Section B

June 14, 2013

Table of Contents

Section B. Data Collection Procedures and Statistical Methods

The purpose of this research study is to examine a model of performance feedback to gain a deeper understanding of feedback characteristics that may be necessary for teachers to move from receiving feedback to improved performance in the context of teacher evaluation systems. Specifically, the study will examine the relationship among the usefulness of feedback, accuracy of feedback, credibility of the person providing feedback, learning opportunities related to the feedback, responsiveness to feedback, and teacher performance. The results of the study can be used to inform future training and guidance on providing feedback in teacher evaluation systems by considering the importance of different feedback characteristics (usefulness, accuracy, and credibility) and learning opportunities in the use of feedback. This information may be helpful for prioritizing needs both at the state and the district level for training and guidance on providing feedback in teacher evaluation systems, and also for informing the state of additional data collection needed to further understand feedback characteristics.

B1. Respondent Universe and Sampling Methods

Pilot Study Sample

The pilot test is intended to test the survey prior to full study implementation on a small sample of teachers in districts that are implementing a teacher evaluation system that has the intention of informing teacher development through feedback. The pilot sample will include teachers from four districts in Colorado that are implementing four different teacher evaluation systems. These four districts served as partner districts in the development of the Colorado Department of Education (CDE) model evaluation system during the initial development of the CDE system. These partner districts have been implementing locally-developed teacher evaluation systems for a few years, and they shared their lessons learned with CDE to inform the development of the model evaluation system. Based on conversations with CDE, it is believed that these districts are further along in the teacher evaluation process and would be likely to have valuable information about the teacher feedback included in their evaluation systems. Table 1 presents demographic information for these districts. Table 2 provides a brief description of the district evaluation systems. A random sample of 900 teachers would be drawn from these districts to represent various grade levels and subject areas. The 900 teachers is a sufficiently large enough sample to conduct the planned analysis, including confirmatory factor analysis (CFA) which would require approximately 265 to 530 participants given the proposed CFA model.

Table 1: Description of partner districts

Partner District	Locale	Number of Teachers
Brighton School District	Rural: Fringe	776
Denver Public Schools	City: Large	4960
Eagle County School Districts	Town: Remote	455
Harrison School District 2	City: Large	696

Table 2: Brief description of partner district evaluation systems

Partner District	Description of Evaluation System
Brighton School District	- Coaches and principals are trained on what good teaching looks like. - Student achievement data is also used to inform evaluation ratings. - Observations are conducted by the principal.
Denver Public Schools	- Locally developed rubric with 12 indicators of effective teaching - Student growth is 50% of performance criteria. - Observations conducted by peer observers and principal
Eagle County School Districts	- Rubric developed based on TAP system with 19 indicators of effective teaching - Student growth measure included - Observations conducted by master teachers and the principal
Harrison School District 2	- Locally-developed rubric with seven standards of effective teaching - Student growth is 50% of performance criteria. - Four observations are conducted each year to inform evaluation ratings.

Full Study Sample

Within each state, researchers will randomly sample approximately 10 to 20% of teachers to obtain a sample large enough to meet the sample size requirements described in Section B2 Table 6, given an expected 50% consent and 80% response rate. In each state, researchers first identified the districts who are participating in the pilot of the state's model evaluation system. Each state sought volunteer districts to participate in the pilot of the statewide evaluation system. These districts were identified to participate in the proposed study because within each state these districts will all be implementing the same evaluation system (the state system), and these districts will have the most experience with implementing this system because they have participated in the pilot and therefore have implemented the state system for one to three years prior to data collection in this study. Using the districts who were involved in the pilot will allow us to identify a sample within each state that is using the same system and has the most experience with implementing the system. Additionally, the districts that are involved in piloting the teacher evaluation system in each state proportionally represent locales across the state in a manner similar to that of the statewide distribution (see Tables 3, 4, and 5).

Researchers will obtain a random sample of teachers from districts participating in each state's evaluation system pilot study, which in each state continues through 2013–14. Using simple random sampling techniques we expect to get representation from various district locales/sizes, subject areas (such as language arts, math, science, social studies, non-core subjects, and ELL/SPED/intervention) and school levels (such as elementary, middle, and high school). This random sampling technique will allow us to generalize findings to the population of teachers in districts piloting the state's evaluation model. For each state, we will obtain a list of all classroom teachers in each participating district. The appropriate number of teachers will then be randomly selected using the random-number-generator feature in Microsoft Excel. Information about the pilot districts and the expected sample for each state is described below.

