. doi:10.18676/caderno-scenpec.v2i2.191

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*. School Redesign Network at Stanford University. National Staff Development Council.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualization and measures. *Educational Researcher*, *38*(3), 181–199. doi:10.3102/0013189x08331140

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis*, *24*(2), 81–112. doi:10.3102/01623737024002081

Desimone, L., Smith, T. M., & Phillips, K. (2013). Linking student achievement growth to professional development participation and changes in instruction: A longitudinal study of elementary students and teachers in Title I schools. *Teachers College Record*, *115*(5), 1–46.

Duessen, T., Coskie, T., Robinson, L., & Autio, E. (2007). *"Coach" can mean many things: Five categories of literacy coaches in Reading First. REL Northwest (REL 2007-No. 005). IES: US Department of Education.*

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., … Sztejnberg, L. (2008). *The impact of two professional development interventions on early reading instruction and achievement*. Washington, D.C.: U.S. Department of Education, NCEE.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*(4), 915–945. doi:10.3102/00028312038004915

Garet, M. S., Wayne, A., Stancavage, F., Taylor, J., Walters, K., Song, M., … Doolittle, F. (2011). *Middle school mathematics professional development impact study: Findings after the second year of implementation*. Washington, DC: National Center for Education Evaluation and Regional Assistance, U.S. Department of Education, NCEE.

Gay, L. R., & Airasian, P. (2000). *Educational research competencies for analysis and application* (6th Ed.). Upper Saddle River, NJ: Prentice Hall.

George, P. S. (2004). Middle school teacher education: Staying the course during trying times. *Middle School Journal*, *35*(4), 13–24. doi: 10.1080/00940771.2004.11461436

Gottfried, M. A., Stecher, B. M., Hoover, M., & Cross, A. B. (2011). *Federal and state roles and capacity for improving schools*. Santa Monica, CA: Rand Corporation.

Guskey, T. R. (2002). Does it make a difference? Evaluating professional development. *Educational Leadership*, *59*(6), 45–51.

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality, and student achievement. *Journal of Public Economics*, *95*(7-8), 798–812. doi: 10.1016/j.jpubeco.2010.11.009

Hartmann, D. P., Barrios, B. A., & Wood, D. D. (2004). Principles of behavioral observation. In S. N. Haynes & E. M. Hieby (Eds.), *Comprehensive handbook of psychological assessment* (Vol. 3, Behavioral assessment) (pp. 108–127). New York, NY: John Wiley & Sons.

Hersen, M., & Barlow, D. (1976). *Single-case experimental designs: Strategies for studying behavior change*. Elmsford, NY: Pergamon Press Inc.

Jacob, B. A., & Lefgren, L. (2004). The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago. *Journal of Human Resources*, *39*(1), 50–79. doi:10.2307/3559005

Kraft, M. A., & Blazar, D. (2017). Individualized coaching to improve teacher practice across grades and subjects: New experimental evidence. *Education Policy*, *31*(7), 1033–1068. doi:10.1177/0895904816631099

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the casual evidence. *Review of Educational Research*, 1–42. doi: 10.3102/0034654318759268 Q12

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation.* Retrieved from: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Knight, D. S. (2012). Assessing the cost of instructional coaching. *Journal of Education Finance*, *38*(1), 52–80. doi:10.1353/jef.2012.0010.

Knight, D. S., Hock, M., & Knight, J. (2016). Designing instructional coaching. In C. M. Reigeluth, B. J. Beatty, & R. D. Myers (Eds.). *Instructional-design theories and models: A learner-centered paradigm of education, volume IV* (pp. 269–286). New York, NY: Routledge.

Lenz, B. K., Alley, G. R., & Schumaker, J. B. (1987). Activating the inactive learner: Advance organizers in the secondary content classroom. *Learning Disability Quarterly*, *10*(1), 53–67. doi: 10.2307/1510755

Marsh, J. A., McCombs, J., & Martorell, F. (2010). How instructional coaches support data-driven decision making: Policy implementation and effects in Florida middle schools. *Educational Policy*, *24*(6), 872–907. doi:10.1177/0895904809341467

Picus, L. O. (2017). Strategic budgeting for school principals. In S. V. Coffin & B. S. Cooper (Eds.), *Sound school finance for educational excellence* (pp. 19–31) New York, NY: Rowan & Littlefield.

