Contract No: ED-IES-12-C-0007

Supporting Justification for OMB Clearance of a Study of Feedback in Teacher Evaluation Systems

Section A

June 14, 2013

Submitted to:
Institutes of Education Sciences
U.S. Department of Education
555 New Jersey Ave., NW
Washington, DC 20208
(202) 219-1385

Project Officer:
Ok-Choon Park

Submitted by:
Marzano Research Laboratory
9000 E. Nichols Ave., Ste. 210
Centennial, CO 80112
(303) 766-9199

Project Director:
Trudy L. Cherasaro

Table of Contents

Section A. Justification

The U.S. Department of Education (ED) requests OMB clearance for data collection related to the Regional Educational Laboratory (REL) program. ED, in consultation with Marzano Research Laboratory under contract ED-IES-12-C-0007, has planned a study of feedback to teachers in teacher evaluation systems in three of the states served by REL Central. OMB approval is being requested for data collection including a pilot survey of teachers, interviews with teachers about the pilot survey, a survey of teachers (revised based on the pilot), and collection of extant teacher evaluation ratings.

A1. Circumstances Necessitating Collection of Information

This data collection is authorized by the Educational Sciences Reform Act (ESRA) of 2002 (see Attachment A). Part D, Section 174(f)(2) of ESRA states that as part of their central mission and primary function, each regional educational laboratory "shall support applied research by … developing and widely disseminating, including through Internet-based means, scientifically valid research, information, reports, and publications that are usable for improving academic achievement, closing achievement gaps, and encouraging and sustaining school improvement, to— schools, districts, institutions of higher education, educators (including early childhood educators and librarians), parents, policymakers, and other constituencies, as appropriate, within the region in which the regional educational laboratory is located."

Statement of Need

The importance of teacher effectiveness has been well-supported by studies demonstrating that teachers vary in their ability to produce student achievement gains; all else being equal, students taught by some teachers experience greater achievement gains than those taught by other teachers (Konstantopoulos & Chung, 2010; Aaronson, Barrow, & Sander, 2007; Nye, Konstantopoulos, & Hedges, 2004; Wright, Horn, & Sanders, 1997). These findings have encouraged wide interest in identifying the most effective programs and practices to enhance teacher effectiveness.

All seven states in the Central Region have either developed performance-based teacher evaluation systems (Colorado, Kansas, and Missouri), are in the process of developing or adopting a model system (Nebraska and South Dakota), or are exploring the development or adoption of a model system (North Dakota and Wyoming). Because states in the Central Region are local control states, districts can choose to adopt the state system or develop or adopt a different model that is aligned with the state guidelines for educator evaluation. Colorado, Kansas, and Missouri are in the process of piloting their state systems. Table 1 provides a timeline of the development of the Colorado, Kansas and Missouri teacher evaluation systems. States intend to use these models to serve both measurement and development

purposes. In order for evaluation systems to improve teaching effectiveness, it is recommended that the first priority for such system design be promotion of teacher development (Papay, 2012). As states and districts prepare to develop and implement teacher evaluation systems, they are exploring ways to use evaluation findings to provide individualized feedback that will facilitate improved teaching and learning practices, which will lead to increased student performance (Kane & Staiger, 2012). Because such new evaluation systems that focus on a system of observation and feedback have not been fully established and tested, research on feedback in teacher evaluation systems is limited. The majority of research on performance feedback is conducted in industries outside of the education profession. The current study will address this gap in the current research by examining theories of performance feedback that have limited research in the education field.

Table 1. Development Timeline of Teacher Evaluation Systems in Colorado, Kansas, and Missouri

	Colorado	Kansas	Missouri
Timeline	**2012–2013:** Pilot **2013–2014:** Validation study **2014–2015:** Full implementation	**2011–2012:** Pilot **2012–2013:** Pilot **2013–2014:** Validity study **2014–2015:** Full implementation	**2011–2012:** Field test **2012–2013/2013–2014:** Pilot **2014–2015:** Full implementation
State Requirements	State developed model system for voluntary adoption by districts.	State developed model system for voluntary adoption by districts.	State developed model system for voluntary adoption by districts.
Components	Fifty percent of the evaluation is based on evidence of professional practice. Fifty percent of the evaluation is based on student growth measures.	Evidence for the four standards will be the basis of teacher evaluation ratings. Student growth will significantly inform the evaluation.	Evaluation scores are determined based on evidence collected for each of the standards. Student growth is a significant factor in the scores.

	Colorado	Kansas	Missouri
Teacher Standards	**State-developed Teacher Quality Standards** 1. Know content 2. Establish environment 3. Facilitate learning 4. Reflect on practice 5. Demonstrate leadership 6. Student growth	**InTASC standards (10 standards grouped into 4 categories)** 1. The learner and learning 2. Content knowledge 3. Instructional practice 4. Professional responsibility	**State-developed standards aligned with InTASC** 1. Content knowledge 2. Student learning, growth, and development 3. Curriculum implementation 4. Critical thinking 5. Positive classroom environment 6. Effective communication 7. Student assessment and data analysis 8. Professional practice 9. Professional collaboration

Educators throughout the Central Region have expressed strong interest in this area. For example, district superintendents in Colorado want to know what types of feedback are best, what the right evidence is to gather about teacher performance, and how to incorporate feedback into educator evaluation systems (Central States Education Laboratory, 2011; Colorado Association of School Executives, n. d.). REL Central's Governing Board members also identified the need to develop teacher evaluation systems that both hold teachers accountable and improve professional practice as one of their top educational challenges (REL Central Governing Board Meeting Minutes, March 19–20, 2012). Conversations with each state liaison have indicated that states want to ensure that their newly-developed educator evaluation systems are supportive in nature, and that they specifically provide for constructive, timely feedback offered to all evaluated individuals.

