



Changes in coaching study design shed light on how features impact teacher practice

WHAT THE STUDY SAYS

Teacher coaching is a powerful form of professional learning that improves teaching practices and student achievement, yet little is known about the specific aspects of coaching programs that are more effective.

Researchers used a blocked randomized experiment to study the effects of one-to-one coaching on teacher practice. When pooled across all teachers in both cohorts, there is no effect of coaching on teacher practice, yet considerable variability exists between the cohorts.

Changes in program design that occurred between the two cohorts provided researchers an opportunity to study how differences in program features influence positive effects in the first cohort on teacher practice and the absence of effects in the second cohort.

Study description

Researchers applied a blocked randomized trial design to study the effects of MATCH Teacher Coaching

Joellen Killion (joellen.killion@learningforward.org) is senior advisor to Learning Forward. In each issue of JSD, Killion explores a recent research study to help practitioners understand the impact of particular professional learning practices on student outcomes.



across two cohorts of volunteer teachers in selected charter schools in the Recovery School District in New Orleans. Three specific areas of teacher practice, behavior management, instructional delivery, and student engagement were examined.

Large positive effects on teacher

practice occurred in cohort 1, yet did not occur in cohort 2. Further exploratory analyses of the features of the coaching program, specifically focus of coaching interactions, dosage of coaching, and the coach, offer possible explanations for the difference in effects between the cohorts.

At a glance

Overall, a study of one-to-one coaching across two cohorts did not significantly lead to improvements in teaching practice. Exploratory analyses of the features and effects of the two cohorts, however, suggest that changes in the design and focus of coaching may explain the large positive effects on teacher practice in one cohort that were absent in the other.

THE STUDY

Blazar, D. & Kraft, M. (2015, December). Exploring mechanisms of effective teaching coaching: A tale of two cohorts from a randomized experiment. *Educational Evaluation and Policy Analysis*, 37(4), 542-566.

Questions

Researchers sought to answer an overarching question about the effects of MATCH Coaching Program on teacher practice. Changes in the program design in cohort 2, primarily as a result of additional teachers and differences in impact between the two cohorts, created an opportunity to explore how the features of the coaching program influence the effects.

Methodology

In a limited blocked randomized trial study, researchers studied the effects of one year of coaching on two different cohorts of volunteer teachers. Cohort 1, with 30 treatment teachers, received coaching in 2011-12, and cohort 2, with 49 treatment teachers, received coaching in the subsequent school year.

Control-group teachers (cohort 1 = 29; cohort 2 = 45) received no coaching. Teachers were randomly assigned by block based on school and geography, and coaches were assigned primarily on teaching level (elementary, middle, and high school). Teachers within each cohort varied on a number of characteristics, including years of experience, demographics, type of preparation programs, and subject areas taught, yet across both cohorts the differences were insignificant.

Three coaches provided coaching each year, with only one coach, the director of the coaching program, remaining the same from cohort 1 to cohort 2. Treatment teachers received four days of training in the summer and then intensive coaching cycles of weeklong observations and feedback. Teachers received four weeks of coaching in cohort 1 and three weeks in cohort 2.

Coaches received training from the program director, who served as one of the coaches, in using the MATCH Classroom Observation Rubric to develop internal consistency and in giving feedback to teachers.

In cohort 1, coaches served about 10 teachers each, with some teachers receiving coaching from more than one coach during the program. In cohort 2, because of the increase in number of participating teachers, the amount of coaching was reduced from four to three weeks, and two coaches worked with about 20 teachers each, while the third coach (the program director) worked with only nine teachers.

Changes in the coaching program design for cohort 2 included a larger number of teacher participants; reduction in the dosage of coaching from four to three weeks; two new coaches; intentional sequencing of the focus within coaching interactions on behavior management until teachers

WHAT THIS MEANS FOR PRACTITIONERS

As a small study of the effects of coaching, the study provides multiple opportunities for examining how to examine impact of a program, as specified within the **Data standard** of Learning Forward's Standards for Professional Learning. New professional learning initiatives require rigorous evaluation to strengthen and refine them and to ensure that they produce the intended results.

