

Making an Impact

January 2018

Impact of a checklist on principal–teacher feedback conferences following classroom observations

Kata Mihaly
RAND Corporation

Isaac M. Opper
RAND Corporation

Luis Rodriguez
Vanderbilt University

Heather L. Schwartz
RAND Corporation

Geoffrey Grimm
RAND Corporation

Louis T. Mariano
RAND Corporation

Key findings

This statewide experiment in New Mexico in 2015/16 tested whether providing principals and teachers a checklist to use in the feedback conferences that principals had with teachers following formal classroom observations would improve the quality and impact of the conferences.

- With two exceptions, the checklist had no clear impact on conference quality, teachers' instruction, or student achievement as of spring 2016.
 - According to teachers, the checklist reduced the degree to which principals dominated the feedback conferences.
 - According to teachers, the checklist made them more likely to follow their principals' professional development recommendations.
- Of principals who received the checklist, 58 percent reported using it.
- The low-cost electronic distribution of a guide and a short video were insufficient to substantially alter feedback conferences and other key outcomes, at least over the short run.

U.S. Department of Education

Betsy DeVos, *Secretary*

Institute of Education Sciences

Thomas W. Brock, *Commissioner for Education Research*
Delegated the Duties of Director

National Center for Education Evaluation and Regional Assistance

Ricky Takai, *Acting Commissioner*
Elizabeth Eisner, *Associate Commissioner*
Amy Johnson, *Action Editor*
Chris Boccanfuso, *Project Officer*

REL 2018–285

The National Center for Education Evaluation and Regional Assistance (NCEE) conducts unbiased large-scale evaluations of education programs and practices supported by federal funds; provides research-based technical assistance to educators and policymakers; and supports the synthesis and the widespread dissemination of the results of research and evaluation throughout the United States.

January 2018

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-12-C-0012 by Regional Educational Laboratory Southwest administered by SEDL. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This REL report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Mihaly, K., Schwartz, H. L., Opper, I. M., Grimm, G., Rodriguez, L., & Mariano, L. T. (2018). *Impact of a checklist on principal–teacher feedback conferences following classroom observations* (REL 2018–285). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from <http://ies.ed.gov/ncee/edlabs>.

This report is available on the Regional Educational Laboratory website at <http://ies.ed.gov/ncee/edlabs>.

Summary

Most states' teacher evaluation systems have changed substantially in the past decade. New evaluation systems typically require school leaders to observe teachers' classrooms two to three times a school year instead of once (Doherty & Jacobs, 2015). The feedback that school leaders provide to teachers after these observations is a key but understudied step in the teacher evaluation cycle. The feedback and subsequent professional development are intended to help teachers change their instructional practices and improve student achievement (Correnti & Rowan, 2007; DeNisi & Sonesh, 2011; Taylor & Tyler, 2012). However, little is known about the feedback that school leaders provide to teachers following classroom observations or about how to train leaders to make that feedback more effective.

This study examined the impact of disseminating a detailed checklist intended to structure an effective feedback conference between a school leader and a teacher following a classroom observation. The feedback conference checklist is a modified version of one created by the Carnegie Foundation for the Advancement of Teaching (Tang & Chow, 2007).

The checklist, along with short testimonial videos, was a low-cost, low-intensity intervention provided to a randomly selected half of 339 participating New Mexico principals in fall 2015 by the study team. These principals' schools constituted the treatment group. Principals in the treatment group schools received an email with an attachment containing a guide and a 24-item feedback conference checklist, plus a hyperlink to a three-minute testimonial video featuring a principal. Principals were encouraged to distribute the checklist to other school leaders and to use the checklist in all their feedback conferences in the 2015/16 school year. Principals were also asked to distribute the checklist to all their teachers in order to promote greater teacher participation in the feedback conference. The study team also emailed the same checklist plus a hyperlink to a three-minute testimonial video featuring a teacher to up to 10 randomly sampled teachers in each treatment group school.

The other half of the principals in the study schools formed the control group. Each of the control group principals received a two-page principal guide as an email attachment in fall 2015. The two-page guide reprised the five tips about feedback included in the summer 2015 New Mexico Public Education Department-sponsored professional development for principals and informed principals about the study. In addition, the study team sent up to 10 randomly sampled teachers in each control group school a two-page teacher guide summarizing the teacher evaluation system (Skandera, 2013) and teachers' right to receive post-observation feedback.

All principals and teachers in both the treatment group and the control group who consented to be in the study were asked to complete an online survey (one for principals, another for teachers) in spring 2015 and again in spring 2016.

The main outcomes of the study were principals' and teachers' reports of the impacts of the checklist and testimonial video on the perceived quality of feedback conferences following formal classroom observations; principals' recommendations for and teachers' take-up of professional development; and the quality of teachers' subsequent instructional practices as measured by principals' formal classroom observation scores and teachers' self-reported scores. Additional exploratory outcomes included the impact of the checklist on student

achievement (school-average math and English language arts scores on the Partnership for Assessment of Readiness for College and Careers assessment) and school report card grades (reported as an A, B, C, D, or F of multiple measures of a school's student achievement) compiled annually by the New Mexico Public Education Department. The study also documented how many recipients reported using the checklist and what they thought about it.

The checklist had few clear impacts on the quality of feedback, professional development outcomes, instructional practice, or student achievement. There were two exceptions: teachers who received the checklist reported that their principals were less likely to dominate the feedback conferences, and they reported that they were more likely to follow their principal's professional development recommendations.

Use of the checklist in the treatment group was moderate: 77 percent of principals surveyed who received the checklist reported viewing it, and 58 percent said they used it with one or more teachers. At the same time, 29 percent of control group principals (who were not emailed the checklist) reported that they had seen the checklist, and 10 percent reported using it with one or more teachers. The relatively moderate use of the checklist by treatment group principals, combined with the reports by some control group school leaders that they were using it, implies that the estimated impacts of using the checklist would be larger than the estimated impacts of receiving it.

Though distribution of the feedback conference checklist to principals and teachers had a few modest impacts, this study indicates that distributing the checklist is unlikely by itself to substantially alter feedback conferences, teachers' classroom practices, or student achievement, at least during the first school year in which the checklist is used. This study suggests that only a fraction of school leaders are likely to use the checklist if it is distributed in the low-cost manner followed in this study. But the checklist may also have failed to help principals overcome common barriers to effective feedback, such as providing critical comments to teachers or recommending appropriate professional development. The study results raise the possibility that additional (or different) investments might be necessary to improve school leaders' feedback conferences with teachers—for example, pairing training with written guidance.

Contents

Summary	i
Why this study?	1
What the study examined	3
The study	3
The research questions	4
What the study found	6
Providing the feedback conference checklist had no clear impact on principals' perceptions about the quality of the post-observation feedback conference	6
Provision of the checklist led to teachers reporting less dominance of the conference by the principal	7
Teachers who received the checklist were more likely to follow their principals' professional development recommendations	8
The feedback conference checklist had no clear impact on teachers' subsequent classroom observation rating scores	8
The feedback conference checklist had no clear impact on student achievement outcomes or on school report card grades	9
A little over half the treatment group principals reported using the checklist, and almost one-third of the control group principals reported seeing the checklist	10
Principals and teachers who used the checklist reported that it was useful but believed that it could lead to formulaic conferences	12
Implications of the study findings	14
Limitations of the study	14
Appendix A. Theory of action and literature about feedback	A-1
Appendix B. Feedback conference checklist	B-1
Appendix C. Control group guides for principals and teachers	C-1
Appendix D. Data, sample, and methodology	D-1
Appendix E. Treatment-on-the-treated analyses	E-1
Notes	Notes-1
References	Ref-1
Boxes	
1 Content of the feedback conference checklist	2
2 Data, sample, and methods	5

Figures

1	Most principals and teachers in New Mexico who used the feedback conference checklist reported that it was useful but that it could make the conference feel formulaic, 2015/16	13
A1	Theorized ideal teacher evaluation cycle	A-1
D1	Consolidated standards of reporting trials diagram for a study on the impact of a feedback conference checklist in New Mexico, 2015/16	D-5

Tables

1	Treatment and control conditions for the current study of a feedback conference checklist in New Mexico public schools, 2015/16	4
2	Impact of receipt of the feedback conference checklist on five aspects of the quality of feedback conferences, as reported by principals in sample New Mexico public schools, 2015/16	7
3	Impact of receipt of the feedback conference checklist on six aspects of the quality of feedback conferences, as reported by teachers in sample New Mexico public schools, 2015/16	8
4	Impact of receipt of the feedback conference checklist on teachers following principals' professional development recommendations in sample New Mexico public schools, 2015/16	9
5	Impact of receipt of the feedback conference checklist on subsequent classroom observation scores and on self-reported measures of teacher instructional practice in sample New Mexico public schools, 2015/16	10
6	Impact of receipt of the feedback conference checklist on student achievement test scores in sample New Mexico public schools, 2015/16	11
7	Impact of receipt of feedback conference checklist on school report card grades in sample New Mexico public schools, 2015/16	11
8	Principals' and teachers' self-reported viewing and use of the feedback conference checklist and accompanying testimonial video in sample New Mexico public schools, 2015/16 (percent)	12
D1	Control variables used in regression analyses in a study on the impact of a feedback conference checklist in sample New Mexico public schools, 2014/15	D-1
D2	Comparison of principal and teacher samples at baseline and of those who responded to both the spring 2015 and spring 2016 surveys, 2014/15 and 2015/16	D-6
D3	Comparison of school, principal, teacher, and student characteristics at baseline, 2014/15	D-7
D4	Baseline summary statistics for principal-reported feedback conference quality, 2014/15	D-8
D5	Baseline summary statistics for teacher-reported feedback conference quality, 2014/15	D-9
D6	Baseline summary statistics for teacher professional development outcomes, 2014/15	D-9
D7	Baseline summary statistics for teacher instructional practice, 2014/15	D-10
D8	Baseline summary statistics for student Partnership for Assessment of Readiness for College and Careers assessment scores, 2014/15	D-10
D9	Principal and teacher indexes on the content, structure, and utility of post-observation feedback conferences, 2014/15 and 2015/16	D-12
E1	Treatment-on-the-treated estimates on principal-reported conference quality, 2015/16	E-1
E2	Treatment-on-the-treated estimates on teacher-reported conference quality, 2015/16	E-2
E3	Treatment-on-the-treated estimates on professional development recommendation and take-up, 2015/16	E-2
E4	Treatment-on-the-treated estimates on teacher instructional practice, 2015/16	E-3
E5	Treatment-on-the-treated estimates on student achievement, 2015/16	E-4

Why this study?

Public school systems have undergone a sea change in how they evaluate teachers' performance. All but six states set timelines to include student achievement as a factor in teacher evaluations by the 2016/17 school year (National Council on Teacher Quality, 2016). New Mexico, the location of this study, launched a revised statewide teacher evaluation system called NMTEACH in the 2013/14 school year. Like revised teacher evaluation systems in other states, in the 2015/16 school year, the year of the study, NMTEACH assigned ratings to teachers on the basis of student achievement growth, scored classroom observations, and locally selected measures approved by the state, such as teacher attendance and student surveys.

A critical stage in the NMTEACH evaluation cycle is the feedback conversation that a school leader has with a teacher after each of two or three annual formal classroom observations. The school leader is to observe a teacher's classroom for at least 20 minutes and complete a 22-item observation rubric from the New Mexico Public Education Department called the NMTEACH Observation Rubric.¹ Within 10 days of the observation, the school leader must provide feedback to the teacher, including reviewing the scores assigned to the teacher on the rubric and recommending improvement and professional development. The feedback conversations have the potential to influence teaching practice by evaluating a teacher's instructional practices at multiple points each year, in place of the once-a-year overall teacher rating.

There is little research evidence about how to help school leaders communicate feedback to teachers in a way that leads to improvements in instruction and, ultimately, in student education outcomes. At the same time, research in behavioral economics has shown that informational interventions, such as "nudges," can be effective at changing behavior (Lavecchia, Liu, & Oreopoulos, 2016; Thaler & Sunstein, 2008).² Therefore, the New Mexico Public Education Department requested that the Regional Educational Laboratory Southwest design a rigorous evaluation of a low-cost 24-item checklist intended to promote practices in the feedback conference that the human resources management research literature has found to be effective (Myung & Martinez, 2013). The checklist is a modification of one created by the Carnegie Foundation for the Advancement of Teaching (Tang & Chow, 2007), adapted by the study team to the New Mexico context.

The changes in the past decade to teacher evaluation systems have increasingly required principals to act not only as managers of school organizations but also as instructional leaders (Green, 2010; Marshall, 2009; Shulman, Sullivan, & Glanz, 2008). Principals are expected to spend more time in classrooms providing feedback to teachers than they did under older evaluation systems. This feedback could improve teachers' instructional practice if the principals' observations included targeted recommendations for professional development in areas needing improvement (Rathel, Drasgow, & Christle, 2008; Taylor & Tyler, 2012). Although the literature on the efficacy of professional development is mixed, limited evidence suggests that teachers improve their instruction when they receive professional learning opportunities that are ongoing and closely connected to curriculum and instruction (Correnti, 2007; Correnti & Rowan, 2007; Supovitz & Turner, 2000). (See appendix A for a discussion of the theory of the teacher evaluation cycle and research related to the effects of feedback on performance.)

Feedback conversations have the potential to influence teaching practice by evaluating a teacher's instructional practices at multiple points each year, but there is little research evidence about how to help school leaders communicate feedback to teachers in a way that leads to improvements in instruction and, ultimately, in student education outcomes

The broader human resources management literature indicates that the features of effective feedback include two-way communication; timeliness, frequency, consistency, and accuracy; a focus on performance improvement; trust in the evaluator; identification of individual strengths and weaknesses; perceived fairness of the process; positive interpersonal treatment during the process; and goal setting (Cawley, Keeping, & Levy, 1998; DeNisi & Sonesh, 2011; Kluger & DeNisi, 1996; Locke & Latham, 2002; London & Smither, 2002).

Nevertheless, school principals have identified barriers to providing effective feedback, including a lack of time, perceived ineffectual performance measures (Donaldson, 2013), and difficulty and unwillingness in providing negative feedback to poorly performing teachers (Donaldson, 2013; Yariv, 2009). In a study of Chicago's teacher evaluation system, administrators listed the provision of useful feedback to teachers as an area in which they needed professional development (Sporte, Stevens, Healey, Jiang, & Hart, 2013).

The feedback conference checklist examined in this study is intended to remedy some of the shortcomings in feedback conferences by offering prompts to guide educators through conversations that include elements regarded as effective in the human resources literature. The checklist aims to structure a feedback conversation characterized by both positive and critical feedback, two-way rather than principal-dominated conversation, evidence from the classroom observation ratings, and concrete next steps (see box 1 for a summary of the checklist features). The feedback conference checklist does not influence the frequency of feedback (set at two or three times a school year in New Mexico) or alter the fundamentals of the teacher evaluation system.

The feedback conference checklist examined aims to structure a feedback conversation characterized by both positive and critical feedback, two-way rather than principal-dominated conversation, evidence from the classroom observation ratings, and concrete next steps

Box 1. Content of the feedback conference checklist

The feedback conference checklist is a version of the Carnegie Foundation Feedback Checklist, modified for the New Mexico context. The modifications did not change the structure of the checklist, but simply replaced generic terms about observation rubrics with references specifically to the NMTEACH Observation Rubric. The Carnegie Foundation checklist first recommends a list of documents that the principal and teacher should bring to the conference. It then guides principals and teachers through the stages of a formal post-observation conference using a 24-item checklist organized in the following sections:

1. Warm and clear opening (for example, "Thanks for meeting with me. What would you like to get out of this conversation?").
2. Focus on what's going well (for example, "What do you think went well for the lesson plan? In addition to what you mentioned, I noticed [POSITIVES]").
3. Identify challenges facing the teacher (for example, "What are some things you feel could have gone better? It sounds like what's challenging you is X, Y, and Z. Is that right?").
4. Generate ideas for addressing the teacher's challenges and prioritize next steps (for example, "Here are some professional development modules for you to consider").
5. End positively (for example, "Was this conversation helpful? Thank you for your insights").

Source: Tang and Chow (2007).

What the study examined

This study examined the impact of providing principals and teachers with the feedback conference checklist, along with a short video, on the perceived quality of feedback conferences following formal classroom observations, principals' recommendations for and teachers' take-up of professional development, and the quality of teachers' subsequent instructional practices as measured by principals' formal classroom observation scores and teachers' self-reported scores. The study also gathered exploratory evidence on the impact of the checklist on student achievement (school-average math and English language arts scores on the Partnership for Assessment of Readiness for College and Careers assessment) and school report card grades (reported as an A, B, C, D, or F of multiple measures of a school's student achievement) compiled annually by the New Mexico Public Education Department. Finally, the study documented how many recipients reported using the checklist and what they thought about it. The feedback conference checklist was distributed to principals and teachers in fall 2015, and all outcomes in the study are for the 2015/16 school year.

The study

In April 2015 the study team invited principals in all 786 of New Mexico's K–12 regular-instruction public schools to participate in the study about providing effective feedback to teachers.³ Of the 339 principals who consented to participate, the study team randomly selected half to be in the treatment group, with the other half constituting the control group. In fall 2015 principals in the treatment group received an email with an attachment containing a guide and a 24-item feedback conference checklist, plus a hyperlink to a three-minute professionally edited testimonial video of a principal who had used the Carnegie Foundation Feedback Checklist in another state.⁴ The guide's introduction encouraged the principal to use the checklist with all teachers in the school, suggested documents to have ready for the conference, and requested that principals not share the checklist with anyone outside the school (see appendix B for the principals' version of the treatment guide).

The principals in the control group received an email in fall 2015 with an attachment containing a two-page guide presenting the five stages of feedback that had been covered in professional development sessions about NMTEACH sponsored by the New Mexico Public Education Department in summer 2015 (see appendix C for the principal and teacher versions of the guide). The five stages start with a reflection or targeted question (for example, "What was your objective for the activity?"), provide evidence to the teacher (for example, "When you framed some questions ... 6 of 20 students were involved"), identify one to three areas of concern, give the teacher actions to take, and set a timeline for the actions.⁵

All study principals were asked to complete two rounds of online surveys—one in spring 2015, prior to random assignment, and one in spring 2016.