Overview of the Study Design

The purpose of the proposed study is to describe teacher experiences with feedback as part of a teacher evaluation system and identify factors that may influence their use of feedback and their performance. Specifically, our study will examine inputs, mediating variables, and outcomes that have been identified in previous research on performance feedback. Figure 1 presents a theoretical model that depicts relationships between inputs, mediating variables, and outcomes. The proposed model of performance feedback (Figure 1) was developed based on seminal work (Ilgen, Fisher, & Taylor, 1978) and emerging research on performance feedback. The theoretical model suggests that the feedback characteristics (usefulness, accuracy, and credibility) affect teachers' responsiveness to feedback and their access to learning opportunities, which, in turn, mediate the effect of feedback characteristics on teachers'

performance. The relationship between the inputs, mediating variables, and outcomes in the theoretical model will be examined to gain a deeper understanding of feedback characteristics that may be necessary for teachers to move from receiving feedback to improved performance.

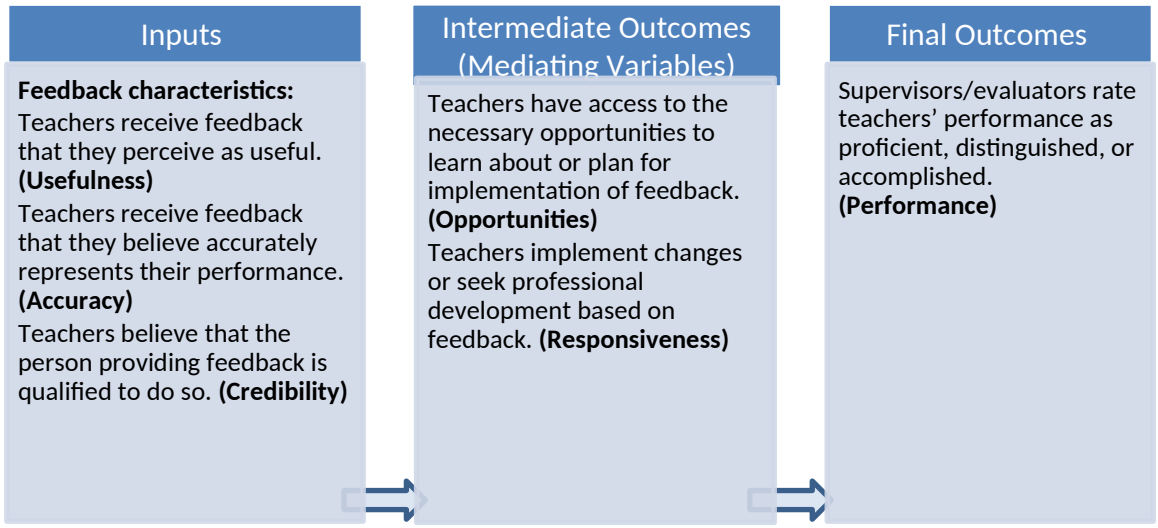

Figure 1. Theoretical model for performance feedback based on previous research

The study will address the following research questions:

1. What are teachers' perceptions of the usefulness, accuracy, and credibility of the feedback they receive as a part of their performance evaluation?
2. What is the relationship among the characteristics of performance feedback (usefulness, accuracy, and credibility)?
3. What is the relationship between feedback characteristics and teacher responsiveness?
4. What is the relationship between teacher responsiveness and teacher performance?

Overview of the Specific Data Collection Plan

This proposed study includes a pilot test of the survey instrument and implementation of the full study. 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.

Pilot Study Data Collection

The pilot serves two purposes: (1) to ensure that the survey items are understood and easy to answer; and (2) to examine the psychometric properties of the survey instrument through confirmatory factor analysis (CFA). The pilot test includes two data collection instruments: (1) a researcher-developed survey, and (2) a teacher interview protocol. The online survey consists of a set of items about feedback as part of a district's teacher evaluation system to be asked of all respondents. The survey includes questions about the usefulness of feedback, accuracy of feedback, credibility of the person providing feedback, learning opportunities related to the feedback, and responsiveness to feedback. The online survey will be administered once in the pilot study to approximately 450 teachers, in 4 school districts

for the CFA. During the pilot study, interviews will also be conducted with approximately 20 teachers to gather their feedback on the clarity of the survey questions. Teachers participating in the interview will have already completed the online survey. During the interview, teachers will be provided with a hard copy of the survey so that they can review the items and directions in order to respond to questions about clarity, such as "In question 1, what does the term 'designated evaluator' mean to you?". The interviews will be transcribed. Copies of the teacher survey and the teacher interview protocol are provided in Attachments B and C, respectively.