Colorado Sampling

There are 25 districts who are participating in the pilot of the Colorado model teacher evaluation system during the 2012–13 school year. Ten of these districts are also participating in the pilot of the new Colorado Academic Standards. This study will include the 15 districts who are participating only in the pilot of the model teacher evaluation system. The 15 districts participating in the pilot represent various locales (Table 3). In the 15 pilot districts, a pool of approximately 7,200 full-time-equivalent (FTE) teachers are available for potential participation in this study. Researchers will identify a random sample of 20% of the FTE teachers, approximately 1,500 teachers.

Table 3: Percent of pilot districts in Colorado by locale

Locale	Pilot Districts	All Districts
% City	0%	7%
% Suburb	13%	7%
% Town	20%	21%
% Rural	67%	65%

Kansas Sampling

There are 22 districts participating in the pilot of the Kansas Educator Evaluation Protocol (KEEP) during the 2012–13 school year. The 22 districts in the pilot represent various locale (Table 4). In the 22 pilot districts, a pool of approximately 8,900 FTE teachers are available for potential participation in this study. Researchers will identify a random sample of 20% of the FTE teachers, approximately 1,800 teachers.

Table 4: Percent of pilot districts in Kansas by locale

Locale	Pilot Districts	All Districts
% City	13%	2%
% Rural	50%	71%
% Suburb	5%	3%
% Town	32%	24%

Missouri Sampling

There are 97 school districts participating in the 2012–13 pilot of Missouri's Educator Evaluation System representing various locales (Table 5). In the 97 pilot districts, a pool of approximately 16,400 FTE teachers is available for potential participation in this study. Researchers will identify 10% of the FTE teachers, approximately 1600 teachers.

Table 5: Percent of pilot districts in Missouri by locale

Locale	Pilot Districts	All Districts
% City	3%	3%
% Rural	70%	74%
% Suburb	10%	7%
% Town	17%	16%

B2. Procedures for the Collection of Information

Statistical Power

Structural equation modeling requires a minimum of 100 to 150 subjects (Loehlin, 1992), or between 5 and 10 subjects per parameter estimated, depending on the distribution of the variables (Tabachnick & Fidell, 2001). We can estimate the number of parameters and degrees of freedom using the proposed models, which include the latent variables and the indicators based on the potential survey items and teacher evaluation rating data. The latent variables are the same in all of the states, and all indicators are the same except the indicators for teacher performance. Because of the difference in indicators of teacher performance, the model will be tested for each state. Table 6 identifies information about the model for each state and presents the sample size needed based on Tabachnick & Fidell's recommendations and also based on the methods for calculating power estimates provided by MacCallum, Browne, and Sugawara (1996). For a test of close fit based on the following assumptions: $\alpha = 0.05$, $\varepsilon_0 = 0.05$ and $\varepsilon_1 = 0.08$, where ε_0 is the null value of the root-mean-square error of approximation, and ε_1 is the alternative value.

Table 6: Sample sizes need to test the model in each state

State	Number of parameters in proposed model	Degrees of freedom in proposed model	Sample size (Tabachnick & Fidell recommendation)	Desired power	Sample size based on desired power
Colorado	69	366	345 to 690	90%	71
Kansas	65	313	325 to 650	90%	78
Missouri	63	288	315 to 630	90%	84

Data Collection

Prior to data collection, teachers will be asked to sign a consent form to participate in the study, which will provide them with information on the procedures for collecting data, including information about how the researchers will protect the confidential data. Researchers plan to hold on-site meetings with teachers to provide information about the study and to collect consent forms. States and districts participating in the study will be members of the research alliance. At the start of the study, researchers will identify a point of contact within each district who will provide the evaluation rating data, including a unique teacher identifier, and teacher email addresses for invitation.

Pilot Tests Data Collection Procedures

The pilot test includes two data collection instruments: a researcher-developed survey and a teacher interview protocol. Researchers will obtain a list of email addresses for teachers who agreed to participate from the districts and will also collect email addresses on the consent forms to verify addresses as necessary. The email addresses will be associated with a unique identification number and uploaded into an online survey software system such as SurveyGizmo. An online survey will be administered to teachers in the pilot sample near the end of the 2013–14 school year. The survey will take approximately 30 minutes to complete. Shortly after the survey is administered, a subset of approximately 20 teachers will also be asked to participate in an interview in which they will be asked to provide feedback on the clarity of the survey. The 20 teachers will be purposively selected to represent various response patterns on the survey (such as levels of agreement or disagreement). The in-person interview will last approximately one hour.

Full Study Data Collection Procedures

The study involves two data collection instruments: a teacher evaluation instrument already in use in the districts and a researcher-developed teacher survey.