The study team solicited up to 10 randomly selected teachers in each school with a participating principal for voluntary participation in the study. Teachers in schools with a principal in the treatment group received an email with the same checklist guide as the principal but with a teacher-oriented introduction (see appendix B for the teachers' version), plus a hyperlink to a three-minute professionally edited testimonial video of a teacher who had used the checklist in a different state. (Treatment group principals were also instructed to distribute the checklist to all teachers in the school.) Teachers in schools with a control

This study examined the impact of providing principals and teachers with a feedback conference checklist on the perceived quality of feedback conferences, principals' recommendations for and teachers' take-up of professional development, and the quality of teachers' subsequent instructional practices

group principal received a two-page guide reminding teachers of the NMTEACH system and of their right to feedback resulting from a classroom observation within 10 calendar days of the observation (see appendix C for this email).

All study teachers were also asked to complete two rounds of online surveys—one in spring 2015, prior to random assignment, and one in spring 2016. Table 1 summarizes the differences between the treatment and the control conditions.

The research questions

The study addressed three research questions responding to the needs of the New Mexico Public Education Department:

1. Does providing the feedback conference checklist intervention, compared with the control condition, affect the quality and time burden of the post-observation feedback conference?
2. Does providing the feedback conference checklist intervention, compared with the control condition, affect principals' recommendations for professional development and the professional development that teachers take?
3. Does providing the feedback conference checklist intervention, compared with the control condition, improve the quality of teachers' instructional practices as rated on the NMTEACH classroom observation rubric?

The study addressed three research questions responding to the needs of the New Mexico Public Education Department, one exploratory question related to the proximal impacts of the intervention, and the extent to which treatment and control groups implemented the intervention

Table 1. Treatment and control conditions for the current study of a feedback conference checklist in New Mexico public schools, 2015/16

Participant and study component	Treatment group	Control group
Principals		
New Mexico Public Education Department–sponsored professional development for principals in summer 2014 with 2 hours devoted to feedback to teachers	✓	✓
List of documents to bring to each feedback conference (see appendix B)	✓	
24-item checklist to use during each feedback conference (see appendix B)	✓	
Three-minute video in which a principal testifies about his or her experience using the checklist (see appendix B)	✓	
Reminder about five stages of feedback to use in conferences, described in a New Mexico Public Education Department–sponsored principal professional development (see appendix C)		✓
Teachers		
List of documents to bring to each feedback conference (see appendix B)	✓	
24-item checklist to use during each feedback conference (see appendix B)	✓	
Three-minute video in which a teacher testifies about his or her experience using the checklist (see appendix B)	✓	
Reminder to teachers of the NMTEACH system and of their right to feedback resulting from a classroom observation within 10 calendar days of the observation (see appendix C)		✓

Source: Authors' compilation.

The study also addressed one exploratory question related to the proximal impacts of the intervention:

4. Does providing the feedback conference checklist intervention, compared with the control condition, raise student achievement on state standardized math and English language arts tests and raise the school report card grade generated by the New Mexico Public Education Department?

Finally, the study addressed the extent to which both the treatment and control groups implemented the intervention:

5. How extensively do principals and teachers in the treatment and control groups report using the feedback conference checklist, and how do they like using it?

See box 2 for a brief summary of the data, sample, and methods used in the study.

Box 2. Data, sample, and methods

Data and outcome measures

Participating principals took online principal surveys and participating teachers took online teacher surveys in spring 2015 and spring 2016. Those surveys are the sources of data used to answer research questions 1, 2, and 5. The data used to answer questions 3 and 4 are from administrative student, teacher, principal, and school records for the 2014/15 and 2015/16 school years for all teachers and students in study schools, including teachers' NMTEACH Observation Rubric scores and student achievement data, provided by the New Mexico Public Education Department.

The study team analyzed multiple measures under each research question. These included indexes of the quality of the feedback conference, indicators of professional development recommendations and take-up, scores on the NMTEACH Observation Rubric to measure instructional practice, and school-average math and English language arts scores on the Partnership for Assessment of Readiness for College and Careers (PARCC) assessment and school report card grades to measure student achievement. The NMTEACH Observation Rubric comprises four domains—planning and preparation, creating an environment for learning, teaching for learning, and professionalism—each of which contains five or six elements scored individually on a five-point scale. School report card grades are a composite, reported as an A, B, C, D, or F, of multiple measures of a school's student achievement compiled annually by the New Mexico Public Education Department. The study used responses on the spring 2016 surveys on whether the principal or teacher had seen the feedback conference checklist guide and used it in one or more feedback conferences to measure implementation. (See appendix D for a more detailed description of the outcome measures.)

Study sample

In April 2015 the study team invited principals in all 786 of New Mexico's K–12 regular-instruction public schools to participate in the study; 339 consented. In summer 2015 the study team selected half the consenting principals to be the treatment group and the other half to be the control group, using a blocked random selection procedure (see appendix D).

(continued)

Box 2. Data, sample, and methodology (continued)

About 63 percent of schools in the study sample were elementary schools, 21 percent were high schools, and 16 percent were middle or junior high schools; 68 percent of students in the sample were eligible for the federal school lunch program. Balance tables indicate that the initial treatment and control groups were equivalent on most key factors and were representative of the state as a whole (see appendix D). The overall attrition rates were 47 percent for principals and 69 percent for teachers, decreasing the power to detect impacts in research questions 1 and 2; attrition in this study means not completing the spring 2016 survey (see appendix D for more information about the attrition rates).

Methodology

One week before the first day of the 2015/16 school year (but after the spring 2015 survey for principals and teachers had been completed), the study team distributed the feedback conference checklist (see appendix B) as an electronic attachment in emails to the treatment group principals and up to 10 teachers at each treatment group school (some teachers received the email two weeks later). The study team distributed the control guides (see appendix C) as electronic attachments in emails to the control group principals and teachers. The feedback conference checklist guide for principals encouraged them to distribute the checklist to all teachers and school leaders within their school but not to disseminate it to anyone outside the school. The study team sent four reminders during the school year to the treatment group principals and teachers to use the feedback conference checklist.

For research questions 1–4 the study team estimated the impacts of providing the feedback conference checklist to principals and to teachers, regardless of whether they used it (called “intent-to-treat effects”) using hierarchical linear modeling to account for the nesting of teachers and students within schools and districts. The study team estimated the impacts of actually using the checklist (called “treatment-on-the-treated effects”) using two-stage least squares regression, with the treatment variable serving as the instrumental variable, which in turn predicted whether the teacher (or principal for the principal treatment-on-the-treated effect estimates) used the checklist. For research question 5 the study team compared the percentage of principals and of teachers who reported on the spring 2016 survey that they had seen and used the checklist, as well as their characterizations of the checklist. (See appendix D for more detail regarding the analytic models.)

What the study found

The following sections present the results on the impact of the feedback conference checklist.

Providing the feedback conference checklist had no clear impact on principals’ perceptions about the quality of the post-observation feedback conference

Neither the encouragement to use (“intent-to-treat”) nor self-reported actual use (“treatment-on-the-treated”) of the feedback conference checklist had a clear impact on principals’ perceptions of the quality of the post-observation conference. Among principals who were encouraged to use the feedback conference checklist, there were no statistically significant differences compared with the control group for indexes of post-observation conference quality (table 2). All of the estimated impact sizes were small, and none of the

Table 2. Impact of receipt of the feedback conference checklist on five aspects of the quality of feedback conferences, as reported by principals in sample New Mexico public schools, 2015/16

Principal-reported conference quality outcome measure	Treatment group mean (standard deviation)	Control group mean (standard deviation)	Estimated impact (standard error ^a)	Effect size ^b	Sample size
Supportive conference index (0–100 scale)	79.27 (11.94)	78.29 (11.26)	–0.459 (1.234)	–0.041	173
Specific feedback index (0–100 scale)	82.10 (12.95)	81.85 (13.72)	–1.517 (1.927)	–0.111	175
Data-driven conference index (0–100 scale)	77.23 (17.55)	76.30 (17.47)	–0.878 (2.621)	0.050	177
Well prepared, collaborative conference index (0–100 scale)	69.63 (14.62)	66.06 (14.02)	1.331 (1.664)	0.095	170
Conference duration (minutes)	31.59 (12.60)	31.27 (12.37)	–2.082 (1.253)	–0.168	167

Note: Although the treatment and control group means reported do not control for any differences in covariates, the results in the estimated impact column were estimated using a two-level hierarchical linear model with an indicator for treatment. See appendix D for a list of the included covariates and a description of how missing values were handled. The analysis sample included only respondents who completed both the spring 2015 and spring 2016 surveys.

a. See appendix D for a description of how standard errors were estimated.

b. Calculated by dividing the estimated impact by the standard deviation of the outcome for the control group.

Source: Authors' analysis of survey data collected for the study; see appendix D for more details.

Among principals who were encouraged to use the feedback conference checklist, there were no statistically significant differences compared with the control group for indexes of post-observation conference quality

coefficients was statistically significant. Among principals who reported using the checklist, there were no statistically significant differences compared with the control group in indexes of the quality of the post-observation feedback conference (see table E1 in appendix E).

Provision of the checklist led to teachers reporting less dominance of the conference by the principal

The feedback conference checklist affected one of the five teacher-reported indexes of post-observation conference quality: teachers in treatment group schools were less likely to report that their feedback conferences were dominated by the school leader than were teachers in schools that were not provided with the checklist intervention. Specifically, the principal-dominated conference index was 3.8 points lower (on a 100-point scale) for teachers in schools where the principal was encouraged to use the checklist (table 3) and more than 19 points lower in schools where the principal reported using the checklist (see table E2 in appendix E), compared with teachers in control group schools.⁶ The remaining four indexes were not affected by the receipt or reported use of the feedback conference checklist. There were no statistically significant differences in the duration of the conferences according to teachers. Hence, the feedback conference checklist did not lead to increased teacher reports of desired feedback practices identified by the research literature, such as more specific or more actionable feedback.

Table 3. Impact of receipt of the feedback conference checklist on six aspects of the quality of feedback conferences, as reported by teachers in sample New Mexico public schools, 2015/16

Teacher-reported conference quality outcome measure	Treatment group mean (standard deviation)	Control group mean (standard deviation)	Estimated impact (standard error ^a)	Effect size ^b	Sample size
Best practices conference index (0–100 scale)	70.32 (21.11)	70.31 (20.86)	0.771 (1.622)	0.037	815
Specific and actionable feedback conference index (0–100 scale)	65.81 (27.95)	66.23 (25.84)	1.280 (1.532)	0.050	840
Data-driven conference index (0–100 scale)	56.18 (24.33)	55.94 (23.66)	0.381 (1.600)	–0.016	815
Principal-dominated conference index (0–100 scale)	25.19 (19.60)	27.47 (20.47)	–3.848** (1.237)	–0.188	832
Well-rounded conference index (0–100 scale)	64.12 (23.65)	64.25 (22.80)	0.906 (1.541)	0.040	801
Conference duration (minutes)	33.85 (19.02)	31.70 (16.50)	0.974 (1.092)	0.059	829

** Statistically significant at $p < .01$.

Note: Although the treatment and control group means reported do not control for any differences in covariates, the results in the estimated impact column were estimated using a three-level hierarchical linear model with an indicator for treatment. See appendix D for a list of the included covariates and a description of how missing values were handled. The analysis sample included only respondents who completed the spring 2015 and spring 2016 surveys.

a. See appendix D for a description of how standard errors were estimated.

b. Calculated by dividing the impact estimate by the standard deviation of the outcome for the control group.

Source: Authors' analysis of survey data collected for the study; see appendix D for more details.

The principal-dominated conference index was 3.8 points lower (on a 100-point scale) for teachers in schools where the principal was encouraged to use the checklist and more than 19 points lower in schools where the principal reported using the checklist

Teachers who received the checklist were more likely to follow their principals' professional development recommendations

Teachers who received the checklist and the subset of teachers who reported that the checklist was used during their feedback conference were more likely to follow their principals' recommendations on professional development (table 4; see also table E3 in appendix E).⁷ For teachers who received the checklist the estimated impact was 5.6 percentage points. This finding is consistent with the checklist's prompts for the principal and teacher to commit to next steps by listing specific professional development opportunities that address challenges that the teacher faces. There were no additional clear impacts of the checklist on recommended professional development or on teachers' self-reported take-up of professional development independent of any recommendation.⁸

The feedback conference checklist had no clear impact on teachers' subsequent classroom observation rating scores

Neither the teachers who received the feedback conference checklist nor the subset of these teachers who reported using the checklist obtained significantly different scores on the NMTEACH Observation Rubric in the 2015/16 school year compared with teachers in the control group (table 5; see also table E4 in appendix E). However, teachers in the treatment group reported marginally higher and statistically significant self-ratings, collected

Table 4. Impact of receipt of the feedback conference checklist on teachers following principals' professional development recommendations in sample New Mexico public schools, 2015/16

Professional development recommendation and take-up outcome measures	Treatment group mean (standard deviation)	Control group mean (standard deviation)	Estimated impact (standard error ^a)	Effect size ^b	Sample size
Observation domain–specific professional development recommended by principal (indicator)	0.021 (0.143)	0.055 (0.227)	–0.029 (0.017)	–0.129	789
General professional development recommended by principal (indicator)	0.112 (0.316)	0.152 (0.359)	–0.022 (0.024)	–0.063	802
Take-up of any professional development by teacher (indicator) ^c	0.840 (0.367)	0.866 (0.342)	–0.029 (0.027)	–0.086	802
Teacher follows principal's professional development recommendation (indicator) ^c	0.943 (0.232)	0.884 (0.320)	0.056* (0.022)	0.174	784

* Statistically significant at $p < .05$.

Note: Although the treatment and control group means reported do not control for any differences in covariates, the results in the estimated impact column were estimated using a probit model with an indicator for treatment status. See appendix D for a list of the included covariates and a description of how missing values were handled. The analysis sample included only respondents who completed both the spring 2015 and spring 2016 surveys.

a. See appendix D for a description of how standard errors were estimated.

b. Calculated by dividing the impact estimate by the standard deviation of the outcome for the control group.

c. A teacher's report of taking up professional development is independent of any recommendation by the principal, whereas following a principal's recommendation for professional development means to take it up when recommended or not to take it up when not recommended.

Source: Authors' analysis of survey data collected for the study; see appendix D for more details.

The estimated impact on the proportion of teachers who received the checklist and who reported that they followed their principals' recommendations on professional development was 5.6 percentage points

in the teacher survey, on the teaching for learning domain compared with control group teachers.⁹ Because teachers are scored on the creating an environment for learning and the teaching for learning domains of the NMTEACH Observation Rubric multiple times during the school year and the feedback conference checklist focuses expressly on these domains, the receipt or use of the checklist may have made teachers more aware of these domains of their classroom practice and led them to work on them more.¹⁰

The feedback conference checklist had no clear impact on student achievement outcomes or on school report card grades

The study included exploratory analyses of the impact of the feedback conference checklist on student achievement outcomes to capture proximal impacts of teacher practice changes resulting directly from the feedback conversation. Students at schools where the principal and teachers received the feedback conference checklist did not score better (or worse) on their spring 2016 Partnership for Assessment of Readiness for College and Careers (PARCC) math and English language arts assessments than did students at the control group schools (table 6). After prior achievement, student demographic characteristics, school characteristics, and the randomization stratum of the school were controlled for and all student test scores were combined into one sample, students at the treatment group schools scored 0.009 standard deviation lower than did students at control group schools, a difference that is not statistically different from zero. The impacts on the school report card grades were positive but not statistically significant (table 7; see appendix D for a discussion of what is included in the school report card grades). However, given that only one

Table 5. Impact of receipt of the feedback conference checklist on subsequent classroom observation scores and on self-reported measures of teacher instructional practice in sample New Mexico public schools, 2015/16

Instructional practice outcome measure (NMTEACH Observation Rubric domains, 1–5 scale)	Treatment group mean (standard deviation)	Control group mean (standard deviation)	Estimated impact (standard error ^a)	Effect size ^b	Sample size
Principal ratings					
Planning and preparation	3.642 (0.625)	3.616 (0.635)	-0.018 (0.027)	-0.028	6,883
Creating an environment for learning	3.656 (0.518)	3.628 (0.518)	0.016 (0.024)	0.030	7,144
Teaching for learning	3.569 (0.538)	3.536 (0.533)	0.010 (0.020)	0.019	7,144
Professionalism	3.699 (0.613)	3.699 (0.618)	-0.013 (0.022)	-0.021	6,852
Teacher self-ratings					
Creating an environment for learning	3.885 (0.561)	3.816 (0.559)	0.056 (0.029)	0.100	860
Teaching for learning	3.747 (0.534)	3.711 (0.533)	0.066* (0.029)	0.123	856

* Statistically significant at $p < .05$.

Note: Although the treatment and control group means reported do not control for any differences in covariates, the results in the estimated impact column were estimated using a three-level hierarchical linear model with an indicator for treatment. See appendix D for a list of the included covariates and a description of how missing values were handled. The analysis sample included only teachers who had an observation score from the previous school year.

a. See appendix D for a description of how standard errors were estimated.

b. Calculated by dividing the impact estimate by the standard deviation of the outcome for the control group.

Source: Authors' analysis of administrative and survey data collected for the study; see appendix D for more details.

Teachers who received the feedback conference checklist did not obtain significantly different scores on the NMTEACH Observation Rubric in the 2015/16 school year compared with teachers in the control group

school year of outcomes was examined in this study, it is premature to conclude that the feedback conference checklist would have no impact on teachers' subsequent instructional practices and student achievement over the course of several years (see the "Limitations of the study" section for further discussion).

A little over half the treatment group principals reported using the checklist, and almost one-third of the control group principals reported seeing the checklist

Three-fourths of treatment group principals who were emailed the feedback conference checklist reported viewing it, and a little more than half (58 percent) reported using it (table 8). But only 28 percent of treatment group principals reported using the checklist with most or all teachers (15.8 percent with most teachers and 12.6 percent with all teachers), and only 28 percent indicated that they had viewed the three-minute video included in a hyperlink in the email that also included the feedback conference checklist as an attachment.