Full Study Data Collection

The full study involves two data collection instruments: (1) a teacher evaluation instrument already in use in the districts, and (2) a researcher-developed teacher survey. The revised survey from the pilot study will be administered once, to approximately 2,446 teachers in approximately 134 districts, in the full study. In addition to the survey, researchers will collect existing teacher evaluation ratings from the teacher evaluation instrument in use in each state to provide information about the teachers' performance as related to standards identified in the each state. Tables 8–10 in Section A.16 present the performance standards for each state.

Timeline for the Data Collection

Instrument Pilot Test	
Timeframe	**Data Collection**
Spring 2014	Administer pilot survey
Spring 2014	Conduct interviews about pilot survey
Full Study	
Timeframe	**Data Collection**
Spring 2015	Administer revised survey
Spring 2015	Collect existing teacher evaluation ratings

A2. How, by Whom, and for What Purpose Information is to Be Used

Pilot Study

The findings from the pilot study will be used by the researchers to refine the survey. The results will inform changes to the survey used in the full study, such as determining that certain indicators or survey items are not good indicators of the factors (usefulness, credibility, accuracy, opportunities, and responsiveness), or determining that there are more or fewer factors than originally conceived.

Full Study

The findings will be used by state and district leaders to prioritize needs both at the state and district level for training and guidance on providing feedback as part of teacher evaluation systems. This study will result in a report intended for district and state leaders who are responsible for selecting, developing, and implementing teacher evaluation systems and overseeing support for teachers' professional growth and effectiveness. The report will provide the findings in an accessible format

describing possible interpretations and implications of the results of the model testing. Technical information such as the correlation matrix used to test the model, the fit statistics for proposed and adjusted models, and significant standardized coefficients for each of the parameters estimated in the model will be provided in an appendix. The report will summarize findings based on results of the model testing, provide a detailed description of methods in an appendix, and discuss limitations and implications for practice and further research. The report will provide findings from each state as well as the synthesized findings from the multi-analytic structural equation modeling to inform leaders across the region. The synthesized findings will provide information about the relationship of the characteristics of feedback from three different evaluation contexts within the Central Region. These findings may be helpful for states and districts in the region who are implementing teacher evaluation systems, because they will identify which variables account for a significant amount of variance in other variables across three different evaluation systems. Researchers will also facilitate discussions about the results with state and district leaders to examine and interpret the results from this study. For example, to identify which variables account for a significant amount of variance in other variables, we will examine the coefficients of the hypothesized relationships (that the feedback characteristics [usefulness, accuracy, and credibility] affect teachers' responsiveness to feedback and their access to learning opportunities, which, in turn, mediate the effect of feedback characteristics on teachers' performance).

A3. Use of Automated, Electronic, Mechanical or Other Technological Collection Techniques

The survey will be administered online using survey software such as SurveyGizmo in order to reduce burden on respondents. Respondents will be sent a unique link via email that will lead them to the online survey. SurveyGizmo allows the creation of a unique link for each email address in the invite list that prevents duplicate responses, because the link cannot be used after the survey is completed. Researchers will also pre-code items using the online survey software which will reduce the time needed to prepare and merge the teacher survey and teacher evaluation rating data files. Reminders will also be sent via email using the survey software.

Participants' contact information (for sending the online survey) will be collected electronically from a district contact using a secure file sharing software such as Dropbox. Teacher evaluation rating data will also be collected in electronic format and stored in electronic databases created by MRL.

A4. Efforts to Identify Duplication.

This project will use existing teacher evaluation rating data. The survey data is unique data that addresses the need for providing targeted feedback as part of a teacher evaluation system. REL Central reviewed the literature and determined that there is no alternative source for the information to be collected through the survey.

A5. Sensitivity to Burden on Small Entities

The data collection does not involve small entities.

A6. Consequence to Federal Program or Policy Activities if the Collection is Not Conducted or is Conducted Less Frequently

If the proposed data were not collected, the goals of the IES Regional Education Laboratory program may not be met. REL Central would not be able to "conduct and support high quality studies on key regional priorities," resulting in decision makers in the region not having access to the research needed to build "an education system in which decisions are firmly grounded in data and research."[1] Additionally, if the proposed data were not collected, the states would not have timely and relevant research on teacher feedback in evaluation systems to inform training and guidance to evaluators on providing feedback to teachers. The data will be collected only one time; thus, there is no "frequency" component to the data collection.

A7. Special Circumstances

There are no special circumstances.

A8. Federal Register Announcement and Consultation

Federal Register Announcement

ED will publish Federal Register Notices to allow both a 60-day and 30-day public comment period. The REL will assist ED in addressing any public comments received.

The 60 day Federal Register notice was published on 7/10/13, Vol. 78, page 41386. No comments were received. The 30-day Federal Register Notice was published on [DATE] (Appendix X)."