Reinke, W. M., Lewis-Palmer, T., & Merrell, K. (2008). The classroom checkup: A classwide teacher consultation model for increasing praise and decreasing disruptive behavior. *School Psychology Review*, *37*(3), 315–332.

Reinking, D., & Bradley, B. A. (2008). *On formative and design experiments: Approaches to language and literacy research*. New York, NY: Teachers College Press.

Rockoff, J. E. (2008). *Does mentoring reduce turnover and improve skills of new employees? Evidence from teachers in New York City. (No. w13868)*. National Bureau of Economic Research.

Sprick, R., Garrison, M., & Howard, L. (2009). *CHAMPS: A Proactive and positive approach to classroom management* (2nd ed.). Eugene, OR: Pacific Northwest Publishing.

Sprick, R., Knight, J., Reinke, W., & McKale, T. (2006). *Coaching classroom management: A toolkit for coaches and administrators*. Eugene, OR: Pacific Northwest Publishing.

Strunk, K. O., McEachin, A., & Westover, T. N. (2014). The use and efficacy of capacity-building assistance for low-performing districts: The case of California's district assistance and intervention teams. *Journal of Policy Analysis and Management*, *33*(3), 719–751. doi:10.1002/pam.21658

The New Teacher Project (2015). *The mirage: Confronting the hard truth about our quest for teacher development*. Washington, DC: Authors.

Wentzel, K. R. (1997). Student motivation in middle school: The role of perceived pedagogical caring. *Journal of Educational Psychology*, *89*(3), 411–419. doi:10.1037//0022-0663.89.3.411

Wilhelm, J. D. (2007). *Engaging readers & writers with inquiry: Promoting deep understandings in language arts and the content areas with guiding questions (theory and practice*). New York, NY: Scholastic Inc.

Appendix

Figure A1. *The classroom observation form.*

Classroom Observation Form

Teacher and Coach name: _____

Date: _____ Observer:_____ Reliability? _____ Lesson Topic:_____

		Comments
1. Planning Content		
Lessons are linked to state standards	Yes ___ No___	
Guiding Questions (GQ) are linked to standards	Yes ___ No___	
Learning Maps (LM) are created and used for the unit	Yes ___ No___	
Within first 10 min., an Advance Organizer is given	Yes ___ No___	
Rationale for lesson is given	Yes ___ No___	
Student expectations for learning are shared by teacher	Yes ___ No___	
Current lesson is placed in context of unit	Yes ___ No___	
5. Students and the teacher co-construct the LM (even though it's already created, they do it again together).	Yes ___ No___	
	Total ____/ 8	
2. Formative Assessment		
The Teacher:		
Selects an informal assessment tool (write type in comments)	Yes ___ No___	
Is assessment tool clearly linked to the lesson target?	Yes ___ No___	
Informally assesses ALL students	Yes ___ No___	
Teacher provides feedback to students	Yes ___ No___	
Adjusts instruction based on the assessment	Yes ___ No___	
	Total ____/ 5	
3. Community Building		
1. Behavioral expectations for all instructional activities	Yes ___ No___	
2. Behavioral expectations for student-student interactions	Yes ___ No___	
3. Behavioral expectations for all transitions	Yes ___ No___	
4. General classroom expectations are posted	Yes ___ No___	
5. Teacher exhibits respectful behavior toward students	Yes ___ No___	
	Total ___/ 5	
6. Number of times students are praised/corrected	Pr.__/ Cor.____	
7. Time on Task	___ Percentage	
7a. At 10 minutes in: ___# on task ____ # in class	___ Percentage	
7b. Half way through: ___# on task ____ # in class	___ Percentage	
7c. At 10 minutes left: ___# on task ____ # in class	____ Average	

Figure A2. *The time-on-task form.*

Time-on-Task Form

Directions: Take 3–5 seconds for each student, one by one, and assign a plus sign if the student is engaged in the behavior the teacher expects of the student and a zero to indicate behavior that can be considered off-task.

Number of students in the class _____ Teacher_____ Date_____

Ten minutes into the class period

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30

Midway through the class period

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30

With 10 minutes remaining in the class period

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30

Source: Reinke et al. (2008).