Despite instructions sent to the treatment group principals not to share the feedback conference checklist outside their school, there was evidence of sharing with control group principals and teachers. About 29 percent of control group principals reported viewing the feedback conference checklist, and about 10 percent reported using it. The analysis included

Table 6. Impact of receipt of the feedback conference checklist on student achievement test scores in sample New Mexico public schools, 2015/16

Student achievement outcomes (PARCC scores)	Treatment group mean (standard deviation)	x mean (standard deviation)	Estimated impact (standard error ^a)	Effect size ^b	Sample size
Elementary school math	0.17 (1.03)	0.19 (1.02)	0.003 (0.031)	0.003	30,004
Elementary school English language arts	-0.11 (0.95)	-0.07 (0.95)	0.013 (0.022)	0.013	29,606
Middle school math	0.02 (0.97)	-0.02 (1.00)	-0.006 (0.030)	-0.006	27,259
Middle school English language arts	-0.02 (0.92)	-0.07 (0.92)	0.017 (0.034)	0.018	27,200
High school math	-0.26 (0.97)	-0.15 (0.91)	-0.043 (0.030)	-0.043	20,330
High school English language arts	0.12 (1.11)	0.15 (1.10)	-0.003 (0.062)	-0.003	20,546

PARCC is Partnership for Assessment of Readiness for College and Careers.

Note: Although the treatment and control group means reported do not control for any differences in covariates, the results for the estimated impact column were estimated using a three-level hierarchical linear model with an indicator for treatment. See appendix D for a list of the included covariates and a description of how missing values were handled. The analysis sample included only students who had an achievement score from the previous school year.

a. See appendix D for a description of how standard errors were estimated

b. The effect on a student's English language arts or math PARCC score divided by the standard deviation of all students' PARCC scores.

Source: Authors' analysis of administrative data collected for the study; see appendix D for more details.

Table 7. Impact of receipt of feedback conference checklist on school report card grades in sample New Mexico public schools, 2015/16

Student achievement outcome (school report card grades)	Treatment group mean (standard deviation)	Control group mean (standard deviation)	Estimated impact (standard error ^a)	Effect size ^b	Sample size
Increased report card grade	0.35 (0.48)	0.30 (0.46)	0.044 (0.052)	0.044	285
Decreased report card grade	0.27 (0.45)	0.38 (0.49)	-0.092 (0.053)	-0.092	285
Overall report card grade	1.93 (1.30)	1.88 (1.18)	0.114 (0.115)	0.096	285

Note: Although the treatment and control group means reported do not control for any differences in covariates, the results in the estimated impact column were estimated using a two-level hierarchical linear model with an indicator for treatment. See appendix D for a list of the included covariates and a description of how missing values were handled. The overall report card grade was quantified in the same way grade point averages are constructed. For example, an A was scored as 4 and a C as 2.

a. See appendix D for a description of how standard errors were estimated.

b. The effect on the overall grade, divided by the standard deviation of all grades.

Source: Authors' analysis of administrative data collected for the study; see appendix D for more details.

Table 8. Principals’ and teachers’ self-reported viewing and use of the feedback conference checklist and accompanying testimonial video in sample New Mexico public schools, 2015/16 (percent)

Spring 2016 survey item	Principals		Teachers	
	Treatment group (n = 95)	Control group (n = 84)	Treatment group (n = 456)	Control group (n = 473)
Saw checklist	74.7	28.6	56.2	25.0
Used checklist	57.9	9.5	30.7	14.6
Saw video	28.4	1.2	21.1	1.7
Used checklist with a few teachers	25.3	1.2	—	—
Used checklist with half of teachers	4.2	2.4	—	—
Used checklist with most teachers	15.8	3.6	—	—
Used checklist with all teachers	12.6	2.4	—	—
Checklist was used in one conference	—	—	10.0	5.1
Checklist was used in two conferences	—	—	14.2	7.7
Checklist was used in three or more conferences	—	—	6.4	1.7

— Not available.

Source: Authors’ compilation of survey data collected for the study; see appendix D for details.

Only 28 percent of treatment group principals reported using the checklist with most or all teachers, and only 28 percent indicated that they had viewed the three-minute testimonial video linked to in the email with the checklist

responses from all participants who reported using the checklist, whether they were in the treatment group or the control group. It is possible, though, that control group principals’ self-reported usage rates were inflated if some principals who reported “yes” on the survey were referring to the control group guide rather than to the treatment group guide. Regardless, the combination of the moderate usage rate among the treatment group principals (58 percent) and the noticeable percentage of control group principals who reported using it implies that the estimated impacts of using the feedback conference checklist would be larger than the estimated impacts of receiving it. The size of the estimates of feedback conference checklist use, though generally statistically insignificant, bear this out (see appendix E).

Teachers’ self-reported use of the feedback conference checklist was much lower than principals’. About 56 percent of treatment group teachers reported seeing the feedback checklist and 31 percent reported using it. About 25 percent of control group teachers reported seeing the feedback conference checklist, and about 15 percent reported using it. Again, the reported usage rate of the feedback conference checklist by control group teachers may have been inflated if teachers were thinking of the control group guide when they answered this question.

Principals and teachers who used the checklist reported that it was useful but believed that it could lead to formulaic conferences

Principals and teachers who reported using the feedback conference checklist tended to agree on its characteristics. A majority agreed that it was easy to use, provided a helpful structure for the feedback conference, and helped teachers commit to a set of next steps (figure 1). However, principals and teachers also agreed that the checklist could make the conference feel formulaic. Principals typically felt that the checklist helped somewhat with providing more critical feedback but not with providing positive feedback. The average teacher reported no appreciable impact on either critical or positive feedback. Approximately equal proportions of principals and teachers agreed that the checklist took too much time as agreed that it did not.

Figure 1. Most principals and teachers in New Mexico who used the feedback conference checklist reported that it was useful but that it could make the conference feel formulaic, 2015/16



Note: The box plots display the distribution of spring 2016 survey responses about use of the feedback conference checklist from all participants who reported using the checklist, regardless of treatment or control group assignment. Not all schools where teachers responded had a principal who also responded and vice versa. When the responses to a survey item are ordered from lowest to highest, the middle value of the responses to a given survey item is shown as the vertical bar in the box, and the 25th and 75th percentiles of the responses at the edges of the box. The whiskers extending from the left and right of the box indicate the range of the bottom quartile and top quartile of principals' and teachers' ratings on the given survey item. The dots indicate outlier values.

Source: Authors' calculations from survey data collected for the study.

Implications of the study findings

The results suggest that distributing a feedback conference checklist and short accompanying video electronically at low cost and sending four reminders to use the checklist during the year do not substantially alter feedback conferences, at least over the short run. It is possible that the feedback conference checklist and video could have greater impact in later school years. It is also possible that the impact would have been greater if the checklist distribution had been supported with more resources, such as training in using the checklist, but further research would be needed to examine this hypothesis.

Providing the checklist had, at best, moderate impacts on a few outcomes. Specifically, teachers viewed the feedback conference as less dominated by the principal, and they followed their principals' professional development recommendations more closely. Teachers also reported higher self-ratings on one domain of the NMTEACH Observation Rubric, the teaching for learning domain.

Use of the feedback conference checklist was moderate overall. Of the principals who were encouraged to use the checklist, about 75 percent reported seeing it, and 58 percent reported using it in post-observation feedback sessions with at least some teachers. Of the teachers who were encouraged to use the checklist, about 56 percent reported seeing it and 31 percent reported using it. This relatively moderate use was achieved on the basis of four encouragement emails from the study team, without any paired professional development or involvement of the state department of education. Both principals and teachers reported that the checklist was easy to use and provided helpful structure for the feedback conference, and both gave mixed responses about whether it took too much time to use or helped provide more critical and positive feedback.

Given that only one school year of outcomes was examined in this study, it is premature to conclude that the feedback conference checklist would have no impact on teachers' subsequent instructional practices and student achievement over the course of several years. For example, teachers' professional development recommended as a result of the feedback conference checklist may not have been completed by the time principals re-rated teachers' instructional practices or students were tested.

The main implication of this study for school districts and state departments of education is that, at least in the first year, a feedback conference checklist in and of itself does not substantially alter the quality of principal–teacher feedback conferences across the board. To boost the quality of feedback conferences, it is likely that, at a minimum, the checklist would need to be paired with more intensive training and encouragement and additional steps to embed it in school and district practices and procedures.

Limitations of the study

The study's measures of teachers' instructional practices and student achievement occurred only months after the initial distribution of the feedback conference checklist, which may be too soon to detect impacts, especially on professional development and subsequent changes to instructional practice and student achievement. Teachers' self-reports of taking professional development came from surveys that they completed at the end of the same school year in which the checklist was disseminated, so it is possible that teachers had

The main implication of this study for school districts and state departments of education is that, at least in the first year, a feedback conference checklist in and of itself does not substantially alter the quality of principal–teacher feedback conferences across the board

not yet been able to complete the professional development recommended in a feedback conference. For example, if a principal recommended in April 2016 that a teacher take professional development, the teacher might not have been able to do so until summer 2016 or later, which was after the spring 2016 survey was completed. Likewise, this study would be unable to identify potential impacts in subsequent years that such professional development could have on teacher practice or student achievement.

A second pair of limitations is that use of the feedback conference checklist was not mandatory and its distribution could not be perfectly restricted to schools in the treatment group. The effect of both these conditions was increased because the checklist was distributed by email and its use was not encouraged through any other support, other than email reminders. The moderate usage rate of the checklist among the treatment group combined with the checklist's spread to a little more than one-quarter of the control group likely lowered the estimated impacts of the checklist on study outcomes and decreased the precision of the estimates. However, the take-up rate provides a useful gauge for school districts and state departments of education as they anticipate usage among principals and teachers of other checklists or information-only guides. Although the moderate usage rate is a serious limitation in estimating the impacts of the checklist, a clear benefit of electronically distributing it is its substantially lower cost compared with in-person or online professional training, which themselves have imperfect take-up rates.

Another limitation comes from the in-person professional development that, theoretically, all New Mexico principals received about giving feedback to their teachers, which took place prior to the distribution of the checklist. The professional development reduced the contrast between the principals who did and those who did not receive the feedback conference checklist in this study. The reduced contrast means that the checklist might have larger impacts where principals receive no prior training about providing feedback.

Last, because of high attrition in the spring 2016 surveys and the relatively high annual teacher turnover, the study could not detect small impacts of the feedback conference checklist for research questions 1 and 2, which used survey measures as outcome variables. However, the rate of attrition was nearly equal for the treatment and control groups, and in general the differences between the outcome measures for the treatment and control groups were not substantively large. So although a larger sample might have made some of the estimated impacts reported in this study statistically significant, the impacts would remain small, assuming that attrition is random.

Although the moderate usage rate is a serious limitation in estimating the impacts of the checklist, a clear benefit of electronically distributing it is its substantially lower cost compared with in-person or online professional training, which themselves have imperfect take-up rates

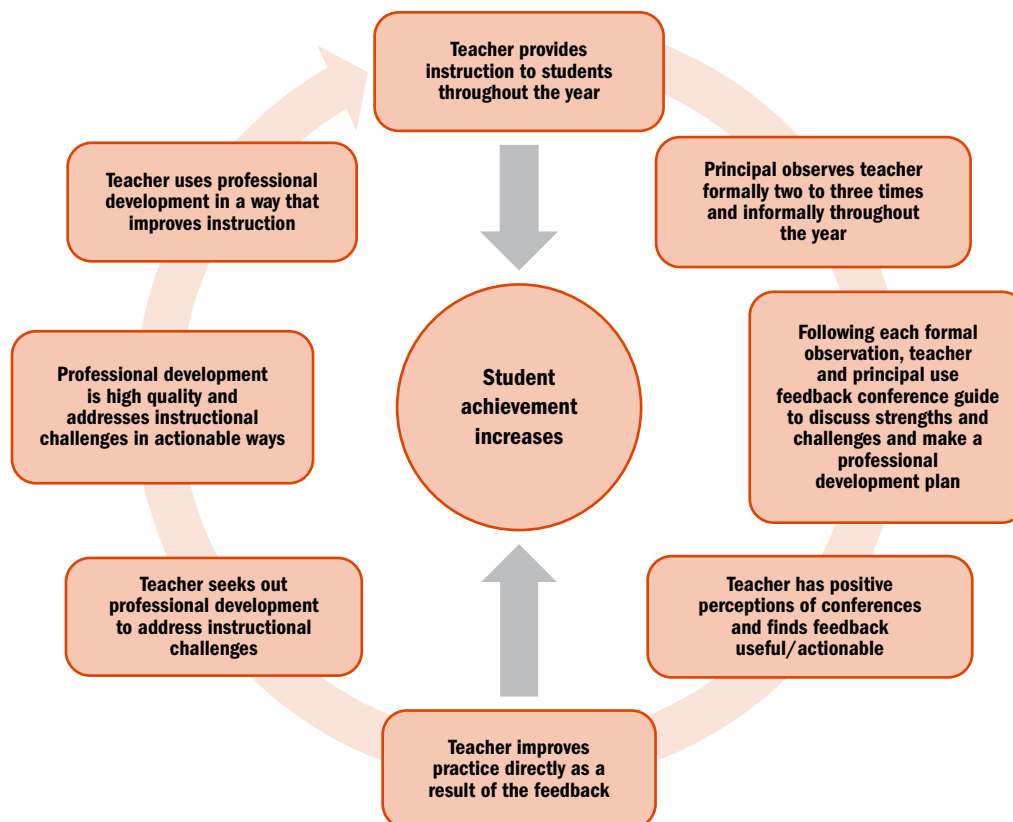
Appendix A. Theory of action and literature about feedback

This appendix describes the theory of action related to the feedback conference checklist intervention examined in this study. Figure A1 shows the theorized ideal teacher evaluation cycle.

Theory of action

Principal feedback may improve teacher instructional practice if principals make targeted recommendations to teachers for professional development in specific areas identified in classroom observations. Such subjective evaluations designed to provide teachers with feedback may have positive lasting impacts on teacher practices and behaviors and on student achievement, according to some studies (Rathel et al., 2008; Taylor & Tyler, 2012). Although the literature on the efficacy of professional development is mixed, some evidence suggests that teachers improve their instruction when they receive ongoing professional learning opportunities that are closely connected to curriculum and instruction (Correnti, 2007; Correnti & Rowan, 2007; Supovitz & Turner, 2000) and to school district priorities (Garet, Porter, Desimone, Birman, & Yoon, 2001; Penuel, Fishman, Yamaguchi, & Gallagher, 2007). If professional development were tightly linked to teachers' observed classroom practices, it could become increasingly relevant to their future practices (Ball & Cohen, 1999; Little, 1993; Wilson & Berne, 1999).

Figure A1. Theorized ideal teacher evaluation cycle



Source: Authors' construction.

Education experts have pointed to concrete strategies for improving principals' communication with teachers about instruction. These strategies are incorporated in the design of the detailed feedback checklist that is the subject of this study. The first strategy is for principals to provide a "learning-oriented assessment" that develops a shared understanding of evaluation criteria (Tang & Chow, 2007) and encourages teachers to take an active role in assessing their own performance so they can see the conference as useful (Chalies, Ria, Bertone, Trohel, & Durand, 2004; Holland, 1989; Tang & Chow, 2007). A second strategy is to use a wide range of prompts for teacher reflection, which may encourage productive teacher–principal communication (Williams & Watson, 2004). And a third strategy is to use objective teacher data during conferences (Holland, 1989; Rockoff, Staiger, Kane, & Taylor, 2012).

These education-specific findings comport with research on human resources management: effective feedback includes two-way communication; timely, frequent, consistent, and accurate feedback; a focus on improving performance; trust in the evaluator; identification of individual strengths and weaknesses; perceived fairness of the process; positive interpersonal treatment during the process; and goal setting (Cawley et al., 1998; DeNisi & Sonesh, 2011; Kluger & DeNisi, 1996; Locke & Latham, 2002; London & Smither, 2002).

Obstacles to productive principal feedback

Newer teacher evaluation systems generally require more frequent and more extensive formal feedback from school leaders to teachers about instruction, but they retain some limitations from older teacher evaluation systems. These include inflated ratings, little substantive feedback, growth plans misaligned with personnel evaluation findings, school leaders not taking responsibility for evaluations, and low validity and reliability of principals' judgments about teaching (Frase & Streshly, 1994; Medley & Coker, 1987; Peterson, 2000; Stodolsky, 1984).

In addition, the reformed teacher evaluation systems pose new challenges. Principals cite a lack of time and the perceived inadequacy of performance measures as reasons for disengagement from regular observation and feedback (Donaldson, 2013) and difficulty and unwillingness in providing negative feedback to poorly performing teachers (Donaldson, 2013; Yariv, 2009). A 2011 study found that the Chicago school district's expectations for principal–teacher conferences before and after observation did not align with principals' and teachers' actual practices (Sartain, Stoelinga, & Brown, 2011). In another study of Chicago's teacher evaluation system, administrators disclosed a need for professional development in providing useful feedback to teachers (Sporte et al., 2013). These obstacles signal the importance of designing strategies to enhance the quality of post-observation conferences intended to improve teacher instruction.

Appendix B. Feedback conference checklist

This appendix presents the feedback conference checklist that was emailed to treatment group principals and teachers. The checklist is identical for principals and teachers except for the introductory material.

Principal version of the treatment group guide to the feedback conference checklist



Checklist for New Mexico Principals' Provision of Feedback to Teachers School year 2015–2016



Dear Principals:

Purpose of the checklist

The checklist is adapted from a guide developed by the Carnegie Foundation for the Advancement of Teachers to the New Mexico evaluation system. It includes effective elements identified by research of performance feedback in teaching and other professions.

We highly encourage you to use the checklist this year during every feedback conversation you have after formally observing your teachers this school year. Please disseminate this to all of your school leaders who conduct formal observations and to all of your teachers. We encourage you to discuss the checklist at one of the first faculty meetings this year.

View this 5-minute video for a principal's viewpoint

To learn more about the checklist and why it might be useful to you, we invite you to view this video {link here} for one principal's testimonial about how the checklist helped her give effective feedback.

The checklist is part of a research study

You are receiving this checklist because you have agreed to be a part of a research study conducted by the Regional Education Laboratory (REL) Southwest. For more information about the study, please see {REL Southwest project website URL here}.

Your participation and feedback on surveys in this study will help REL Southwest researchers give independent feedback to New Mexico PED about NMTEACH. The study is testing out two types of guidance for principals about formative evaluation feedback to teachers. To ensure the success of the study, we therefore ask that you **not share or forward this document or the checklist to anyone outside your school**. It is critical that not all principals receive this checklist so that we may compare outcomes in schools depending on which of the two types of guidance they received. Sharing the guide with others outside your school undermines the study.

To encourage its use, REL Southwest has also disseminated this checklist to teachers in your school who have also consented to be in the study. Depending on the number of teachers in your school, REL Southwest researchers may not have asked all of them to participate in the study, and thus not all teachers will have received a copy of this guide from REL Southwest.