Consultations Outside the Agency

REL Central has consulted with research alliance members on the availability of data, the clarity of the instructions, and the survey instrument. Additionally, we have consulted with content and methods experts to develop the study. The following individuals were consulted on the statistical, data collection, and analytic aspects of this study:

Name	Title	Organization	Contact Information
Dr. Fatih Unlu	Senior Scientist	Abt Associates	Fatih_Unlu@abtassoc.com 617-520-2528
Katy Anthes	Executive Director of Educator Effectiveness	Colorado Department of Education	Anthes_K@cde.state.co.us
Britt Wilkenfeld	Strategic Data Fellow	Colorado Department of Education	Wilkenfeld_B@cde.state.co.us
Dr. Jean Williams	Evaluation Design Specialist	Colorado Department of Education	Williams_J@cde.state.co.us

[1] REL goals were described on the program website: http://ies.ed.gov/ncee/edlabs/about/.

As appropriate, we will consult with representatives of those from whom information is to be obtained, such as research alliance members, and Technical Working Group members to address any public comments received on burden. The Technical Working Group members are experts in content and methodology and they provide consultation on the design, implementation, and analysis of the study. The research alliance members are educators and stakeholders in the Central Region who share a specific education concern. The research alliance members have a shared interest in the proposed study, and they provide feedback on the design and implementation of the study.

A9. Payment or Gift to Respondents

Teachers participating in the pilot will be offered a $30 Visa gift card to participate in the study. Teachers participating in the full study also will be offered an incentive of a $30 Visa gift card to participate in the study. This incentive was determined using NCEE guidance, which suggests a $30 incentive for a 30-minute survey about interventions (NCEE, 2005). Additionally, the results of the Reading First Impact Study (RFIS) incentives study were considered, which found that incentives significantly increase response rates in surveys (Gamse, Bloom, Kemple, & Jacob, 2008). Teachers will receive the incentive upon completion of the survey.

A10.Confidentiality of the Data

REL Central will be following the new policies and procedures required by the Education Sciences Reform Act of 2002, Title I, Part E, Section 183: "All collection, maintenance, use, and wide dissemination of data by the Institute" are required to "conform with the requirements of section 552 of title 5, United States Code, the confidentiality standards of subsection (c) of this section, and sections 444 and 445 of the General Education Provision Act (20 U.S.C. 1232g, 1232h)." These citations refer to the Privacy Act, the Family Educational Rights and Privacy Act, and the Protection of Pupil Rights Amendment. Subsection (c) of section 183 referenced above requires the Director of IES to "develop and enforce standards designed to protect the confidentiality of persons in the collection, reporting, and publication of data." Subsection (d) of section 183 prohibits the disclosure of individually identifiable information, and makes any publishing or communicating of individually identifiable information by employees or staff a felony.

REL Central will protect the confidentiality of all information collected for the study and will use it for research purposes only. No information that identifies any study participant will be released. Information from participating institutions and respondents will be presented at aggregate levels in reports. Unique identification numbers will be assigned to each participating teacher and used to identify all survey responses. The ID number/name association files will be kept secure in a confidential file separate from the data analysis file. No information identifying respondents will be included in the study data files or reports. Data from the online survey software system will be downloaded and deleted from the online system within one week after the survey window closes. These data files will then be stripped of any identifying information. Survey Gizmo, the online survey software, has Safe Harbor Certification and is HIPAA-compliant. SurveyGizmo uses an encrypted (SSL/HTTPS) connection to transfer and collect survey data. Per a report examining SurveyGizmo's security features, third parties

will not access nor be granted access to data, and the survey data can be scrubbed on request (Web Accessibility Center, 2008).

Information will be reported in aggregate so that individual responses are not identifiable. All identification lists will be destroyed at the end of the project. The research team is trained to follow strict guidelines for soliciting consent, administering data collection instruments, and preserving respondent confidentiality. All members of the research team have successfully completed the Collaborative Institutional Training Initiative (CITI) course in the Protection of Human Research Subjects through Liberty IRB.

REL Central will collect consent forms from study participants which will also be stored in a locked file. Copies of the consent form for the pilot and full study are provided in Attachments D and E, respectively. A copy of the affidavit of non-disclosure is provided in Attachment F for each researcher who will have access to the data.

All study materials will include the following language:

Per the policies and procedures required by the Education Sciences Reform Act of 2002, Title I, Part E, Section 183, responses to this data collection will be used only for statistical purposes. The reports prepared for this study will summarize findings across the sample and will not associate responses with a specific district or individual. Any willful disclosure of such information for nonstatistical purposes, except as required by law, is a class E felony.

A11. Additional Justification for Sensitive Questions

No questions of a highly sensitive nature are included in the survey or interview.

A12. Estimates of Hour Burden (total burden is being divided by the two years of actual burden collection to establish an annual figure)

Pilot Study

The total reporting burden associated with the data collection for the pilot study is 672 hours (see Table 1). A stratified random sample of 900 teachers in four districts from one state will be drawn from approximately 6,887 records. Sampled teachers will be invited to attend a 30-minute in-person meeting in which they will learn about the study and sign consent forms if they agree to participate. Of the 900 sampled teachers, we expect that approximately 450 will agree to participate in the study. Of the 450 teachers, we expect that approximately 180 (40%) will complete the survey without any email reminders; the remaining 270 teachers will require email reminders. Each email reminder is estimated to take 2.5 minutes (0.04 hours), with an estimated average of 2 additional emails sent per teacher. The first email notification is estimated as part of the survey completion time. Of the 450 who agree to participate in the survey, an estimated 360 (80%) will complete the survey, with an estimated burden of 30 minutes (0.50 hours) per teacher. In addition to the survey, 20 participants will agree to participate in a 1-hour interview.