Teacher Evaluation Instrument. Within each of the states, teachers currently receive a performance rating based on their state's teacher evaluation instrument. At the start of the study, researchers will identify a point of contact within each district who will provide the teacher evaluation rating data, including a unique teacher identifier. Prior to study implementation, researchers will work with their contact in the district to set up protocols for collecting this data to ensure that the data is in the needed format. A copy of the current draft of the teacher evaluation protocol for each state is provided in Attachment H (Colorado), Attachment I (Missouri) and Attachment J (Kansas).

Teacher Survey. Researchers will obtain a list of email addresses for teachers who agreed to participate from the district contact and will also collect email addresses on the consent forms to use to verify email addresses as necessary. The email addresses will be associated with a unique identification number and uploaded into an online survey software system such as SurveyGizmo. Online surveys will be administered to teachers who agree to participate in the study. The survey takes approximately 30 minutes to complete. In order to maximize response rates, pre-invitation and reminder emails will be sent.

Researchers will combine evaluation data from each district into one data set for each state. Teacher survey data and evaluation rating data will be combined using a relational database that links the data based on the teacher's assigned identification number.

B3. Methods to Maximize Response Rates and To Deal With Non-Response

In order to maximize the response rate, we will employ techniques identified in the tailored design survey methods (Dillman, 2000). The tailored design method emphasizes developing easily-understood, brief surveys, and using personalized and repeated communication with participants as methods for increasing response rates.

We will aim to minimize the burden on the participants by designing a survey that is easily understood and takes a minimal amount of time to complete, and by working with educators to determine a window of time within the school year to administer the online survey that would allow for greater response. We will involve members of our research alliance, including key personnel in the state and districts, early in the research process. The involvement of key personnel will help with appropriate messaging about the study, and communication about the importance of the study to states and districts.

In order to obtain an 80% or higher response rate, as suggested by the Office of Management and Budget (2006), we will focus on personalized and repeated communication. We will first meet in-person with teachers, followed by personalized email invitations, and targeted email reminders. In addition, we will offer incentives for participation. Because we are collecting teacher evaluation rating in addition to the survey data, and teacher evaluation rating data is somewhat sensitive information, it will be important for the participants to have an understanding of how this data will be used. Teachers who are identified in the sample will be invited to attend an in-person meeting in which they will receive information on the study and procedures for collecting data, including information about how the researchers will protect the confidential data. Consent forms will be collected during this meeting; therefore, teachers will provide consent to participate prior to receiving the survey email invitation. We will communicate with our district contacts to coordinate the in-person meeting, so that the invitation will come from the district and teachers will meet in a central location in the district in which the district typically convenes teachers. The district contact will send a letter about the study with the invitation to the in-person meeting. In this letter, the district contact will show their support and endorsement of the study and explain to teachers the benefits of their participation. Researchers will also send up to three targeted email reminders in order to increase the response rate (see Attachment G for the email invitation and reminders). If teachers do not respond after two email reminders, we will also send a reminder in the mail to the teacher's school address. Teachers will be provided with the option to request a paper survey rather than completing an online survey in the initial invitation. Additionally, teachers will be offered an incentive of a $30 Visa gift card upon completion of the survey as part of the study. As part of the Reading First Impact study, researchers conducted a study to examine the effects of monetary incentives on survey response rates for teachers and found that incentives significantly increased response rates (Games, Bloom, Kemple & Jacob, 2008). This incentive was determined using NCEE guidance, which suggests a $30 incentive for a 30-minute survey about interventions (NCEE, 2005). The NCEE guidance noted that collective bargaining agreements in districts may not allow teachers to complete surveys during school time, which may result in a decreased response rate without the offer of an incentive.

In order to address the potential of incorrect email addresses in the list provided by the district, we will collect email addresses on the consent forms as well. If an email is "bounced back" as undeliverable, we will verify the email address against the address provided on the consent form and make necessary changes. If the problem cannot be resolved by checking the email address on the consent form, we will contact the district to identify a correct email address and/or the reason why the email address is not working.

B4. Tests of Procedures or Methods to be Undertaken

During the 2013–14 school year, researchers will conduct a pilot test of the survey instrument. The pilot test will be used to gather data to refine the survey and assess reliability and validity. This refined survey will then be used to conduct the full study in the 2014–15 school year. Information about the sample, data collection, and analysis for the pilot test is presented below.