What to expect from the research study

Participation in the study is voluntary and will not impact principals' or teachers' effectiveness ratings. REL Southwest invited all public schools in the state to participate in the study. Among those that agreed to participate, REL Southwest selected at random half to receive this checklist. You are among the schools selected to obtain the checklist at the beginning of the 2015–2016 school year.

In each school that agrees to participate in the study, REL Southwest will solicit principals and teachers to fill out an on-line, 30-minute survey once in spring 2015 and again in spring 2016. We will email an Amazon gift card in the amount of \$25 to principals and to teachers each time they complete the survey. Answers will be confidential, and will only be reported in aggregate form in a public research report.

If you have any questions about the research study, please do not hesitate to contact us at teacher_feedback_study@relsouthwest@sedl.org.

How the guide fits into New Mexico's teacher evaluation system

New Mexico's Public Education Department requires that principals (or school leaders) observe teachers formally two or three times per year (with 20 minute observations), and informally throughout the school year (with 3–5 minute “walk-throughs”). The enclosed conversation protocol is for use after each formal observation.

The formal observation occurs three times per school year if the teacher is being observed by a single observer, and twice a year if they are observed by two observers (such as a principal and assistant principal). For teachers being observed three times, the observations must take place by October 15, December 20, and April 15. Teachers being observed twice must be observed by December 20 and April 15.

When formally observing teachers, principals must use the NMTEACH Observation Rubric (available at [URL here].) The principal must provide feedback to the teacher within 10 calendar days of each formal observation. The formal, formative feedback contains three types of information: (1) scores from each of the domains in the observation rubric, (2) how these scores are tied to the narrative feedback from the observer, and (3) recommendations for professional development through online modules. ***The enclosed guide walks you through these steps.***

In addition, all teachers must create a professional development plan with their principal within the first 40 days of the school year. ***The enclosed guide walks you through the creation of all elements of a professional development plan.***

In addition, teachers who receive a rating of ineffective or minimally effective must be placed on growth plans, which require more frequent observations of teachers, and support for teachers to improve through instructional coaches or professional development courses. Teachers who do not show improvement after 90 days of being placed on the growth plan can be recommended for dismissal or reassignment. ***Because school districts have different guidance about growth plans, this guide does not include prompts for the creation of growth plans.***

The formal observations that are the subject of this guide are one part of a larger teacher evaluation system that was mandated in all New Mexico public schools starting in 2013–2014. For details on the teacher evaluation rating system and how it works, see {URL}.

We hope you find the conversation protocol useful to your practice!

New Mexico Principal-Teacher Post-Observation Conversation Checklist

Applies to Teachers in Groups A, B, and C

Key
Green text: Principal's prompt
Purple text: Teacher's prompt

Teacher
Principal
Date

Documents to have in hand for the conversation

Principal should have:

The completed hard copy of NMTEACH Observation Rubric or else the print-out of observation scores & notes from <i>Reflect</i> system
Teacher's most recent online report card
A copy of the teacher's most recent professional development plan
If applicable, a copy of the teacher's professional growth plan

Teacher should have:

Artifacts of student work and/or students' teaching and learning
A hard copy of his or her lesson plan for the lesson that that principal observed
A copy of the teacher's most recent professional development plan
If applicable, a copy of the teacher's professional growth plan
If different from the PDP, a list of professional development activities the teacher has participated in the past two school years

A. Warm and clear opening

1. Both teacher and principal acknowledge each other's time. Thanks for meeting with me.	<input type="checkbox"/>
2. Principal provides summary overview of the conversation. I would like to discuss your lesson, review your scores overall, and then discuss elements where your practice is strong, elements where your practice could improve, and link those to how you can take your instruction to the next level.	<input type="checkbox"/>
3. Principal asks and then teacher clearly states aim for the conversation. In this conversation I am looking forward to ...	<input type="checkbox"/>
4. Teacher states the lesson's objective and learning goals. My aim for the lesson was ...	<input type="checkbox"/>
5. Principal paraphrases and affirms the teacher's (1) goal of the lesson, and (2) aim for this conversation. I hear that in this lesson you hoped students would learn {XYZ} and that you hope to discuss {XYZ}.	<input type="checkbox"/>
6. Principal summarizes the scores and the narrative feedback from each scored domain of the NMTEACH Observation rubric.	<input type="checkbox"/>

B. Focus on what's going well

7. **Principal asks teacher to reflect on what went well in the lesson overall, using student artifacts if possible. I noticed students were...**
-
8. **Principal paraphrases what the teacher identifies as going well. So what I heard you say was...**
-
9. **Principal comments on concrete, specific things that went well.** Looking at the observation rubric, principal identifies all elements from Domains 2 and 3 rated highly effective or exemplary. If no elements were so rated, principal identifies the 3 elements where the teachers' practices are most effective. *I noticed your lesson was relatively strong in establishing a culture for learning. I rated it as exemplary because your practice improved from an already strong position last time I observed you ...*
-
10. **THE THREE STRONGEST ELEMENTS FROM DOMAIN 2 OR 3 IN THE OBSERVED LESSON.** *Principal writes answers here.*
-
11. **(ONLY APPLICABLE FOR THE LAST CONFERENCE OF THE YEAR.) Principal comments on concrete, specific things that went well related to teacher's professionalism.** Principal identifies all elements from Domains 1 and 4 rated highly effective or exemplary. If no elements were so rated, principal identifies the 2–3 elements where the Principal judges the teacher to be most effective. Note whether these positive findings link with action steps in teacher's PDP. *Over this school year during my observations and walkthroughs, I've noticed your growing knowledge of NM's content standards for [XX] and how you are orienting lessons around those standards. Did that online course about Common Core standards help?*
-
12. **(ONLY APPLICABLE FOR THE LAST CONFERENCE OF THE YEAR.) THE THREE STRONGEST ELEMENTS FROM DOMAIN 1 OR 4 IN EITHER THE OBSERVED LESSON OR OVER THE COURSE OF THE CURRENT SCHOOL YEAR.** *Principal writes answers here.*
-

C. Identify challenges facing the teacher

13. **Principal asks teacher to reflect on what changes she should make to improve the lesson next time, using student artifacts if possible. Next time, I would change how I introduced the standard... I would like some help addressing student actions such as ...**
-
14. **Principal paraphrases the teacher's identified challenges. Is sounds like what's challenging you is X, Y, & Z. Is this right?**
-
15. **Principal comments on concrete, specific challenges. Teacher responds.** Principal lists all elements from Domains 2 and 3 where the teacher's level of performance was rated as ineffective or minimally effective. If no elements were so rated, principal identifies the 1–3 elements where the teacher could continue to improve. If Domain 2 or 3 is rated effective or minimally effective, then principal and teacher must identify a professional growth plan. *I noticed the lesson included negative interactions between you and students. I rated element 2A ineffective because ...*
-
16. **THE ONE TO THREE ELEMENTS FROM DOMAIN 2 OR 3 IN THE OBSERVED LESSON THAT COULD MOST IMPROVE.** *Principal writes answers here.*
- 1.
- 2.
- 3.
-

-
17. (ONLY APPLICABLE FOR THE LAST CONFERENCE OF THE YEAR.) *Principal comments on concrete, specific things that could improve related to teacher's professionalism. Teacher responds.* Principal lists all elements from Domains 1 and 4 where the teacher's level of performance was rated as ineffective or minimally effective. If no elements were so rated, principal identifies the 1–3 elements where the teacher could continue to improve.

Over this school year during my observations and walkthroughs, I've noticed that you are struggling to connect to the non-English speaking families of your students. Let's discuss how to access translation services from the district to help.



-
18. (ONLY APPLICABLE FOR THE LAST CONFERENCE OF THE YEAR.) THE ONE TO THREE ELEMENTS FROM DOMAIN 1 OR 4 IN EITHER THE OBSERVED LESSON OR OVER THE COURSE OF THE CURRENT SCHOOL YEAR THAT COULD MOST IMPROVE. *Principal writes answers here.*

1.

2.

3.

D. Generate ideas for addressing teacher's challenges

19. *Principal offers ideas for addressing the teacher's challenges from Steps 16 & 18. The following online professional development modules might address these challenges...*



20. *Teacher responds to ideas by either adding or suggesting amendments.*



21. *Principal and teacher collaborate to prioritize the ideas and commit to next steps. List specific professional development modules if applicable. Principal writes answers here. Teacher prompts for clarification: Can you elaborate on that? Can you give me an example?*

Top priority:

2nd priority:

3rd priority:

One thing teacher suggests she will try differently tomorrow. _____

E. End positively

- | | | |
|-----|---|--------------------------|
| 22. | Principal asks if this conversation was helpful. Teacher gives feedback on what worked and what didn't work. My goal for this conversation was {AIM} and I appreciated your {specific feedback} about what did work and {specific feedback} about what didn't work. | <input type="checkbox"/> |
| 23. | Principal makes a final positive statement, recognizing growth and progress. | <input type="checkbox"/> |
| 24. | Teacher thanks principal for time and insights. | <input type="checkbox"/> |



Checklist for New Mexico Post-Observation Feedback Conversation Between Teachers and Principals

School year 2015–2016



Dear Teachers:

Purpose of the checklist

Enclosed is a checklist for you and your school leader to use at all of your post-observation feedback conversations this school year.

The checklist is adapted from a guide developed by the Carnegie Foundation for the Advancement of Teachers to the New Mexico evaluation system. It includes effective elements identified by research of performance feedback in teaching and other professions.

We highly encourage you to use the checklist this year during every feedback conversation you have after a school leader formally observes and rates your classroom this year.

Since the checklist is part of a research study to compare two different types of guidance educators about feedback conversations, we ask that you NOT disseminate it to anyone outside your school building.

View this 3-minute video for a teacher's viewpoint

To learn more about the checklist and why it might be useful to you, we invite you to view this video (https://youtu.be/Rabqn5an_jE).

The checklist is part of a research study

You are receiving this checklist because you have agreed to be a part of a research study conducted by the Regional Education Laboratory (REL) Southwest. For more information about the study, please see <http://relsouthwest.sedl.org/nmpf>.

Your participation and feedback on surveys in this study will help REL Southwest researchers give independent feedback to New Mexico PED about NM TEACH. The study is testing out two types of guidance for principals and teachers about formative evaluation feedback to teachers.

To ensure the success of the study, we therefore ask that you ***not share or forward this document or the checklist to anyone outside your school.*** It is critical that not all teachers and principals receive this checklist so that we may compare outcomes in schools depending on which of the two types of guidance they received. Sharing the guide with others outside your school undermines the study.

To encourage its use, REL Southwest has given this same checklist to your school principal and asked him or her to use it in all the post-observation feedback sessions this school year. Depending on the number of teachers in your school, REL Southwest researchers may not have asked all of them to participate in the study, and so not all teachers will have received a copy of this guide from REL Southwest.

What to expect from the research study

Participation in the study is voluntary and will not impact teachers' or principals' effectiveness ratings. REL Southwest invited principals in all public schools in the state to

participate in the study. Among those that agreed to participate, REL Southwest selected at random half to receive this checklist. You are among the schools selected to obtain the checklist at the beginning of the 2015–2016 school year.

In each school that agrees to participate in the study, REL Southwest asked principals and teachers to fill out an on-line, 30-minute survey once in spring 2015, and we will again a final time in spring 2016. We will email an Amazon gift card in the amount of \$25 to teachers and to principals each time they complete the survey. Answers will be confidential, and will only be reported in aggregate form in a public research report.

If you have any questions about the research study, please do not hesitate to contact us at FeedbackStudy-rels@rand.org.

How the guide fits into New Mexico's teacher evaluation system

New Mexico's Public Education Department requires that principals (or school leaders) observe teachers formally two or three times per year (with 20 minute observations), and informally throughout the school year (with 3–5 minute “walk-throughs”).

The formal observation occurs three times per school year if the teacher is being observed by a single observer, and twice a year if the teacher is observed by two observers (such as a principal and assistant principal). For teachers being observed three times, the observations must take place by October 15th, December 20, and April 15th. Teachers being observed twice must be observed by December 20 and April 15th.

When formally observing your classroom, principals must use the NM Teach Observation Rubric (available at http://www.nctq.org/docs/NMTEACH_Rubric.pdf). The principal must provide feedback to you within 10 calendar days of each formal observation. The formal, formative feedback contains three types of information: (1) scores from each of the domains in the observation rubric, (2) how these scores are tied to the narrative feedback from the observer, and (3) recommendations for professional development through online modules. ***The enclosed guide walks you through these steps.***

In addition, all teachers must create a professional development plan with their principal within the first 40 days of the school year. ***The enclosed guide walks you through the creation of all elements of a professional development plan.***

Teachers who receive a rating of ineffective or minimally effective must be placed on growth plans, which require more frequent observations of teachers, and support for teachers to improve through instructional coaches or professional development courses. Teachers who do not show improvement after 90 days of being placed on the growth plan can be recommended for dismissal or reassignment. ***Because school districts have different guidance about growth plans, this guide includes prompts for the creation of growth plans.***

The formal observations that are the subject of this guide are one part of a larger teacher evaluation system that was mandated in all New Mexico public schools starting in 2013–2014. For details on the teacher evaluation rating system and how it works, see <http://www.ped.state.nm.us/ped/NMTeachIndex.html>.

We hope you find the conversation protocol useful to your practice!

New Mexico Principal-Teacher Post-Observation Conversation Checklist

Applies to Teachers in Groups A, B, and C

Key
Green text: Principal's prompt
Purple text: Teacher's prompt

Teacher
Principal
Date

Documents to have in hand for the conversation

Principal should have:

The completed hard copy of NM Teach Observation Rubric or else the print-out of observation scores & notes from <i>Reflect</i> system
Teacher's most recent online report card
A copy of the teacher's most recent professional development plan
If applicable, a copy of the teacher's professional growth plan

Teacher should have:

Artifacts of student work and/or students' teaching and learning
A hard copy of his or her lesson plan for the lesson that that principal observed
A copy of the teacher's most recent professional development plan
If applicable, a copy of the teacher's professional growth plan
If different from the PDP, a list of professional development activities the teacher has participated in the past two school years

A. Warm and clear opening

1. Both teacher and principal acknowledge each other's time. Thanks for meeting with me.	<input type="checkbox"/>
2. Principal provides summary overview of the conversation. I would like to discuss your lesson, review your scores overall, and then discuss elements where your practice is strong, elements where your practice could improve, and link those to how you can take your instruction to the next level.	<input type="checkbox"/>
3. Principal asks and then teacher clearly states aim for the conversation. In this conversation I am looking forward to ...	<input type="checkbox"/>
4. Teacher states the lesson's objective and learning goals. My aim for the lesson was ...	<input type="checkbox"/>
5. Principal paraphrases and affirms the teacher's (1) goal of the lesson, and (2) aim for this conversation. I hear that in this lesson you hoped students would learn {XYZ} and that you hope to discuss {XYZ}.	<input type="checkbox"/>
6. Principal summarizes the scores and the narrative feedback from each scored domain of the NM TEACH Observation rubric.	<input type="checkbox"/>

B. Focus on what's going well

7. **Principal asks teacher to reflect on what went well in the lesson overall, using student artifacts if possible. I noticed students were...**
-
8. **Principal paraphrases what the teacher identifies as going well. So what I heard you say was...**
-
9. **Principal comments on concrete, specific things that went well.** Looking at the observation rubric, principal identifies all elements from Domains 2 and 3 rated highly effective or exemplary. If no elements were so rated, principal identifies the 3 elements where the teachers' practices are most effective. *I noticed your lesson was relatively strong in establishing a culture for learning. I rated it as exemplary because your practice improved from an already strong position last time I observed you ...*
-
10. **THE THREE STRONGEST ELEMENTS FROM DOMAIN 2 OR 3 IN THE OBSERVED LESSON.** *Principal writes answers here.*
-
11. **(ONLY APPLICABLE FOR THE LAST CONFERENCE OF THE YEAR.) Principal comments on concrete, specific things that went well related to teacher's professionalism.** Principal identifies all elements from Domains 1 and 4 rated highly effective or exemplary. If no elements were so rated, principal identifies the 2–3 elements where the Principal judges the teacher to be most effective. Note whether these positive findings link with action steps in teacher's PDP. *Over this school year during my observations and walkthroughs, I've noticed your growing knowledge of NM's content standards for [XX] and how you are orienting lessons around those standards. Did that online course about Common Core standards help?*
-
12. **(ONLY APPLICABLE FOR THE LAST CONFERENCE OF THE YEAR.) THE THREE STRONGEST ELEMENTS FROM DOMAIN 1 OR 4 IN EITHER THE OBSERVED LESSON OR OVER THE COURSE OF THE CURRENT SCHOOL YEAR.** *Principal writes answers here.*
-

C. Identify challenges facing the teacher

13. **Principal asks teacher to reflect on what changes she should make to improve the lesson next time, using student artifacts if possible. Next time, I would change how I introduced the standard... I would like some help addressing student actions such as ...**
-
14. **Principal paraphrases the teacher's identified challenges. Is sounds like what's challenging you is X, Y, & Z. Is this right?**
-
15. **Principal comments on concrete, specific challenges. Teacher responds.** Principal lists all elements from Domains 2 and 3 where the teacher's level of performance was rated as ineffective or minimally effective. If no elements were so rated, principal identifies the 1–3 elements where the teacher could continue to improve. If Domain 2 or 3 is rated effective or minimally effective, then principal and teacher must identify a professional growth plan. *I noticed the lesson included negative interactions between you and students. I rated element 2A ineffective because ...*
-
16. **THE ONE TO THREE ELEMENTS FROM DOMAIN 2 OR 3 IN THE OBSERVED LESSON THAT COULD MOST IMPROVE.** *Principal writes answers here.*
- 1.
- 2.
- 3.
-

-
17. (ONLY APPLICABLE FOR THE LAST CONFERENCE OF THE YEAR.) *Principal comments on concrete, specific things that could improve related to teacher's professionalism. Teacher responds.* Principal lists all elements from Domains 1 and 4 where the teacher's level of performance was rated as ineffective or minimally effective. If no elements were so rated, principal identifies the 1–3 elements where the teacher could continue to improve.

Over this school year during my observations and walkthroughs, I've noticed that you are struggling to connect to the non-English speaking families of your students. Let's discuss how to access translation services from the district to help.

-
18. (ONLY APPLICABLE FOR THE LAST CONFERENCE OF THE YEAR.) THE ONE TO THREE ELEMENTS FROM DOMAIN 1 OR 4 IN EITHER THE OBSERVED LESSON OR OVER THE COURSE OF THE CURRENT SCHOOL YEAR THAT COULD MOST IMPROVE. *Principal writes answers here.*

1.