Table 1. Estimated annualized burden hours (Pilot)

Instrument	Person Incurring Burden	Number of Respondents	Responses per Respondent	Total Responses	Hours per Response	Total Burden Hours
Recruitment and Consent	Teacher	900	1	900	.5	450
Email reminder	Teacher	270	2	540	.04	22
Pilot Survey	Teacher	360	1	360	.5	180
Interview	Teacher	20	1	20	1	20
					Total	672

Table 2. Estimated annualized cost to respondents (Pilot)

Instrument	Person Incurring Burden	Total Burden Hours	Hourly Wage Rate	Total Respondent Costs
Recruitment and Consent	Teacher	450	$28	$12,600
Email reminder	Teacher	22	$28	$616
Pilot survey	Teacher	180	$28	$5,040
Interviews	Teacher	20	$28	$560
Total		672		$18,816

Bureau of Labor Statistics, Occupational Employment Statistics, estimates annual teacher salary to be $57,810. To calculate an hourly wage, a 40-hour work week was assumed, resulting in an average hourly wage of $27.79.

Full Study

The total reporting burden associated with the data collection for the full study is 3,541 hours (see Table 3). A stratified random sample of 4,892 teachers in 134 districts on three Central states will be drawn from approximately 32,500 records. Sampled teachers will be invited to attend a 30-minute in-person meeting, in which they will learn about the study and sign consent forms if they agree to participate. Of the 4,892 sampled teachers, we estimate that approximately 2,446 will agree to participate in the study and therefore will be invited to take the survey. Of the 2,446 teachers, we estimate that approximately 978 (40%) teachers will complete the survey without any email reminders; the remaining 1,468 teachers (60%) will require up to two email reminders. Each email reminder is estimated to take 2.5 minutes

(0.04 hours), with an estimated average of 2 additional emails sent per teacher. Of the 2,446 teachers who agree to participate in the survey, an estimated 1,956 will complete the survey, with an estimated burden of 30 minutes (0.50 hours) per teacher.

Table 3. Estimated annualized burden hours (Full Study)

Instrument	Person Incurring Burden	Number of Respondents	Responses per Respondent	Total Responses	Hours per Response	Total Burden Hours
Recruitment and Consent	Teacher	4,892	1	4,892	.5	2,446
Email reminder	Teacher	1,468	2	2,936	.04	117
Survey	Teacher	1,956	1	1,956	.5	978
					Total	3,541

Table 4. Estimated annualized cost to respondents (Full Study)

Instrument	Person Incurring Burden	Total Burden Hours	Hourly Wage Rate	Total Respondent Costs
Recruitment and Consent	Teacher	2,446	$28	$68,488
Email reminder	Teacher	117	$28	$3,276
Pilot survey	Teacher	978	$28	$27.384
Total		3,541		$99,148

Bureau of Labor Statistics, Occupational Employment Statistics, estimates Annual Teacher Salary to be $57,810. To calculate an hourly wage, a 40-hour work week was assumed, resulting in an average hourly wage of $27.79.

Annual Responses and Burden

The annual reporting burden is 2,106 hours with 5802 responses.

A13. Estimate for the Total Annual Cost Burden to Respondents or Record Keepers

There are no start-up costs for this collection.

A14.Estimates of Annualized Costs to the Federal Government.

The total estimated cost for this study is $373,250 over three years. The average yearly cost is $124,417. The costs in the pilot study include developing, administering, analyzing, and refining the survey, and conducting and analyzing the interview results. The costs in the full study include administering the survey, collecting and cleaning teacher evaluation rating data, and analyzing and reporting results.

In addition to the evaluation costs, there are personnel costs of several Federal employees involved in the oversight and analysis of information collection that amount to an annualized cost of $5,000 for Federal labor. The total annualized cost for the evaluation is therefore the sum of the annual contracted evaluation cost ($124,417 and the annual Federal labor cost ($5,000), or a total of $129,417 per year.

A15.Reasons for Program Changes or Adjustments

This is a new study.

A16. Plans for Tabulation and Publication.

Tabulation Plans

Pilot Study

The pilot test is intended to test the survey prior to full study implementation on a small sample of teachers in four districts that are implementing a teacher evaluation system that has the intention of informing teacher development through feedback. The pilot study will include the administration of an online survey and interviews.

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 conduct item analysis to identify problematic items and to examine the 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 (see Section A, Figure 3) 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 indices 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 added constraints five-factor model significantly reduces model fit, Chi-square difference tests will be performed.

Full Study

The full study will address the following research questions:

1. What are teachers' perceptions of the usefulness, accuracy, and credibility of the feedback they receive as a part of their performance evaluation?
2. What is the relationship among the characteristics of performance feedback (usefulness, accuracy, and credibility)?
3. What is the relationship between feedback characteristics and teacher responsiveness?
4. What is the relationship between teacher responsiveness and teacher performance?