It is unclear how the design of the coaching program studied meets the other Standards for Professional Learning, yet the study offers an example of how to assess a professional learning program. In addition, it provides insights into the features of effective coaching programs that contribute to positive effects on teacher practice. Because coaching is an increasingly common professional learning practice and one that is not inexpensive, decision makers and policymakers want to consider thoughtfully how to design, implement, and evaluate coaching programs to increase their effects on both educators and students.

demonstrated mastery before addressing instructional delivery and student engagement; more explicit guidance and direct feedback for cohort 2 coaches on debriefing observations; and greater emphasis by coaches on teachers practicing and watching video on behavior management.

Analysis

In the spring before randomization and training and coaching, coaches

observed all participating teachers and rated their performance in three areas of teacher practice, behavior management, instructional delivery, and student engagement using the MATCH Classroom Observation Rubric. The rubric provides a holistic score in two areas, achievement of lesson aim and behavioral climate. Coaches used the rubric during the coaching cycle to evaluate teacher practice in three areas.

One additional outcome measure was the Tripod student survey, administered to upper elementary and secondary students at the end of the coaching year. The survey focused on two areas: challenge and control — the areas most predictive of teacher value-added scores in reading and math — and the specific item, “In this class, we learn a lot every day.”

Other outcome measures included a principal survey based on teacher evaluations in 11 areas that were aggregated into an overall effectiveness composite and external-observer evaluation ratings from two classes at the end of the school year using the MATCH Classroom Observation Rubric. These measures, rather than student test data, provided a way to examine teacher practice across multiple subjects and grade levels and to focus on teaching practice specifically in a generalized way that guaranteed similar data for both control and treatment teachers.

All five scores from the various outcome measures, two Tripod student survey items, principal surveys, and external observer evaluation scores in the two domains of the MATCH Classroom Observation Rubric, were aggregated into a summary index. Qualitative data from interviews with coaches and some teachers complemented the quantitative analyses and informed findings and explanations.

Results

Pooled effects of the coaching program across cohorts 1 and 2 are

not significant on any of the measures, including the summary index. Yet the pooled finding provides limited insight into the effects of variation in program features between cohorts. Further explanatory analyses examine the differences and offer explanations for effects in cohort 1.

To examine the variations in the cohorts, researchers applied substantive statistical analyses to examine the effects of multiple features and offer possible explanations for why teachers in cohort 1 received a statistically significantly higher score on all measures with the exception of overall composite index and control at the end of the year of coaching than teachers in cohort 2, who showed no statistically significant differences at the end of their year of coaching.

Exploratory analyses of the effects of the variations in the coaching program features suggest that the treatment effect differences may be largely the result of the program features. In addition, researchers examine attenuation of spillover, school contexts, teacher characteristics, missing data, and participant dropout to eliminate other possible explanations for the effect differences.

Coaching program features affect results. Differences in the dosage of coaching; the sequence of coaching topics; the coaching techniques used, such as direct feedback, lesson planning, unpacking beliefs, practice, and video watching; and who the coach is offer promising explanations for the differences.

Teachers in cohort 1 received more coaching than those in cohort 2. In interactions with coaches, teachers in cohort 1 focused more on all three areas represented by the outcome measures rather than predominantly on behavior management, as they did in cohort 2.

Researchers suggest that “an additional week spent on instructional delivery [in cohort 1] is associated with positive and mostly statistically significant improvements in teachers’

practices” while “the time spent on behavior management [in cohort 2] is associated with negative and often statistically significant decrements in teachers’ practice” (p. 561).

Coaches in cohort 1 used less practice and video watching than coaches in cohort 2. There was a positive and statistically significant difference on the summary index between coaches in cohort 1 and cohort 2 (.87 standard deviation) and among coaches within each cohort.

Limitations

Researchers acknowledge some limitations in this study, including the lack of randomization of coach assignments and the potential effects of school context. Obviously the change in the program features presents another limitation, yet opened the door to unanticipated and informative exploration about how various features of the coaching program may influence effects.

Other limitations that may exist are the lack of intensive training and support for coaches, the structure of the coaching in a specific cycle focused around observation and feedback in intensive blocks, among others. Disappointing, yet understandable, is the decision to measure effects based on teacher practice without considering student achievement. A small concession to student learning is the component of the MATCH Classroom Observation Rubric focused on achieving the lesson aim and student response on the Tripod survey item on learning every day.

While the randomized trial experiment informs finding about coaching as a form of professional learning, the inclusion of subjects exclusively from charter schools who volunteered to participate limits the generalizability of the findings to those conditions and to this particular coaching approach. ■