2.

3.

D. Generate ideas for addressing teacher's challenges

19. *Principal offers ideas for addressing the teacher's challenges from Steps 16 & 18. The following online professional development modules might address these challenges...*

-
20. *Teacher responds to ideas by either adding or suggesting amendments.*

-
21. *Principal and teacher collaborate to prioritize the ideas and commit to next steps. List specific professional development modules if applicable. Principal writes answers here. Teacher prompts for clarification: Can you elaborate on that? Can you give me an example?*

Top priority:

2nd priority:

3rd priority:

One thing teacher suggests she will try differently tomorrow. _____

E. End positively

- | | | |
|-----|---|--------------------------|
| 22. | Principal asks if this conversation was helpful. Teacher gives feedback on what worked and what didn't work. My goal for this conversation was {AIM} and I appreciated your {specific feedback} about what did work and {specific feedback} about what didn't work. | <input type="checkbox"/> |
| 23. | Principal makes a final positive statement, recognizing growth and progress. | <input type="checkbox"/> |
| 24. | Teacher thanks principal for time and insights. | <input type="checkbox"/> |

Appendix C. Control group guides for principals and teachers

This appendix presents copies of the guidance that was sent to control group principals and teachers.

Principal version of control group guide



Guidance for New Mexico Principals About Provision of Feedback to Teachers School year 2015–2016



Dear Principals:

Purpose of this guide

This guide summarizes training offered by the New Mexico PED to principals about effective feedback to teachers after formal classroom observations using NM TEACH Observation rubric.

We encourage you to use the enclosed five stages for effective feedback to teachers. We hope that you will find it useful, and we encourage you to adapt it to your needs.

Five Stages of Feedback from Principals to Teachers

1. Start with a reflection or targeted question.

Example: “What was your objective for the activity?”

2. Present evidence to the teacher.

Example: “When you framed some questions to promote student achievement 6 of 20 students were involved.”

3. Identify 1–3 areas of concern.

Example: “The discussion about the word problem was teacher centered, providing minimal opportunity for students to discuss in pairs or in small groups.”

4. Give the teacher actions they should take.

Example: “As you plan your lessons, identify sample problems for students to discuss and analyze in pairs or groups.”

5. Set a timeline by which the action should be taken.

This guide is part of a research study

You are receiving this guide because you have agreed to be a part of a research study conducted by the Regional Education Laboratory (REL) Southwest. The study is testing out two types of guidance for principals about feedback to teachers. For more information about the study, please see <http://relsouthwest.sedl.org/nmpf>.

Your participation and feedback on surveys in this study will help REL Southwest researchers give independent feedback to New Mexico PED about NM TEACH. To ensure the success of the study, we ask that you **not share or forward this document to anyone outside your school**. It is critical that not all principals receive this guide so that we may compare outcomes in schools depending on the type of guidance they received. Sharing the guide with others outside your school undermines the study.

What to expect from the research study

Participation in the study is voluntary and will not impact principals' or teachers' effectiveness ratings. REL Southwest invited all public schools in the state to participate in the study. Among those that agreed to participate, REL Southwest selected at random half to receive this guide. You are among the schools selected to obtain this guide at the beginning of the 2015–2016 school year.

In each school that agrees to participate in the study, REL Southwest asked principals and teachers to fill out an on-line, 30-minute survey once in spring 2015, and we will again a final time in spring 2016. We will email an Amazon gift card in the amount of \$25 to principals and to teachers each time they complete the survey. Answers will be confidential, and will only be reported in aggregate form in a public research report.

If you have any questions about the research study, please do not hesitate to contact us at FeedbackStudy-relsw@rand.org.

How the guide fits into New Mexico's teacher evaluation system

New Mexico's Public Education Department requires that principals (or school leaders) observe teachers formally two or three times per year (with 20 minute observations), and informally throughout the school year (with 3–5 minute “walk-throughs”). The 5 stages for feedback list above are intended to help you structure the conversations that occur after the formal classroom observations.

The formal observation occurs three times per school year if the teacher is being observed by a single observer, and twice a year if the teacher is observed by two observers (such as a principal and assistant principal). For teachers being observed three times, the observations must take place by October 15th, December 20, and April 15th. Teachers being observed twice must be observed by December 20 and April 15th.

When formally observing teachers, principals must use the NM Teach Observation Rubric (available at http://www.nctq.org/docs/NMTEACH_Rubric.pdf). The principal must provide feedback to the teacher within 10 calendar days of each formal observation. The formal, formative feedback contains three types of information: (1) scores from each of the domains in the observation rubric, (2) how these scores are tied to the narrative feedback from the observer, and (3) recommendations for professional development through online modules.

In addition, teachers who receive a rating of ineffective or minimally effective must be placed on growth plans, which require more frequent observations of teachers, and support for teachers to improve through instructional coaches or professional development courses. Teachers who do not show improvement after 90 days of being placed on the growth plan can be recommended for dismissal or reassignment.

The formal observations that are the subject of this guide are one part of a larger teacher evaluation system that was mandated in all New Mexico public schools starting in 2013–2014. For details on the teacher evaluation rating system and how it works, see <http://www.ped.state.nm.us/ped/NMTeachIndex.html>.



Guidance for Post-Observation Feedback to Teachers

School year 2015–2016



Dear Teachers:

Purpose of this document

This document is to remind you that you have a right to receive feedback from your schools leaders within 10 days of their formal classroom observations that are to occur 2–3 times in the 2015–2016 school year as a part of the state teacher evaluation system called NMTEACH.

As a part of NMTEACH, a school leader is supposed to formally observe and rate your classroom 2–3 times this school year. They are supposed to observe your class for a minimum of 20 minutes each time.

The formal observation occurs three times per school year if the teacher is being observed by a single observer, and twice a year if they are observed by two observers (such as a principal and assistant principal). For teachers being observed three times, the observations must take place by October 15th, mid-January and April 15th. Teachers being observed twice must be observed by mid-January and April 15th.

When formally observing teachers, principals must use the NMTEACH Observation Rubric (available at http://www.nctq.org/docs/NMTEACH_Rubric.pdf). The principal must provide feedback to the teacher within 10 calendar days of each formal observation. The formal, formative feedback contains three types of information: (1) scores from each of the domains in the observation rubric, (2) how these scores are tied to the narrative feedback from the observer, and (3) recommendations for professional development through online modules.

In addition, teachers who receive a rating of ineffective or minimally effective must be placed on growth plans, which require more frequent observations of teachers, and support for teachers to improve through instructional coaches or professional development courses. Teachers who do not show improvement after 90 days of being placed on the growth plan can be recommended for dismissal or reassignment.

For details on the teacher evaluation rating system and how it works, see <http://www.ped.state.nm.us/ped/NMTeachIndex.html>.

This document is part of a research study

You are receiving this because you have agreed to be a part of a research study conducted by the Regional Education Laboratory (REL) Southwest. The study is testing out two types of guidance about post-observation feedback for principals and teachers, along with a reminder of how often formal observations and post-observation feedback should occur.

For more information about the study, please see <http://relsouthwest.sedl.org/nmpf>.

To ensure the success of the study, we ask that you **not share or forward this document to anyone outside your school**. It is critical that not all teachers receive this document so that we may compare outcomes in schools depending on the type of information they received.

What to expect from the research study

Participation in the study is voluntary and will not impact teachers' or principals' effectiveness ratings. REL Southwest invited all public schools in the state to participate in the study. Among those that agreed to participate, REL Southwest selected at random half of schools to receive this document. You are among the schools selected to obtain the guide at the beginning of the 2015–2016 school year.

In each school that agrees to participate in the study, REL Southwest asked principals and teachers to fill out an on-line, 30-minute survey once in spring 2015, and we will again a final time in spring 2016. We will email an Amazon gift card in the amount of \$25 to teachers and to principals each time they complete the survey. Answers will be confidential, and will only be reported in aggregate form in a public research report.

If you have any questions about the research study, please do not hesitate to contact us at FeedbackStudy-relsw@rand.org.

Appendix D. Data, sample, and methodology

This appendix describes the study data, analysis sample, and methodology.

Study data

This study used both primary and secondary data sources. Primary data collected for the study consisted of principal and teacher surveys administered in spring 2015 and spring 2016. The secondary data consisted of administrative data from the New Mexico Public Education Department (NM PED) about schools (such as school level, district name, and charter status), principals (such as demographic characteristics, years of experience, and education), teachers (such as NMTEACH Observation Rubric scores, NMTEACH summative scores, demographic characteristics, years of experience, and education attainment), and students (such as demographic characteristics and Partnership for Assessment of Readiness for College and Careers [PARCC] assessment scores) for the 2014/15 and 2015/16 school years. In addition, the publicly available school report card grade measure was used as a complementary outcome measure for student achievement at the school level.

Table D1 lists variables used as controls for each research question along with the source of the data for the variable (principal survey, teacher survey, or administrative data). Variables for which treatment and control groups were equivalent at baseline (that is, prior to random assignment) were omitted from the impact estimates (for example, the percentage of students in a school eligible for the federal school lunch program in the 2014/15 school year).

Table D1. Control variables used in regression analyses in a study on the impact of a feedback conference checklist in sample New Mexico public schools, 2014/15

Covariate	Research question					Data source
	1. Conference feedback quality	2. Teacher professional development	3. Quality of instruction	4. Student achievement	5. Perception of checklist	
Student-level covariates						
English learner student indicator				✓		NM PED student demographic file
Eligibility for the federal school lunch program indicator				✓		NM PED student demographic file
Poverty indicator ^a				✓		NM PED student demographic file
Four indicators for race/ethnicity (American Indian/Alaska Native, Black, Hispanic, other race/ethnicity). White is the reference category.				✓		NM PED student demographic file
Baseline student PARCC scores				✓		NM PED student demographic file
Baseline school report card grade				✓		NM PED administrative file

(continued)

Table D1. Control variables used in regression analyses in a study on the impact of a feedback conference checklist in New Mexico, 2014/15 (continued)

Covariate	Research question					Data source
	1. Conference feedback quality	2. Teacher professional development	3. Quality of instruction	4. Student achievement	5. Perception of checklist	
Principal-level covariates						
Male	✓	✓	✓			NM PED principal demographic file
Years of service	✓	✓	✓			NM PED principal demographic file
Compensation	✓	✓	✓			NM PED principal demographic file
Four indicators for race/ethnicity (American Indian/Alaska Native, Asian, Black, Hispanic). White is the reference category.	✓	✓	✓			NM PED principal demographic file
Three indicators for highest degree (doctorate, master's, education specialist). Bachelor's is the reference category.	✓	✓	✓			NM PED principal demographic file
Baseline outcome index score of quality of conference	✓					Principal survey data
Six measures of principal-reported professional development quality (sufficiently resourced, easy to access, easy to customize, sufficiently available, convenient, aligned with observation rubric)		✓				Principal survey data
Teacher-level covariates						
Male	✓	✓	✓			NM PED teacher demographic file
Years of service	✓	✓	✓			NM PED teacher demographic file
Compensation	✓	✓	✓			NM PED teacher demographic file
Four indicators for race/ethnicity (American Indian/Alaska Native, Asian, Black, Hispanic). White is the reference category)	✓	✓	✓			NM PED teacher demographic file
Three indicators for highest degree (doctorate, master's, education specialist). Bachelor's is the reference category.	✓	✓	✓			NM PED teacher demographic file
NMTEACH, creating an environment for learning (domain 2) average, principal rating			✓			NMTEACH evaluation data
NMTEACH, teaching for learning (domain 3) average, principal rating			✓			NMTEACH evaluation data
NMTEACH, creating an environment for learning (domain 2) average, self-rating			✓			Teacher survey data
NMTEACH, teaching for learning (domain 3) average, self-rating			✓			Teacher survey data
Baseline outcome index score about quality of conference	✓					Teacher survey data
Baseline measure of given professional development outcome		✓				Teacher survey data

(continued)

Table D1. Control variables used in regression analyses in a study on the impact of a feedback conference checklist in New Mexico, 2014/15 (continued)

Covariate	Research question					Data source
	1. Conference feedback quality	2. Teacher professional development	3. Quality of instruction	4. Student achievement	5. Perception of checklist	
School-aggregate covariates						
Three indicators for level of school (high; junior high; middle). Elementary is the reference category.	✓	✓	✓			NM PED school demographic file
Five indicators for race/ethnicity (American Indian, Asian, Black, Hispanic, Hawaiian/Pacific Islander). White is the reference category.	✓	✓	✓	✓		NM PED student demographic file
Study stratum	✓	✓	✓	✓		Study variable
Treatment status	✓	✓	✓	✓		Study variable

NM PED is New Mexico Public Education Department. NMTEACH is New Mexico's state system for educator evaluation. PARCC is Partnership for Assessment of Readiness for College and Careers.

Note: Control variables are for 2014/15, and the study outcomes are for 2015/16. PARCC scores are used to measure student achievement.

a. This is an indicator for whether student receives services such as the Supplemental Nutrition Assistance Program and Temporary Assistance for Needy Families.

Source: Authors' compilation of administrative data obtained from New Mexico Public Education Department.

Outcome measures

For research question 1 the study team constructed four indexes using principal survey data and five indexes using teacher survey data to summarize principals' and teachers' perceptions of the quality of post-observation conferences. (See the methodology subsection of this appendix for a description of how these indexes were developed.) In addition, the average duration of conferences reported by principals and teachers in the surveys provided an outcome measure of time burden.

For research question 2 the outcome measures included teacher responses from the spring 2016 survey on whether their principals recommended they take professional development during the 2015/16 year on general topics or on specific topics aligned with items in the NMTEACH Observation Rubric. Additional outcome measures came from teacher survey responses about whether teachers completed any professional development. Last, the study team created an indicator to measure whether a teacher followed the principal's professional development recommendations. The indicator equaled zero only if the teacher did not take professional development that was recommended by the principal; otherwise, it equaled 1 (so that teachers are not penalized for taking professional development that was not recommended by the principal). If the principal did not recommend professional development and the teacher did not take professional development, the teacher was also coded as following recommendations.

For research question 3 the study team used ratings on the 2015/16 NMTEACH Observation Rubric to construct outcome measures of the quality of teacher instruction. Item-level scores in each domain were averaged across the two or three teacher observations within the school year and combined into four domain scores. Also, the study team collected

teachers' self-reports in the surveys on the 10 items in two domains of NMTEACH—creating an environment for learning and teaching for learning—and created domain-level averages of the self-reported scores.

For research question 4 student spring 2016 PARCC scores in math and English language arts measured at the school level were the outcome measures of student achievement. Students in New Mexico in grades 3–11 take the PARCC assessments annually, so the study team considered 2015/16 scale scores in grades 4–11 as outcomes, controlling for spring 2015 scores. In addition, the study team also used as a second outcome measure the school report card grade. Each school report card grade, published annually by NM PED, is a composite of multiple measures of student achievement in reading, math, and English language arts and is reported as an A, B, C, D, or F. These measures include value-added measures; the percentage of students who are proficient in a given year; the rate at which an individual student's test scores grow; the rate at which average test scores grow; and, for high schools, the graduation rate. An additional 5 percent of the school grade is determined by attendance measures and student responses to an annual survey designed to determine whether teachers are using good learning practices.

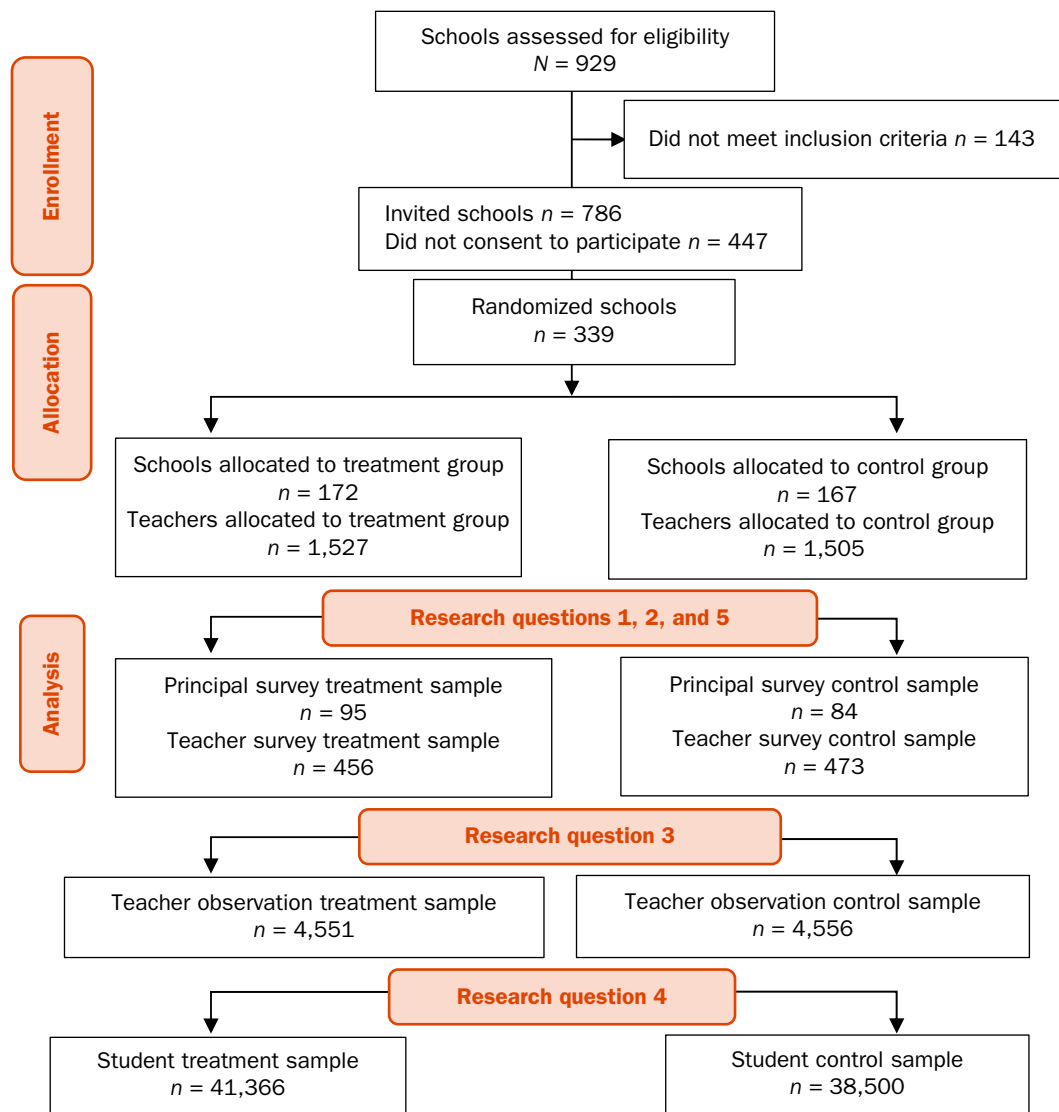
Research question 5 measures implementation fidelity using responses to spring 2016 surveys completed by principals and teachers on whether the study participants had seen the feedback conference checklist and whether they had used it.