The study will use descriptive methods, as described below, to address research question 1. For survey items 2 and 3, the median response and the percentage of teachers selecting each response choice will be reported. The median response will be reported because the scales are ordinal and the median is the most appropriate central tendency for ordinal data.

The principal investigator considered different analytic strategies that would provide the best information toward understanding the role of various feedback variables in teacher performance in a teacher evaluation system (research questions 2, 3, and 4). The principal investigator determined that an exploration of the structure of the relationship of the feedback variables as they relate to performance would be most relevant and therefore necessitated the use of structural equation modeling (SEM).

SEM techniques will be used to examine a model of performance feedback within each state. The structural model (Figure 2) and the underlying measurement model for all of the latent variables except the teacher performance variable will be the same for each state (Figure 3). However, because each state collects slightly different evidence of teacher performance in their teacher evaluation system, the indicators will vary as presented in Figure 4. In the model, straight arrows represent a direct effect, while curved arrows represent covariance of the latent variable due to common predictors, which lie outside of the model. The latent variables in the model are the input, mediating, and outcome variables that were identified previously based on research on effective feedback. Table 5 presents the name of each latent variable, the name of each latent variable as identified in the proposed model, and the instrument that will be used to provide a measure of each latent variable. The model suggests that performance (PERF) is directly influenced by both access to learning opportunities (OPP) and responsiveness to feedback (RESP), and also indirectly by usefulness (USE), accuracy (ACC), credibility (CRED), and opportunities (OPP), which influence performance via responsiveness (RESP). The survey instruments and teacher evaluation ratings described in the data collection section are used as indicators of the latent variables present in the proposed structural model. Table 6 presents example teacher survey

items aligned to the variables in the model (variable names are included in parentheses) that are linked to survey data. For the performance variable, teacher evaluation rating data will be collected from the districts. Tables 7 through 9 present a brief description of the indicators on which teachers are rated for each state's teacher evaluation system. Ratings on each of the states' indicators will be collected to serve as the indicators for teacher performance. Prior teacher performance ratings are not included in the analysis because they are not precise due to changes in criteria and evidence of effectiveness made while the states were piloting their evaluation models over the last few years. The loading of indicators on latent variables is presented in the proposed measurement model. As appropriate, based on an examination of comparing descriptive statistics for the variable groupings, background information will also serve as covariates in the model analysis. The measurement model may be modified once the final survey instrument is developed to include all items related to the latent variables. SEM allows for the examination of the latent structure underlying a set of observed variables (Byrne, 1998).