The online survey will be administered to teachers in the pilot sample near the end of the 2013–14 school year. Shortly after the survey is administered, a subset of approximately 20 teachers will also be asked to participate in an interview in which they will be asked to provide feedback on the clarity of the survey. The interviews will be transcribed. Teachers will be provided with a copy of the survey and asked to review it and respond to questions such as those proposed by the National Service Knowledge Network (2008):

- What problems, if any, did you have completing the survey?
- Are the directions clear?
- Are there any words or language in the instrument that teachers might not understand?
- Did you find any of the questions to be unnecessary or redundant?
- Were any questions difficult to answer?
- What did you think this question was asking? How would you phrase it in your own words?
- Did the answer choices allow you to answer as you intended?
- Is there anything you would change about the instrument?

Analysis

Researchers will examine the clarity and psychometric properties of the survey. Data from the interviews will be analyzed to help determine if the respondents interpreted the directions and questions as the researchers intended. Interview transcripts will be coded using MAXQDA, a qualitative analysis program, to identify patterns in responses to the interview questions. Researchers will modify the survey based on common responses about the clarity of the survey.

Researchers will conduct item analysis to identify problematic items and to examine internal consistency of the survey. Cronbach's alpha coefficients will be computed for each scale and correlation analysis will be used to determine the extent to which the items correlate with the scales and the effect on Cronbach's alpha if the item were removed. Items will be deleted if removal of the items would increase Cronbach's alpha and increase the average inter-item correlation. Following the criterion proposed by Nunnally & Bernstein (1994), alpha coefficients of 0.70 or higher will be considered as reliable.

Following the item analysis, researchers will analyze the proposed measurement model in Figure 3, Section A, using confirmatory factor analysis (CFA). The factors in this measurement model will be assumed to covary, with no hypothesis that any one factor causes the other. The methods for CFA proposed by Kline (2005) will be followed. First, researchers will test for a single factor. This step allows researchers to examine the feasibility of testing a more complex model, and to determine if the fit of the single-factor model is comparable. Selected fit indexes will be examined to determine if the results indicate a poor fit, which would substantiate the need for a more complex model. Next, the five-factor

model will be examined. Researchers will examine the model fit indices, the indicator loadings, and the factor correlations to determine not only if the model fits the data, but also if the indicators have significant loadings on the associated factors, and if the estimated correlations between the factors are not "excessively high (e.g., > .85)"(Kline, 2005; p. 73). If the model does not fit the data, the model will be respecified based on an examination of modification indices, correlation residuals, and practical and theoretical considerations. To determine if the fit of the five-factor significantly reduces model fit, Chi-square difference tests will be performed. These results could inform changes to the final study, such as determining that certain indicators or survey items are not good indicators of the factors, or determining that there are more or fewer factors than originally conceived.

B5. Individuals Consulted on Statistical Aspects of the Design

The following individuals were consulted on the statistical, data collection, and analytic aspects of this study:

Name	Title	Organization	Contact Information
Dr. Steve Meyer	Senior Researcher	RMC Research Corporation	meyer@rmcdenver.com 303-305-4261
Dr. Fatih Unlu	Senior Scientist	Abt Associates	Fatih_Unlu@abtassoc.com 617-520-2528

The following individuals will be involved in the study implementation:

Name	Role	Organization	Contact Information
Dr. Trudy Cherasaro	Principal Investigator	Marzano Research Laboratory	trudy.cherasaro@marzanoresearch.com 303-766-9199
Dr. Helen Apthorp	Co-Principal Investigator	Marzano Research Laboratory	helen.apthorp@marzanoresearch.com 303-766-9199
Dr. Kerry Englert	Technical Working Group Member	Seneca Consulting	kenglert@senecaconsulting.org

References

Dillman, D. A., (2000). *Mail and internet surveys: The tailored design method* (2nd ed.). New York: Wiley.

Gamse, B.C., Bloom, H.S., Kemple, J.J., Jacob, R.T., (2008). *Reading First impact study: Interim report* (NCEE 2008-4016). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Kline, R.B. (2005). *Principles and practices of structural equation modeling* (2nd ed.). New York: Guilford.

Loehlin, J. C. (1992). *Latent variable models; An introduction to factor, path, and structural analysis.* Mahwah, NJ: Erlbaum.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130–149.

National Center for Education Evaluation (2005). *Guidelines for incentives for NCEE impact evaluations*. Washington, DC: Author.

National Service Knowledge Network (2008). *Planning and conducting a pilot test.* Scotts Valley, CA: Author. Retrieved July 18, 2012 from http://www.nationalserviceresources.org/practices/19498.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Office of Management and Budget (2006). Guidance on agency survey and statistical information collections. Washington, DC: Author. Retrieved July 18, 2012 from: http://www.whitehouse.gov/sites/default/files/omb/inforeg/pmc_survey_guidance_2006.pdf.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn and Bacon.