Sample

Recruitment of principals and schools into the study started in spring 2015, when the study team assessed the 929 schools in the state for eligibility. Next, the study team invited 786 public school principals (all kindergarten through grade 12 public school principals of regular-instruction public schools in the state, including charter schools but excluding such special-purpose schools as credit recovery schools, special education–only schools, and preschools) to participate in the research study. Among the 786 invited principals, 339 consented to participate in the study (figure D1). In summer 2015 the study team conducted a blocked random assignment of schools to the treatment and control groups. Each school was assigned to one of three levels (elementary, middle, or high school) and to one of four geographic locations (Metro Albuquerque, North Central, Northwest, or Southeast). Charter schools were assigned to their own stratum. Within each of the resulting 13 strata, half the schools were randomly assigned to the treatment group, and half were assigned to the control group.

Outcome data for research questions 1, 2, and 5 rely on surveys of principals and teachers. Of the 339 principals who consented to participate in the study at baseline, 179 completed both the spring 2015 and the spring 2016 surveys. In each school where the principal consented to participate in the study as of spring 2015, the study team randomly sampled up to 10 teachers¹¹ for recruitment to participate in the teacher surveys. Recruitment yielded 929 teachers who completed the spring 2015 and spring 2016 surveys. To answer research question 3, NM PED provided the study team teacher observation data for the 2014/15 and 2015/16 school years for 4,551 teachers in treatment group schools and 4,556 teachers in control group schools. For research question 4, 2015/16 achievement test data for 41,366 students in treatment group schools and 38,500 students in control group schools were obtained from NM PED administrative records.

Figure D1. Consolidated standards of reporting trials diagram for a study on the impact of a feedback conference checklist in New Mexico, 2015/16



Source: Authors' compilation

Attrition is a concern for research questions 1 and 2, for which the analysis relies on survey data. Overall data attrition was 47 percent for principals: of the 339 principals who consented to participate in the study, 179 completed the spring 2016 survey (45 percent attrition for the treatment group and 50 percent for the control group). Attrition was even higher for teachers: of the 3,032 teachers who were contacted in the 339 study schools, 929 completed the spring 2016 survey—70 percent attrition for the treatment group and 69 percent for the control group.¹² The difference in attrition rates across treatment group schools and control group schools is 5 percentage points for principals and 1 percentage point for teachers. The study team examined whether attrition affected the baseline equivalence when schools were assigned to the treatment and control groups (table D2). At time of assignment, statistically significant differences in characteristics between principals and teachers in treatment and control groups were found for principals whose highest degree was a bachelor's and for teachers for years in current district, compensation, percentage

who were White, and percentage who were Hispanic. At time of the spring 2016 survey, no significant differences in characteristics were found among principals. Among teachers, significant differences were found in percentage who were White, percentage who were Hispanic, those whose highest degree was a bachelor's, and those whose highest degree was a master's. The study team included controls for all covariates that at assignment showed statistically significant differences across the treatment group and the control group.

Table D2. Comparison of principal and teacher samples at baseline and of those who responded to both the spring 2015 and spring 2016 surveys, 2014/15 and 2015/16

Principal and teacher characteristics	Treatment group		Control group		Significance at time of assignment	Significance after follow up
	At time of assignment	Responded to spring 2015 and spring 2016 surveys	At time of assignment	Responded to spring 2015 and spring 2016 surveys		
Principal characteristics						
Male (percent)	28	26	32	32		
Years in district (mean)	6.3	6.4	5.7	6.6		
Years of service (mean)	12	13	12	12		
Compensation amount (mean \$)	74,556	74,333	73,550	74,229		
American Indian/Alaska Native (percent)	5	4	2	2		
Asian (percent)	0	0	2	2	*	
Black (percent)	< 1	1	< 1	1		
Hispanic (percent)	36	35	39	34		
White (percent)	58	61	56	59		
Doctoral degree (percent)	6	6	4	4		
Master's degree (percent)	80	78	81	85		
Bachelor's degree (percent)	7.10	8	15	10	**	
No degree (percent)	1.20	1	0	0		
Total number	169	95	165	84		
Teacher characteristics						
Male (percent)	21	16	21	14		
Years in district (mean)	7.4	7.2	8	7.3	**	
Years of service (mean)	11	10	11	10		
Annual compensation (mean \$)	46,138	44,691	44,753	45,271	*	
American Indian/Alaska Native (percent)	5	5	4	3		
Asian (percent)	2	3	2	2		
Black (percent)	1	< 1	1	1		
Hispanic (percent)	35	33	30	25	***	***
White (percent)	57	59	63	68	***	***
Doctoral degree (percent)	< 1	< 1	< 1	< 1		
Master's degree (percent)	39	37	42	45		**
Bachelor's degree (percent)	59	59	57	53		*
No degree (percent)	< 1	2	< 1	1		
Certified teacher (percent)	99	99	99	99		
Total number	1,527	456	1,505	473		

* Statistically significant at $p < .05$; ** statistically significant at $p < .01$; *** statistically significant at $p < .001$.

Note: Significance at time of assignment indicates whether there were statistically significantly different group means among baseline survey respondents. Significance after follow up indicates whether there were statistically significantly different group means among spring 2016 survey respondents.

Source: Authors' calculations based on administrative data from the New Mexico Public Education Department.

To check the success of randomization, the study team compared baseline characteristics with overall state characteristics for schools, principals, teachers, and students (table D3). The purposes were to test for the baseline statistical equivalence of the treatment and control groups for research questions 3 and 4 and to validate that treatment and control group schools represented the state substantively. Randomization was successful in balancing all of the school-, principal-, teacher-, and student-level characteristics with the exception of receiving a C on the school report card. Significantly more schools in the treatment group received a C on their school report card than in the control group.

Tables D4–D8 present baseline summary statistics for all outcome measures.

Table D3. Comparison of school, principal, teacher, and student characteristics at baseline, 2014/15

Characteristic	Statewide	Treatment group	Control group
School characteristic			
Average number of teachers	26	26	26
Average number of students	441	442	443
School report card grade A (percent)	16	11	16
School report card grade B (percent)	20	16	23
School report card grade C (percent)	26	35***	20
School report card grade D (percent)	23	21	22
School report card grade F (percent)	16	16	17
High school (percent)	24	21	20
Middle school (percent)	17	17	14
Elementary school (percent)	53	61	64
Total number of schools	892	171	167
Principal characteristic			
Male (percent)	36	26	32
Hispanic (percent)	35	34	37
White (percent)	59	62	56
Other race/ethnicity (percent)	6	4	7
Years in district (mean)	6	7	7
Years of service (mean)	13	13	13
Doctorate degree (percent)	4	6	6
Master's degree (percent)	82	79	82
Bachelor's degree (percent)	10	8	11
Total number of schools	762	171	167
Teacher characteristic			
Male (percent)	22.7	22	21
Years in district (mean)	7.5	7.2	7.5
Years of service (mean)	11.1	11	11
Compensation amount (mean \$)	44,889	45,063	44,669
American Indian/Alaska Native (percent)	3.1	4.4	3.8
Asian (percent)	2	2.3	2
Black (percent)	1.1	1.3	1.5
Hispanic (percent)	33.8	35	32

(continued)

Table D3. Comparison of school, principal, teacher, and student characteristics at baseline, 2014/15 (continued)

Characteristic	Statewide	Treatment group	Control group
White (percent)	59.8	56	61
Doctorate degree (percent)	< 1	< 1	< 1
Master's degree (percent)	41.9	40	41
Bachelor's degree (percent)	54.6	57	57
No degree (percent)	2.2	1.7	1.5
Certified teacher (percent)	96.8	96	96
Teacher observation score	3.4	3.4	3.4
Total number of schools	862	169	164
Student characteristic			
Poverty level (percent)	68	69	68
English learner students (percent)	15	17	16
Students in special education (percent)	13	13	13
Gifted students (percent)	4	4	4
American Indian/Alaska Native (percent)	12	14	12
Asian (percent)	9	9	10
Black (percent)	17	16	19
Hispanic (percent)	58	58	58
White (percent)	26	25	25
Standardized math PARCC	0.00	0.00	0.04
Standardized English language arts PARCC	0.00	-0.02	-0.01
Total number of schools	876	163	157

*** Statistically significant at $p < .001$.

PARCC is the Partnership for Assessment of Readiness for College and Careers assessments.

Source: Authors' calculations based on administrative data from the New Mexico Public Education Department

Table D4. Baseline summary statistics for principal-reported feedback conference quality, 2014/15

Outcome	Data source	Baseline treatment group mean (standard deviation)	Baseline control group mean (standard deviation)	Number in treatment group analysis sample	Number in control group analysis sample
Supportive conference (0–100 scale)	Principal survey	78.59 (12.19)	76.12 (12.70)	94	79
Specific feedback conference (0–100 scale)	Principal survey	83.80 (14.10)	81.22 (15.81)	95	80
Data-driven conference (0–100 scale)	Principal survey	76.92 (19.79)	75.15 (19.68)	95	81
Well-prepared, collaborative conference (0–100 scale)	Principal survey	65.84 (16.19)	61.50 (15.67)	94	76
Conference duration (minutes)	Principal survey	34.24 (13.72)	32.23 (11.77)	92	75

Note: Baseline treatment and control group means presented are baseline summary statistics for the sample included in the analysis.

Source: Authors' calculations based on survey data collected for this study.

Table D5. Baseline summary statistics for teacher-reported feedback conference quality, 2014/15

Outcome	Data source	Baseline mean (standard deviation)		Number of teachers in analysis sample		Number of teachers at assignment		Number of schools in analysis sample	
		Treatment group	Control group	Treatment group	Control group	Treatment group	Control group	Treatment group	Control group
Best practices conference (0–100 scale)	Teacher survey	75.32 (18.88)	73.72 (19.26)	394	421	1,361	1,347	147	142
Data-driven conference (0–100 scale)	Teacher survey	63.81 (21.86)	61.91 (22.09)	394	421	1,361	1,365	147	144
Specific and actionable feedback conference (0–100 scale)	Teacher survey	68.99 (24.21)	69.71 (24.49)	406	434	1,365	1,349	148	142
Principal-dominated conference (0–100 scale)	Teacher survey	35.22 (19.29)	33.05 (18.96)	400	432	1,365	1,365	148	144
Well-rounded conference (0–100 scale)	Teacher survey	69.23 (21.17)	68.51 (21.05)	391	410	1,361	1,331	147	140
Conference duration (minutes)	Teacher survey	31.42 (16.16)	30.39 (14.47)	402	427	1,389	1,355	150	143

Note: Baseline and control group means presented are baseline summary statistics for the sample included in the analysis. Assignment is assignment of schools to treatment and control groups.

Source: Authors' calculations based on survey data collected for this study.

Table D6. Baseline summary statistics for teacher professional development outcomes, 2014/15

Outcome	Data source	Baseline mean (standard deviation)		Number of teachers in analysis sample		Number of teachers at assignment		Number of schools in analysis sample	
		Treatment group	Control group	Treatment group	Control group	Treatment group	Control group	Treatment group	Control group
Observation domain-specific professional development recommended by principal	Teacher survey	0.031 (0.174)	0.055 (0.227)	386	403	1,173	1,186	145	139
General professional development recommended by principal	Teacher survey	0.150 (0.357)	0.200 (0.400)	393	409	1,315	1,297	146	140
Take-up of any professional development by teacher	Teacher survey	0.821 (0.383)	0.851 (0.357)	393	409	1,348	1,329	146	140
Teacher follows principal's professional development recommendation	Teacher survey	0.899 (0.302)	0.862 (0.346)	386	398	343	482	146	138

Note: Baseline and control group means presented are baseline summary statistics for the sample included in the analysis. Assignment is assignment of schools to treatment and control groups.

Source: Authors' calculations based on survey data collected for this study.

Table D7. Baseline summary statistics for teacher instructional practice, 2014/15

Teacher instructional practice (NMTEACH domains)	Data source	Baseline mean (standard deviation)		Number of teachers in analysis sample		Number of teachers at assignment		Number of schools in analysis sample	
		Treatment group	Control group	Treatment group	Control group	Treatment group	Control group	Treatment group	Control group
Planning and preparation domain, principal rating (1–5 scale)	Administrative data	3.18 (1.15)	3.24 (1.12)	3,390	3,493	4,482	4,548	165	161
Creating an environment for learning domain, principal rating (1–5 scale)	Administrative data	3.18 (1.11)	3.19 (1.08)	3,541	3,603	4,551	4,556	170	162
Teaching for learning domain, principal rating (1–5 scale)	Administrative data	3.15 (1.12)	3.15 (1.09)	3,541	3,603	4,551	4,556	170	162
Professionalism domain, principal rating (1–5 scale)	Administrative data	3.26 (1.18)	3.26 (1.14)	3,360	3,492	4,511	4,548	166	161
Creating an environment for learning domain, teacher self-rating (1–5 scale)	Teacher survey	3.47 (1.10)	3.46 (1.17)	420	440	1,373	1,365	149	144
Teaching for learning domain, teacher self-rating (1–5 scale)	Teacher survey	3.37 (1.08)	3.30 (1.14)	418	438	1,373	1,365	149	144

Note: Baseline and control group means presented are baseline summary statistics for the sample included in the analysis. Assignment is assignment of schools to treatment and control groups.

Source: Authors' calculations based on administrative data from the New Mexico Public Education Department.

Table D8. Baseline summary statistics for student Partnership for Assessment of Readiness for College and Careers assessment scores, 2014/15

School level	Subject	Data source	Baseline mean (standard deviation)		Number of students in analysis sample		Number of students at assignment		Number of schools in analysis sample	
			Treatment group	Control group	Treatment group	Control group	Treatment group	Control group	Treatment group	Control group
Elementary	Math	Administrative data	0.125 (1.04)	0.160 (1.04)	9,255	9,898	15,197	16,213	97	102
	English language arts	Administrative data	-0.046 (1.01)	0.001 (1.02)	9,042	9,709	14,973	16,024	97	102
Middle	Math	Administrative data	0.084 (0.95)	0.076 (0.95)	14,162	10,582	15,663	11,752	64	60
	English language arts	Administrative data	0.000 (0.89)	-0.018 (0.90)	14,208	10,501	15,671	11,684	64	60
High	Math	Administrative data	-0.230 (0.96)	-0.132 (1.04)	9,324	9,341	10,506	10,535	35	34
	English language arts	Administrative data	-0.011 (1.07)	0.105 (1.10)	9,467	9,439	10,606	10,655	35	34

Note: Baseline and control group means presented are baseline summary statistics for the sample included in the analysis. Assignment is assignment of schools to treatment and control groups.

Source: Authors' calculations based on administrative data from the New Mexico Public Education Department.

Methodology

Starting in April 2015, the study team invited 786 public school principals in New Mexico to participate in the study and complete the spring 2015 survey; 339 consented to participate.

Principals completed the spring 2015 surveys between May 1, 2015, and September 1, 2015. Completion of the baseline survey signaled consent to participate. About 80 percent of consenting principals had responded by the end of May 2015.

When a principal completed the spring 2015 survey (defined as responding to the survey through at least the items that compose outcome measures in the analysis), the study team emailed on the same date a survey to up to 10 randomly selected teachers in that principal's school. Teachers completed the spring 2015 teacher surveys between May 1, 2015, and September 18, 2015. About 80 percent of consenting teachers had responded by August 31, 2015.

Once the spring 2015 principal survey was closed and the final set of schools established in fall 2015, the study team emailed to all treatment group principals on the same date the checklist guide contained in appendix B. Also on that date the study team emailed to control group principals the two-page guide contained in appendix C. In fall 2015 the study team sent to teachers who had consented in spring or summer 2015 to participate in the study and who were working in treatment group schools the teacher checklist guide, which was the same checklist the principals received but with a teacher-oriented introduction (see appendix B). The study team sent teachers who consented to participate and who were working in control group schools a short reminder of their rights to classroom observations (see appendix C).

Participating teachers received their materials in two waves. Most received their checklist or control materials from the study team electronically on the same date in fall 2015 that their principal did. But the balance of participating teachers got their materials two weeks later when the spring 2015 teacher survey was closed. The delay gave teachers whose principals had recently consented to the study more time to complete the teacher survey before receiving the checklist guide.

The study team emailed all participating principals and teachers a request to complete the spring 2016 survey on the same date in April 2016 along with up to seven email reminders (sent every second week) to nonrespondents. The survey remained open for approximately three months and was closed on July 26, 2016.