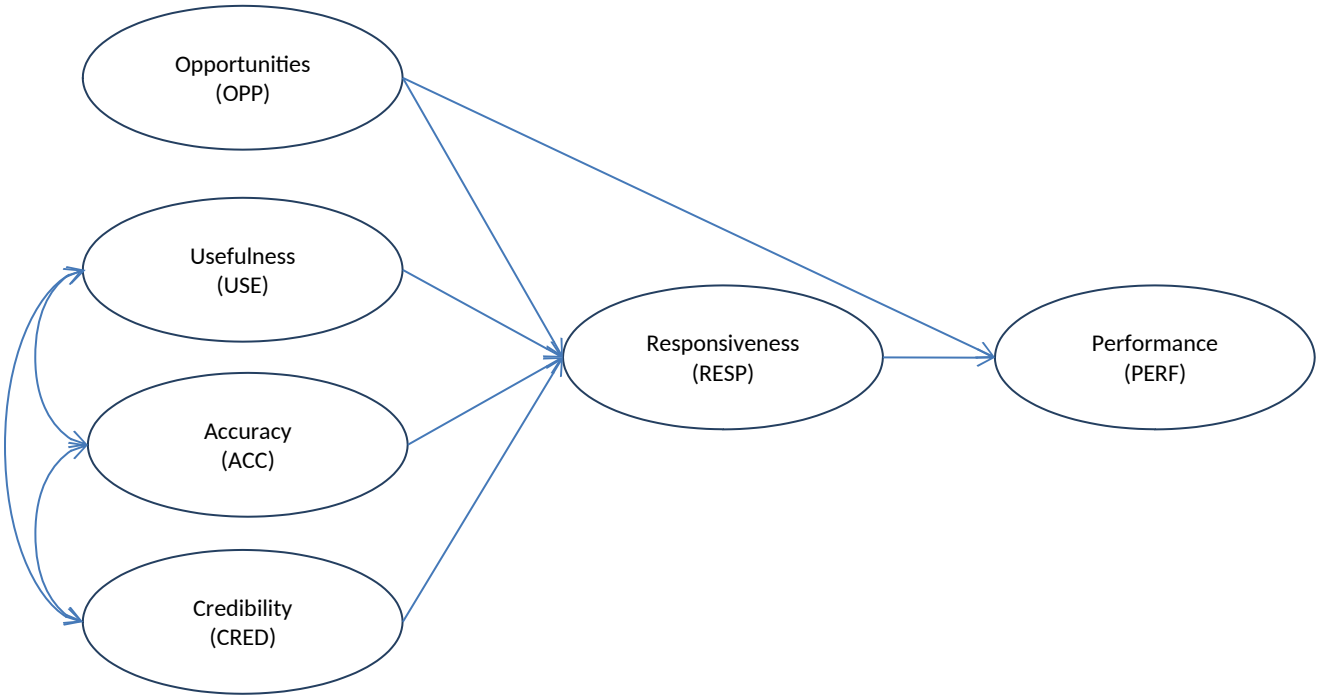

Figure 2. Hypothesized structural model of performance feedback

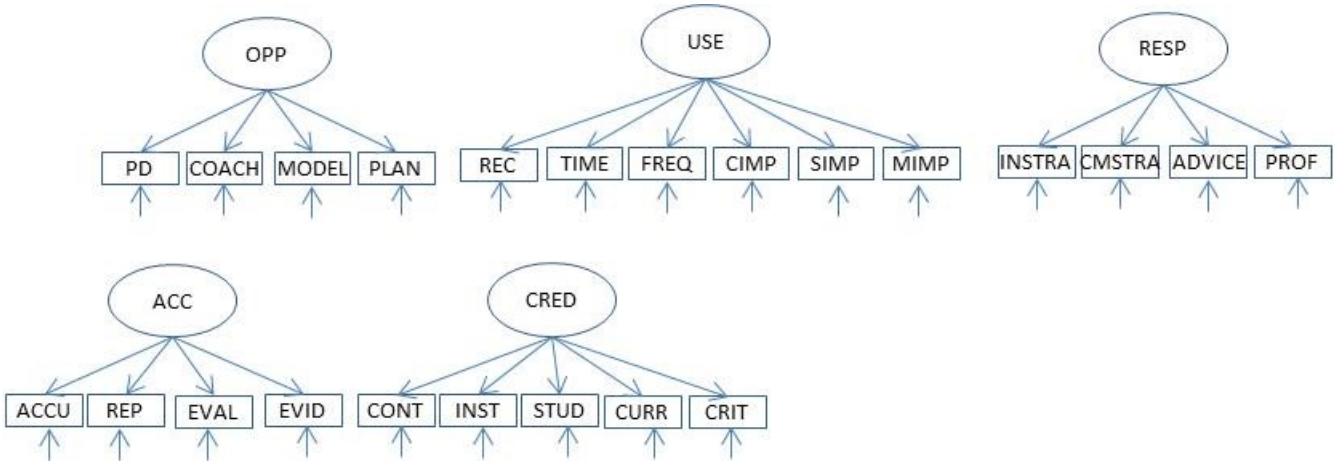

Figure 3. Hypothesized measurement model except performance variable

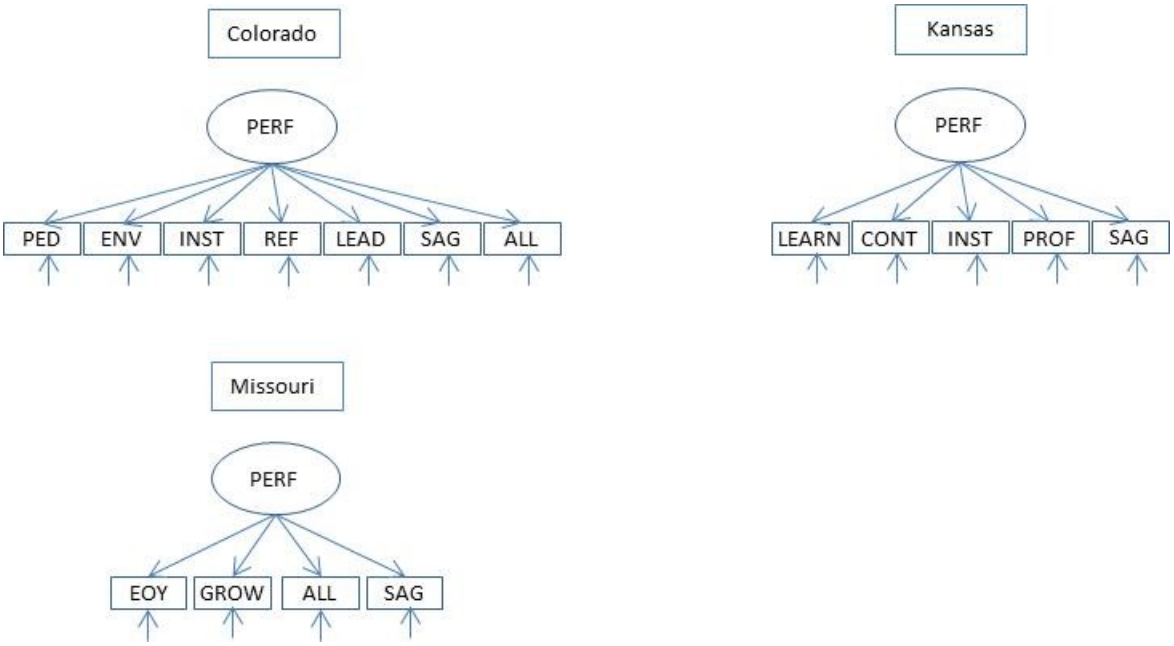

Figure 4. Hypothesized measurement model for performance variable by state

Table 5: Data collection instrument for each variable

Latent Variable	Latent Variable Name in the Model	Instrument
Usefulness of Feedback	Usefulness (USE)	Usefulness items from the researcher developed teacher survey
Accuracy of Feedback	Accuracy (ACC)	Accuracy items from the researcher developed teacher survey
Credibility of the Person Providing Feedback	Credibility (CRED)	Credibility items from the researcher developed teacher survey
Responsiveness to Feedback	Responsiveness (RESP)	Responsiveness items from the researcher developed teacher survey
Access to Opportunities to Learn	Opportunities (OPP)	Opportunities items from the researcher developed teacher survey
Performance Ratings	Performance (PERF)	Ratings on the state's teacher evaluation standards, including professional standards and the student growth standard (provided by the district)

Table 6: Example teacher survey items by model variable

Model Variable	Example Survey Items (model variable names provided in parenthesis)
Usefulness (USE)	- My evaluator recommended next steps for finding professional development to improve my teaching. (REC) - My evaluator provided feedback within an appropriate timeframe. (TIME) - My evaluator provided feedback as frequently as I needed it. (FREQ) - My evaluator's feedback included specific suggestions to improve my content/subject knowledge. (CIMP) - My evaluator's feedback included specific instructional strategies that I could use to improve my teaching. (SIMP) - My evaluator's feedback included specific classroom management strategies that I could use to improve my teaching. (MIMP)
Accuracy (ACC)	- The feedback I received was an accurate portrayal of my teaching. (ACCU) - The classroom observations that informed the feedback I received represented a typical day in my classroom. (REP) - I would receive the same feedback from a different evaluator. (EVAL) - I would receive the same feedback if my evaluator had examined different evidence (for example, if they observed a different lesson, or reviewed additional evidence). (EVID)
Credibility (CRED)	- My evaluator had sufficient knowledge of my subject/content. (CONT) - My evaluator had sufficient knowledge of instructional strategies. (INST) - My evaluator had sufficient knowledge of how my students learn. (STUD) - My evaluator had a sufficient understanding of the curriculum being observed. (CURR) - My evaluator had a sufficient understanding of the established teacher evaluation criteria. (CRIT)
Opportunities (OPP)	- I had access the professional development that I needed in order to implement suggestions provided in my feedback. (PD) - I had access to a coach/mentor who supported me in implementing suggestions provided in my feedback. (COACH) - I was able to observe expert teachers modeling skills that related to my feedback. (MODEL) - I had time during the school day to plan for implementing new strategies based on my feedback. (PLAN)