Development of outcome measures for research question 1

The study team constructed nine indexes from principal and teacher survey data to summarize principal and teacher perceptions of the quality of post-observation conferences—four of the indexes used principal survey data and five used teacher survey data. These indexes were derived through exploratory factor analysis of the 2014/15 data using multiple survey items written by the study team to measure the intended impacts of the modified Carnegie Foundation Feedback Checklist. The precise number of items retained in each scale was determined using a principal components method to examine the factor loading for each item (table D9). As an initial step, factors for which the minimum eigenvalue was greater

Table D9. Principal and teacher indexes on the content, structure, and utility of post-observation feedback conferences, 2014/15 and 2015/16

Indexes	Coefficient alpha
Principal indexes	
Supportive conference	Spring 2015: .83 Spring 2016: .86
<ul style="list-style-type: none"> • Ended the conference on a positive note • High level of collaboration in feedback conferences • Reverse coded: high level of conflict in feedback conferences • Large majority of teachers seemed to trust and accept feedback • Felt positively about feedback to teachers in conferences • Enjoyed most of the post-observation feedback conferences • Feedback session, separate from professional development, helped teacher improve instruction 	
Specific and actionable feedback conference	Spring 2015: .79 Spring 2016: .80
<ul style="list-style-type: none"> • Identified at least one positive practice that the teacher did well • Identified at least one challenge facing the teacher • Provided all teachers with a written or online summary of observation with comments • Provided specific feedback to teachers about their performance • Provided actionable feedback to teachers about their performance 	
Data-driven conference	Spring 2015: .75 Spring 2016: .79
<ul style="list-style-type: none"> • Identified at least one challenge facing the teacher • Used rubric scores to praise or critique instructional practices • Used rubric scores to recommend professional development 	
Well-prepared, collaborative conference	Spring 2015: .70 Spring 2016: .76
<ul style="list-style-type: none"> • Teacher brought documents to the conference (for example, lesson plan or professional development plan) • School leader brought documents to the conference (teacher's report card or professional development plan) • Mutually developed next steps for instruction • High level of collaboration in feedback conferences • Feedback session, separate from professional development, helped teacher improve instruction 	
Teacher indexes	
Best practices conference	Spring 2015: .92 Spring 2016: .91
<ul style="list-style-type: none"> • Teacher brought documents to the conference (for example, lesson plan or professional development plan) • School leader brought documents to the conference (teacher's report card or professional development plan) • School leader identified at least one positive practice • School leader identified at least one challenge • School leader used rubric scores to praise or critique • School leader ended conference on a positive note • Each conference followed a predictable format • Walked away with a clear understanding of school leader's feedback • School leader listened to teacher during conference • Reverse coded: high level of conflict in feedback conference • Provided with a written or online summary of observation with comments 	
Data-driven conference	Spring 2015: .93 Spring 2016: .92
<ul style="list-style-type: none"> • Teacher brought documents to the conference (for example, lesson plan or professional development plan) • School leader brought documents to the conference (teacher's report card or professional development plan) • School leader identified at least one challenge • School leader used rubric scores to praise or critique instructional practices • School leader used rubric scores to recommend professional development • Mutually developed next steps for instruction • Received actionable feedback about performance • Committed to specific next steps to improve instruction • Obtained tailored recommendations for professional development • Feedback session, separate from professional development, helped teacher improve instruction 	

(continued)

Table D9. Principal and teacher indexes on the content, structure, and utility of post-observation feedback conferences, 2014/15 and 2015/16 (continued)

Indexes	Coefficient alpha
Specific and actionable feedback conference <ul style="list-style-type: none"> • Received specific feedback about performance • Received actionable feedback about performance • Trusted and accepted feedback • Felt positive about feedback from conference • Enjoyed most of the post-observation feedback conference • Feedback session, separate from professional development, helped teacher improve instruction 	Spring 2015: .95 Spring 2016: .95
Principal-dominated conference <ul style="list-style-type: none"> • Observations done to teacher and not for teacher • School leader speaks for most of the time • High level of conflict in feedback conference 	Spring 2015: .60 Spring 2016: .61
Well-rounded conference <ul style="list-style-type: none"> • Teacher brought documents to the conference (for example, lesson plan or professional development plan) • School leader brought documents to the conference (teacher's report card or professional development plan) • School leader identified at least one positive practice • School leader identified at least one challenge • School leader used rubric scores to praise or critique • School leader used rubric scores to recommend professional development • School leader ended conference on a positive note • Each conference followed a predictable format • Mutually developed next steps for instruction • Walked away with a clear understanding of school leader's feedback • School leader listened to teacher during conference • Received specific feedback about performance • Received actionable feedback about performance • Committed to specific next steps to improve instruction • Obtained tailored recommendations for professional development • Trusted and accepted feedback • Felt positive about feedback from conference • Enjoyed most of the post-observation feedback conference • Provided with a written or online summary of observation with comments • Feedback session, separate from professional development, helped teacher improve instruction 	Spring 2015: .97 Spring 2016: .96

Source: Authors' construction and calculations based on survey data collected for this study.

than or equal to 1 were retained. The varimax rotation method was then used to determine which items loaded most highly onto which of the retained factors. Operationally, the respondent-level sums of item-level responses from these survey items were averaged to generate the final indexes. The constructed indexes yielded a coefficient alpha ranging from .60 to .97 (.70 or greater is generally considered an acceptable level of internal consistency within a given factor, with lower values indicating a potential lack of adequate reliability).¹³

Analytic approach and statistical adjustments

To analyze the impact of the detailed checklist on principal, teacher, and student outcomes, the study team compared differences in outcomes between principals randomly assigned to treatment groups and those assigned to control groups, between teachers in treatment group schools and those in control group schools, and between students in treatment group schools and those in control group schools. The data for the evaluation are hierarchical, with students and teachers nested within schools (or principals) that are nested within districts. Because units within a group are not statistically independent, hierarchical linear modeling (HLM) was used to account for the statistical dependence of the error terms. The

study estimated the impact of receiving the checklist, known as the intent-to-treat effect in econometric terminology, in a two-level model for principals and a three-level model for teachers and students and used HLM for continuous principal, teacher, and student outcomes and a probit model for binary principal and teacher outcomes.

For research question 1 the intent-to-treat effect of the guide on continuous principal outcomes that are nested within districts was estimated with the following two-level hierarchical model:

$$\text{Level 1 (Principals):} \quad Y_{ij} = \beta_{0j} + \beta_{1j} \text{Treat}_i + \sum_{p=2}^P \beta_{pj} X_{p,ij} + \varepsilon_{ij}, \quad (\text{D1a})$$

$$\text{Level 2 (Districts):} \quad \beta_{0j} = \gamma_{00} + \sum_{q=1}^Q \gamma_{q01} W_{qj} + \omega_{0j}, \quad (\text{D1b})$$

$$\beta_{pj} = \gamma_{p0}, \quad p = 1, \dots, P \quad (\text{D1c})$$

where Y_{ij} is the continuous outcome measure for principal i in district j and Treat_i is an indicator variable taking a value of 1 for treatment group schools and 0 for control group schools. The $X_{p,ij}$ term represents principal- and school-level covariates ($p = 2, \dots, P$), while the W_{qj} term represents district characteristics ($q = 1, \dots, Q$). The covariates included for each research question are listed in table D1. The error term in equation D1a, ε_{ij} , is assumed to be distributed $N(0, \sigma^2)$, and the error term in equation D1b, ω_{0j} , is assumed to be distributed $N(0, \tau^2)$. The intent-to-treat effect is given by β_{1j} and is the difference in the outcome measure Y_{ij} between principals who were randomly assigned to the treatment group and principals who were assigned to the control group, after any differences in the covariates were controlled for.

For research questions 1 and 3 about the quality of the feedback conference and about subsequent instructional practices, the intent-to-treat effect of the guide on teacher outcomes was modeled with a three-level hierarchical model. Because schools are randomly assigned the guide, the treatment effect is included in the level 2 model. The three-level model for continuous teacher outcomes is given by:

$$\text{Level 1 (Teachers):} \quad Y_{ijk} = \pi_{0jk} + \sum_{p=1}^P \pi_{pjk} (a_{p,ijk}) + \varepsilon_{ijk}, \quad (\text{D2a})$$

$$\text{Level 2 (Schools):} \quad \pi_{0jk} = \beta_{00k} + \beta_{01k} \text{Treat}_j + \sum_{q=2}^{Q_p} \beta_{0qk} X_{q,jk} + \omega_{00k}, \quad (\text{D2b})$$

$$\pi_{pjk} = \beta_{p0k}, \quad p = 1, \dots, P \quad (\text{D2c})$$

$$\text{Level 3 (District):} \quad \beta_{00k} = \gamma_{000} + \sum_{s=1}^{S_{pq}} \gamma_{00s} W_{sk} + u_{00k}, \quad (\text{D2d})$$

$$\beta_{p,qk} = \gamma_{p,q0}, \quad p = 0, \dots, P; q = 1, \dots, Q_p \quad (\text{D2e})$$

where Y_{ijk} is the continuous outcome measure for teacher i in school j and district k , Treat_j is an indicator variable taking a value of 1 for treatment group schools and 0 for control group schools, $X_{q,jk}$ represents principal- and school-level covariates, and W_{sk} represents district characteristics. The $a_{p,ijk}$ term represents teacher characteristics (see table D1), that influence the outcome of interest. The error term in equation D2a, ε_{ijk} , is assumed to be distributed $N(0, \sigma^2)$; the error term in equation D2b, ω_{00k} , is assumed to be distributed $N(0, \tau^2)$; and the error term in equation D2d, u_{00k} , is assumed to be distributed $N(0, \nu^2)$.

Teacher dichotomous outcomes for research question 2 about professional development were modeled through a probit function with controls for district, school, principal, and teacher covariates.

For research question 4 about student achievement, student assessment outcomes were modeled with a three-level HLM similar to the model in the teacher-level analyses. Ideally, student-level models would incorporate four levels (students at level 1, classrooms at level 2, schools at level 3, and district at level 4). However, NM PED was unable to provide complete classroom linkages to accompany the student assessment data. Thus, student-level models exclude the classroom level and do not account for the cluster structure of classrooms within schools. Instead, student-level analyses use standard generalized estimating equation techniques (Liang & Zeger, 1986) to capture this structure.

Because schools are randomly assigned to treatment and control groups, the treatment effect is again included in level 2 (that is, the school level) of the model. The three-level model estimating the intent-to-treat effect of the feedback conference checklist relative to the control guide on continuous student achievement outcomes is given by:

$$\text{Level 1 (Students):} \quad Y_{ijkt} = \pi_{0jk} + \pi_{pj1} Y_{ijkt(t-1)} + \sum_{p=2}^P \pi_{pjk} a_{pjk} + \varepsilon_{ijk}, \quad (\text{D3a})$$

$$\text{Level 2 (Schools):} \quad \pi_{0jk} = \beta_{00k} + \beta_{01k} \text{Treat}_j + \sum_{q=2}^{Q_p} \beta_{0qk} X_{qjk} + \omega_{00k}, \quad (\text{D3b})$$

$$\pi_{pjk} = \beta_{p0k}, \quad p = 1, \dots, P \quad (\text{D3c})$$

$$\text{Level 3 (District):} \quad \beta_{00k} = \gamma_{000} + \sum_{s=1}^{S_{pq}} \gamma_{00s} W_{sk} + u_{00k}, \quad (\text{D3d})$$

$$\beta_{pqk} = \gamma_{pq0}, \quad p = 0, \dots, P; q = 1, \dots, Q_p \quad (\text{D3e})$$

where Y_{ijkt} is the student assessment for student i in school j and district k in the outcome year $t = 2016$, and $Y_{ijkt(t-1)}$ is the student's prior year score on the same subject assessment. Similar to the variable in the teacher model, a_{pjk} represents student variables that influence the outcome of interest, Treat_j is an indicator variable taking a value of 1 for treatment group schools and 0 for control group schools, X_{qjk} represents principal- and school-level covariates, and W_{sk} represents district characteristics. The error term in equation D3a, ε_{ijk} , is assumed to be distributed $N(0, \sigma^2)$; the error term in equation D3b, ω_{00k} , is assumed to be distributed $N(0, \tau^2)$; and the error term in equation D3d, u_{00k} , is assumed to be distributed $N(0, \nu^2)$. Similarly, covariates in the level 2 model are centered at the district mean, so school-level parameters were estimated by using within-district variation. Standard errors were again clustered at the school level and calculated with the Huber-White procedure (Greene, 2003).

The analyses for research question 5 compare the responses between the treatment and control groups to principal and teacher survey questions about whether the respondent had seen the feedback conference checklist, whether the respondent had used the checklist, and the number of teachers or conferences for which the checklist was used. For the implementation analyses, box plots were created to summarize the responses to principal and teacher survey questions eliciting opinions of the respondent about the checklist, such as ease of use and time burden.

Sensitivity analyses

The study team conducted a number of sensitivity analyses to test the extent to which estimates were driven by model assumptions. First, sensitivity analyses were conducted to examine whether estimating the models by using linear regression techniques, as opposed to HLM, changed the coefficient estimates for the treatment effect. In this analysis, the coefficients were estimated using ordinary least squares, but the study team accounted for the hierarchical structure of the data when estimating the standard errors by clustering the error terms at the district level. A second set of sensitivity analyses ensured that the results here are not sensitive to which control variables are included. In all cases, the results of the sensitivity analysis were broadly consistent with the results presented in the report. Using the linear regression techniques gave similar estimates to the HLM and probit models. In fact, in none of the cases where the study reported a statistically significant result did the significance level change when running linear regressions instead of HLM or probit models.

Across all the estimated effects, the only change that occurs when running linear regressions instead of HLM or probit models is that the positive but statistically insignificant effects of the treatment on the creating an environment for learning and teaching for learning domains of teacher practice, as reported in table 5 in the main text, become larger and statistically significant (at the 5 percent level). Likewise, once the spring 2015 survey results were controlled for, it did not matter which other covariates were included. In addition, because the randomization succeeded reasonably in ensuring that the treatment and control groups had similar responses on the spring 2015 survey, the findings reported here are similar to those that compared the mean of the treatment group schools to the mean of the control group schools. The only change in statistical significance was that the effect on whether teachers reported that their conference was dominated by the principal was no longer significant at the 1 percent level, only at the 5 percent level; this was due mostly to a larger standard error around the estimate.

Exploratory subgroup analyses

Although this study does not have a sufficient sample size to randomize the feedback conference checklist by subgroups of interest, the study team conducted exploratory analyses on subgroups to better understand the heterogeneity of the impacts of providing a detailed checklist for feedback conversations. For principals the differential impact of the feedback conference checklist on the content and structure of the feedback conversation, the quality of feedback provided, and the alignment of professional development recommendations with needs identified in the formal observation were examined by frequency of use of the feedback protocol; school accountability grades; school characteristics (such as percentage of American Indian/Alaska Native students and percentage of English learner students); training on the NMTEACH Observation Rubric; and qualifications of the principals, including years of experience and certification. To estimate the subgroup effects for principal continuous outcomes, equation D1a was modified to include an interaction term between treatment status and the subgroup of interest.

Similarly, the study examined teacher subgroups to test differences in the impact of the feedback conference checklist on the quality of feedback received by frequency of use of the feedback protocol; teacher tenure status (self-reported); whether the teacher taught

core/tested versus noncore/nontested subjects; and qualifications, including years of experience and certification. Separate subgroup analyses also examined whether teachers were more likely to attend professional development courses and find these courses useful by teaching arrangement, teacher experience, teacher characteristics, observation frequency, professional development opportunities offered by the school district, and school demographic composition. Finally, the study team examined differences in the impact of the feedback conference checklist on teacher practice as measured by the NMTEACH Observation Rubric by teaching arrangement, teacher experience, teacher characteristics, school demographic composition, and use of the checklist.

All subgroup analyses were conducted separately by subgroup to allow both the coefficient on the treatment and the coefficients on all of the covariates to vary by subgroup.

Of the 242 interaction effects estimated, only 20 (8 percent) were statistically significant, which is roughly what would be expected by chance. After a Benjamini-Hochberg correction to account for the many hypotheses being tested, interaction effects for four remained statistically significant. The first two were that treatment group principals who reported that their professional development was not useful saw a larger increase in the way they rated teachers on the planning and preparation and teaching for learning NMTEACH domains than did principals who reported that their professional development was useful. The third effect was that receiving the checklist increased the probability that teachers would take professional development in schools with a smaller proportion of English learner students than in schools with a larger proportion of English learner students. Fourth, receiving the checklist had a larger effect on the probability that principals would recommend specific actions in schools with a larger proportion of American Indian/Alaska Native students than in schools with a smaller proportion.

Because there were no clear trends in the subgroup analyses and the four results that remained statistically significant after the Benjamini-Hochberg correction did not indicate a meaningful pattern, the full set of results are not included in this report.

Treatment of missing data

To prevent loss in the sample because of missing covariates, the missing indicator method was used in the impact analysis (White & Thompson, 2005). Indicators were created for each covariate that included missing data such that the indicator equaled 1 if the covariate was missing for that observation and 0 otherwise, and missing values in the covariates themselves were recoded to a constant. Both the recoded covariate and the missing indicator were included in the regression model at the level of the initial covariate. Observations with missing data on the outcome of interest were not included in the analysis.

Treatment of crossovers

The intent-to-treat analyses presented in the main text used the original random assignment group irrespective of whether the principal or teacher used the feedback conference checklist. However, because use of the checklist was not mandatory, multiple crossovers occurred. To examine the sensitivity of the results to crossovers, the study team conducted a treatment-on-the-treated analysis by using instrumental variable regressions in which the treatment group assignment was used as an instrument for whether the principal or

teacher used the feedback conference checklist. Results for these analyses are presented in appendix E.

Specifically, for teacher-level outcomes, the following pair of equations was estimated using the two-stage least squares methodology:

$$D_{ijkt} = a_{ijt} \tilde{\pi} + X_{jt} \tilde{\beta} + W_{kt} \tilde{\gamma} + \phi Treat_j + \mu_{ijkt} \quad (D4)$$

$$Y_{ijkt} = \delta D_{ijkt} + \alpha_{ijt} \pi + X_{jt} \beta + W_{kt} \gamma + \epsilon_{ijkt} \quad (D5)$$

where D_{ijkt} is an indicator for whether teacher i in school j in district k at time t reported using the checklist, Y_{ijkt} is the continuous outcome measure for the teacher, a_{ijt} represents teacher characteristics, X_{jt} represents school/principal characteristics, W_{kt} represents district characteristics, and the random assignment to the treatment group ($Treat_j$) is used as an instrumental variable for having used the checklist in the second stage. The coefficient of interest is ϕ , which is the regression-adjusted estimate of the treatment-on-the-treated effect, after teacher, school, principal, and district characteristics to improve the precision of the estimates are controlled for. All of the controls used in the instrumental variable estimates are identical to those used in the intent-to-treat estimates. Principal-level outcomes are estimated with a similar methodology, with teacher characteristics excluded from those equations.

Appendix E. Treatment-on-the-treated analyses

This appendix contains the results from the treatment-on-the-treated analyses of the impact of the feedback conference checklist on principal-, teacher-, and school-level outcomes. In general, the overall findings from these analyses are similar to those of the intent-to-treat analyses, and outcomes that saw a statistically significant effect in the intent-to-treat analysis also saw statistically significant outcomes in this analysis. However, there are some important differences. Notably, the relatively low take-up rates meant that the magnitudes of the estimated effect here are larger than those in the main text. It also means that the standard errors are much larger for these estimates than for the intent-to-treat estimates, which has important implications. Therefore, although the study team can be reasonably confident about what impact providing the checklist has on the school, there is less certainty about what impact using the checklist has on the school.

Table E1. Treatment-on-the-treated estimates on principal-reported conference quality, 2015/16

Principal quality of conference outcome measure	Used checklist mean (standard error)	Did not use checklist mean (standard error)	Estimated impact (standard error ^a)	Effect size ^b	Sample size
Supportive conference (0–100 scale)	79.65 (10.36)	78.36 (12.28)	20.904 (2.418)	–0.074	173
Specific feedback conference (0–100 scale)	82.63 (11.49)	81.62 (14.21)	–2.904 (3.913)	–0.204	175
Data-driven conference (0–100 scale)	79.56 (17.57)	75.28 (19.18)	–1.809 (5.348)	–0.094	177
Well-prepared, collaborative conference (0–100 scale)	72.41 (12.22)	65.52 (15.03)	2.690 (3.310)	0.179	170
Conference duration (minutes)	33.86 (12.09)	30.19 (12.52)	–4.319 (2.649)	–0.345	167

Note: Although the treatment and control group means reported do not control for any differences in covariates, the difference is estimated by using an instrumental variable regression in which the treatment group assignment was used as an instrument for whether the principal used the feedback conference checklist. See appendix D for covariates included in the model and treatment of missing data. The analysis sample included only principals who completed both surveys.

a. See appendix D for a description of how standard errors were estimated.

b. The effect on a principal's conference measure divided by the standard deviation of all principals' conference measures.

Source: Authors' analysis of survey data collected for this study.

Table E2. Treatment-on-the-treated estimates on teacher-reported conference quality, 2015/16

Teacher-reported quality of conference outcome	Used checklist mean (standard error)	Did not use checklist mean (standard error)	Estimated impact (standard error ^a)	Effect size ^b	Sample size
Best practices conference (0–100 scale)	78.98 (13.89)	67.79 (21.97)	3.557 (7.626)	0.162	812
Specific and actionable feedback (0–100 scale)	78.05 (18.72)	62.60 (27.81)	5.502 (9.049)	0.198	837
Data-driven conference (0–100 scale)	68.49 (16.84)	52.39 (24.50)	1.793 (8.783)	0.073	812
Principal-dominated conference (0–100 scale)	20.85 (17.07)	27.81 (20.48)	-19.357** (7.361)	-0.945	829
Well-rounded conference (0–100 scale)	75.68 (14.70)	60.78 (24.13)	4.102 (8.040)	0.170	798
Conference duration (minutes)	38.99 (14.73)	31.11 (18.20)	5.353 (7.346)	0.294	826

** Statistically significant at $p < .01$.

Note: Although the treatment and control group means reported do not control for any differences in covariates, the difference is estimated by using an instrumental variable regression in which the treatment group assignment was used as an instrument for whether the teacher used the feedback conference checklist. See appendix D for covariates included in the model and treatment of missing data. The analysis sample included teachers who completed both surveys.

a. See appendix D for a description of how standard errors were estimated.

b. The effect on a teacher’s conference measure divided by the standard deviation of all teachers’ conference measures.

Source: Authors’ analysis of survey data collected for this study.

Table E3. Treatment-on-the-treated estimates on professional development recommendation and take-up, 2015/16

Professional development recommendation and take-up outcome	Used checklist mean (standard error)	Did not use checklist mean (standard error)	Estimated impact (standard error ^a)	Effect size ^b	Sample size
Observation domain–specific professional development recommended by principal (indicator)	0.02 (0.13)	0.04 (0.21)	-0.154* (0.005)	-0.747	789
General professional development recommended by principal (indicator)	0.15 (0.36)	0.13 (0.33)	-0.172 (0.144)	-0.516	802
Take-up of any professional development by teacher (indicator)	0.91 (0.29)	0.84 (0.37)	-0.159 (0.141)	-0.432	802
Teacher follows principals’ professional development recommendation (indicator)	0.90 (0.31)	0.92 (0.427)	0.284 (0.125)	1.036	784

* Statistically significant at $p < .05$.

Note: Although the treatment and control group means reported do not control for any differences in covariates, the difference is estimated by using an instrumental variable regression in which the treatment group assignment was used as an instrument for whether the teacher used the feedback conference checklist. See appendix D for covariates included in the model and treatment of missing data. The analysis sample included only respondents who completed both surveys.

a. See appendix D for a description of how standard errors were estimated.

b. The effect on a teacher’s professional development recommendation or take-up divided by the standard deviation of all teachers’ professional development recommendations or take-up.

Source: Authors’ analysis of survey data collected for this study.

Table E4. Treatment-on-the-treated estimates on teacher instructional practice, 2015/16

Instructional practice outcome (NMTEACH Observation Rubric domains, 1–5 scale)	Used checklist mean (standard error)	Did not use checklist mean (standard error)	Estimated impact (standard error ^a)	Effect size ^b	Sample size
Principal ratings					
Planning and preparation	3.79 (0.64)	3.72 (0.63)	0.093 (0.238)	0.147	836
Creating an environment for learning	3.78 (0.51)	3.69 (0.54)	0.159 (0.194)	0.297	864
Teaching for learning	3.69 (0.52)	3.61 (0.55)	0.267 (0.190)	0.491	864
Professionalism	3.91 (0.60)	3.83 (0.65)	0.004 (0.231)	0.007	830
Teacher self-ratings					
Creating an environment for learning	3.81 (0.55)	3.86 (0.56)	0.283 [†] (0.155)	0.503	859
Teaching for learning	3.66 (0.53)	3.75 (0.53)	0.325* (0.151)	0.608	855

* Statistically significant at $p < .05$. † Statistically significant at $p < .10$.

Note: Although the treatment and control group means reported do not control for any differences in covariates, the difference is estimated by using an instrumental variable regression in which the treatment group assignment was used as an instrument for whether the teacher used the feedback conference checklist. See appendix D for covariates included in the model and treatment of missing data. The analysis sample included only respondents who completed both surveys.

a. See appendix D for a description of how standard errors were estimated.

b. The effect on a teacher's evaluation rating divided by the standard deviation of all teachers' evaluation ratings.

Source: Authors' analysis of administrative data from the New Mexico Public Education Department and survey data collected for this study.

Table E5. Treatment-on-the-treated estimates on student achievement, 2015/16

Student achievement outcomes	Used checklist mean (standard error)	Did not use checklist mean (standard error)	Estimated impact (standard error ^a)	Effect size ^b	Sample size
Elementary school math PARCC scores	0.25 (1.05)	0.21 (1.00)	0.035 (0.076)	0.035	17,125
Elementary school English language arts PARCC scores	-0.03 (0.96)	-0.07 (0.94)	0.015 (0.058)	0.016	16,972
Middle school math PARCC scores	0.07 (1.02)	-0.06 (0.95)	0.133 (0.069)	0.133	16,025
Middle school English language arts PARCC scores	0.11 (0.95)	-0.09 (0.90)	0.151* (0.076)	0.167	15,922
High school math PARCC scores	-0.28 (0.93)	-0.11 (0.99)	-0.113 (0.092)	-0.113	11,962
High school English language arts PARCC scores	0.05 (1.10)	0.24 (1.14)	0.037 (0.068)	0.033	12,060

* Statistically significant at $p < .05$.

PARCC is Partnership for Assessment of Readiness for College and Careers.

Note: Although the treatment and control group means reported do not control for any differences in covariates, the difference is estimated by using instrumental variable regression in which the treatment group assignment was used as an instrument for whether the principal used the feedback conference checklist. See appendix D for covariates included in the model and treatment of missing data. The analysis sample included only students who have an achievement score from the previous school year.

a. See appendix D for a description of how standard errors were estimated.

b. The effect on a student's English language arts or math PARCC score divided by the standard deviation of all students' PARCC scores.

Source: Authors' analysis of administrative data from the New Mexico Public Education Department.

Notes

1. The NMTEACH Observation Rubric is based on the Framework for Teaching rubric developed by Charlotte Danielson (Danielson, 2011). The rubric contains four domains: planning and preparation, creating an environment for learning, teaching for learning, and professionalism. Each domain contains five or six elements that are scored individually on a five-point scale. Immediately following the classroom observation, the observer is supposed to enter the scores on each element of the rubric into the statewide online system, REFLECT, which produces an output for the teacher and principal to review.
2. For example, presenting information about school-level academic achievement to parents who are selecting schools can affect which school their child attends and their child's achievement scores (Hastings and Weinstein, 2008), and sending parents information about their child's missed assignments and grades via email and text messages can improve both student effort and subsequent grades (Bergman, 2017; Kraft & Rogers, 2015).
3. The invitation email included the information that principals would be randomly assigned one of two types of guidance but did not include details about the guidance.
4. Principals and teachers in the treatment group received four reminders throughout the 2015/16 school year to use the feedback checklist in their post-observation conferences. Each reminder email included a copy of the 24-item feedback checklist.
5. Because only principals in the control group received the control guide, the treatment-control comparison in this study differs modestly from a treatment-business-as-usual comparison.
6. For all analyses of the main effects in the study (that is, for research questions 1–3), the study team applied the Benjamini-Hochberg correction to correct for the potential of a false discovery of statistical significance due to testing multiple comparisons. Whenever this correction changes the statistical significance of an outcome, it is reported in an endnote.
7. After the Benjamini-Hochberg correction for testing multiple hypotheses was applied, the coefficient was no longer statistically significant.
8. To follow a principal's recommendation for professional development means to take it up when recommended or not to take it up when not recommended, whereas a teacher's report of taking up professional development means that the decision was made independent of any recommendation.
9. The survey did not collect teacher self-reported measures for the planning and preparation and professionalism domains. After the Benjamini-Hochberg correction for false discovery rate was applied to account for multiple comparisons, the coefficient was no longer statistically significant for both domains.
10. To test whether the receipt or self-reported use of the checklist had differential effects on principals and teachers based on their experience levels and their school contexts, the study team also conducted exploratory analyses to estimate intent-to-treat and treatment-on-the-treated models, where the indicator for treatment was interacted with teacher, principal, or school characteristics to examine subgroup effects. These analyses of subgroups did not yield consistent or policy-relevant patterns and almost all of the statistically significant results may have occurred by chance. Consequently, subgroup results are not reported. See appendix D for more detail.
11. In schools with 10 or fewer teachers, all teachers were sampled. In schools with more than 10 teachers, 10 were selected at random for recruitment.

12. Because the teacher and the principal surveys were fielded independently, the attrition rate for teachers is calculated for all schools, regardless of whether the principal continued to participate.
13. Although the alpha presented for the principal-dominated conference index is below .70, exploratory factor analysis revealed this factor to be unique when examining both the within-school variance and the between-school variance. Moreover, the number of factors extracted that best fit the data in terms of conceptual understanding and model fit statistics include this construct for principal-dominated conferences. Alpha is also partially driven by the number of items in the factor (for example, if inter-item correlations are held constant, adding items will always result in increased alpha). Given that the principal-dominated conference index contains only three items and the results of factor analysis fit statistics, an alpha of .60 is admissible.

References

- Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In G. Sykes & L. Darling-Hammond (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3–22). San Francisco, CA: Jossey-Bass.
- Bergman, P., 2017. *Parent-child information frictions and human capital investment: Evidence from a field experiment*. Working paper. New York, NY: Teachers College, Columbia University.
- Cawley, B. D., Keeping, L. M., & Levy, P. E. (1998). Participation in the performance appraisal process and employee reactions: A meta-analytic review of field investigations. *Journal of Applied Psychology*, 83(4), 615–633.
- Chalies, S., Ria, L., Bertone, S., Trohel, J., & Durand, M. (2004). Interactions between pre-service and cooperating teachers and knowledge construction during post-lesson interviews. *Teaching and Teacher Education*, 20(8), 765–781. <https://eric.ed.gov/?id=EJ697927>
- Correnti, R. (2007). An empirical investigation of professional development effects on literacy instruction using daily logs. *Educational Evaluation and Policy Analysis*, 29(4), 262–295. <https://eric.ed.gov/?id=EJ782078>
- Correnti, R., & Rowan, B. (2007). Opening up the black box: Literacy instruction in schools participating in three comprehensive school reform programs. *American Educational Research Journal*, 44(2), 298–338. <https://eric.ed.gov/?id=EJ782088>
- Danileson, C. (2011). *The Framework for Teaching evaluation instrument, 2011 edition*. Princeton, NJ: The Danielson Group.
- DeNisi, A. S., & Sonesh, S. (2011). The appraisal and management of performance at work. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology, vol. 2: Selecting and developing members for the organization* (pp. 255–279). Washington, DC: American Psychological Association.
- Doherty, K. M., & Jacobs, S. (2015). *State of the states 2015: Evaluating teaching, leading and learning*. Washington, DC: National Council on Teacher Quality.
- Donaldson, M. (2013). Principals' approaches to cultivating teacher effectiveness: Constraints and opportunities in hiring, assigning, evaluating and developing teachers. *Educational Administration*, 49(5), 838–882. <https://eric.ed.gov/?id=EJ1019091>
- Frase, L. E., & Streshley, W. (1994). Lack of accuracy, feedback and commitment in teacher evaluation. *Journal of Personnel Evaluation in Education*, 8(1), 47–57. <https://eric.ed.gov/?id=EJ482611>
- 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 Journal of Educational Research*, 38(4), 915–945.

- Green, E. (2010, March 2). Building a better teacher. *New York Times Magazine*, 1–9.
- Greene, W. H. (2003). *Econometric analysis*. Delhi, India: Pearson Education India.
- Hastings, J. S. & Weinstein, J. M., (2008). Information, school choice, and academic achievement: Evidence from two experiments. *Quarterly Journal of Economics*, 123, 1373–1414. <https://eric.ed.gov/?id=ED501991>
- Holland, P. E. (1989). Implicit assumptions about the supervisory conference: A review and analysis of the literature. *Journal of Curriculum and Supervision*, 4(4), 362–379.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284.
- Kraft, M. A., & Rogers, T. (2015). The underutilized potential of teacher-to-parent communication: Evidence from a field experiment. *Economics of Education Review* 47, 49–63.
- Lavecchia, A. M., Liu, H., & Oreopoulos, P. (2016). Behavioral economics of education: Progress and possibilities. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education*, vol. 5 (pp. 1–74). Amsterdam: North Holland.
- Liang, K., & Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13–22.
- Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational evaluation and policy analysis*, 15(2), 129–151. <https://eric.ed.gov/?id=EJ466295>
- Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9), 705–717. <https://eric.ed.gov/?id=EJ466295>
- London, M., & Smither, J. W. (2002). Feedback orientation, feedback culture, and the longitudinal performance management process. *Human Resource Management Review*, 12(1), 81–100.
- Marshall, K. (2009). A how-to plan for widening the gap. *Phi Delta Kappan*, 90(9), 650–655.
- Medley, D. M., & Coker, H. (1987). The accuracy of principals' judgments of teacher performance. *Journal of Educational Research*, 12(3), 257–268. <https://eric.ed.gov/?id=EJ354931>
- Myung, J., & Martinez, K. (2013). *Strategies for enhancing the impact of post-observation feedback for teachers*. Stanford, CA: Carnegie Foundation for the Advancement of Teaching. <https://eric.ed.gov/?id=ED560122>
- National Council on Teacher Quality. (2016). *State-by-state evaluation timeline briefs*. Washington, DC: Author. Retrieved March 10, 2017, from http://www.nctq.org/dmsStage/Evaluation_Timeline_Brief_Overview.

- Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Journal of Educational Research*, 44(4), 921–958. <https://eric.ed.gov/?id=EJ782062>
- Peterson, K. D. (2000). *Teacher evaluation: A comprehensive guide to new directions and practices*. Thousand Oaks, CA: Sage. <https://eric.ed.gov/?id=ED445087>
- Rathel, J., Drasgow, E., & Christle, C. C. (2008). Effects of supervisor performance feedback on increasing preservice teachers' positive communication behaviors with students with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders*, 16(2), 67–77. <https://eric.ed.gov/?id=EJ794426>
- Rockoff, J. E., Staiger, D. O., Kane, T. J., & Taylor, E. S. (2012). Information and employee evaluation: Evidence from a randomized intervention in public schools. *American Economic Review*, 102(7), 3184–3213. <https://eric.ed.gov/?id=ED511139>
- Sartain, L., Stoelinga, S. R., & Brown, E. R. (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal–teacher conferences, and district implementation*. Chicago, IL: Consortium on Chicago School Research. <https://eric.ed.gov/?id=ED527619>
- Shulman, V., Sullivan, S., & Glanz, J. (2008). The New York City school reform: consequences for supervision of instruction. *International Journal of Leadership in Education*, 11(4), 407–425. <https://eric.ed.gov/?id=EJ821825>
- Skandera, H. (2013). *NMTEACH Observation Protocol workbook*. Santa Fe, NM: New Mexico Public Education.
- Sporte, S. E., Stevens, W. D., Healey, K., Jiang, J., & Hart, H. (2013). *Teacher evaluation in practice: Implementing Chicago's REACH Students*. Chicago, IL: Consortium on Chicago School Research.
- Stodolsky, S. S. (1984). Teacher evaluation: The limits of looking. *Educational Researcher*, 13(9), 11–18. <https://eric.ed.gov/?id=EJ309391>
- Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching*, 37(2), 963–980. <https://eric.ed.gov/?id=EJ615659>
- Tang, S., & Chow, A. (2007). Communicating feedback in teaching practice supervision in a learning-oriented field experience assessment framework. *Teaching and Teacher Education*, 23(7), 1066–1085. <https://eric.ed.gov/?id=EJ770305>
- Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review*, 102(7), 3628–3651.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, Conn: Yale University Press.

- White, I. R., & Thompson, S. G. (2005). Adjusting for partially missing baseline measurements in randomized trials. *Statistics in Medicine*, 24(7), 993–1007.
- Williams, M., & Watson, A. (2004). Post-lesson debriefing: Delayed or immediate? An investigation of student teacher talk. *Journal of Education for Teaching*, 30(2), 85–96. <https://eric.ed.gov/?id=EJ680908>
- Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *American Educational Research Association*, 24(1), 173–209.
- Yariv, E. (2009). The appraisal of teachers' performance and its impact on the mutuality of principalteacher emotions. *School Leadership and Management*, 29(5), 445–461. <https://eric.ed.gov/?id=EJ864700>

The Regional Educational Laboratory Program produces 7 types of reports



Making Connections

Studies of correlational relationships



Making an Impact

Studies of cause and effect



What's Happening

Descriptions of policies, programs, implementation status, or data trends



What's Known

Summaries of previous research



Stated Briefly

Summaries of research findings for specific audiences



Applied Research Methods

Research methods for educational settings



Tools

Help for planning, gathering, analyzing, or reporting data or research