Responsiveness (RESP)*	- Because of the feedback I received from my evaluator, I tried new instructional strategies in my classroom. (INSTRA) - Because of the feedback I received from my evaluator, I tried new classroom management strategies in my classroom. (CMSTRA) - Because of the feedback I received from my evaluator, I sought advice from a coach or mentor. (ADVICE) - Because of the feedback I received from my evaluator, I sought professional development opportunities. (PROF)

* The responsiveness items were based on items used in Tuytens & Devos (2011), and Geijsel, Sleegers, Stoel, & Krüger (2009).

Table 7: Indicators of performance in the Colorado Model Evaluation System

Name of Indicator for Performance in the Model	Colorado Standard	Description of Indicator
PED	Demonstrate pedagogical expertise in the content	Rating using a rubric based on observations and at least one of the following measures: (a) student perception measures (such as surveys), (b) peer feedback, (c) feedback from parents or guardians; and/or (d) reviews of teacher lesson plans or student work samples
ENV	Establish a safe and respectful learning environment	
INST	Plan and deliver effective instruction	
REF	Reflect on practice	
LEAD	Demonstrate leadership	
SAG	Student academic growth	Rating based on multiple measures of student academic growth, such as median student growth percentiles
ALL	Overall	Weighted average of the ratings received for each of the professional practice standards (50%) and the student growth standard (50%)

Table 8: Indicators of performance in the Kansas evaluation system

Name of Indicator for Performance in the Model	Kansas Standard	Description of Indicator
LEARN	The Learner and Learning	Rating using a rubric based on observations and other evidence, such as student work samples, lesson plans and professional learning plans.
CONT	Content Knowledge	
INST	Instructional Practice	
PROF	Professional Responsibility	
SAG	N/A	Student growth is a significant factor in the teacher evaluation rating. Kansas is currently determining how student growth will be measured and how it will be incorporated into the evaluation system

Table 9: Indicators of performance in the Missouri evaluation system

Name of Indicator for Performance in the Model	Description of Indicator
EOY	End-of-year rating using a rubric on three selected indicators based on evidence including data on professional commitment (such as credentials and professional development plans), professional practice (such as classroom observation data), and/or professional impact (such as student performance measures)
GROW	Growth rating based on difference between the end-of-year and baseline ratings on three selected indicators
ALL	Overall rating on the rubric based 50% on overall growth and 50% on end-of-year ratings
SAG	Student growth is identified as a significant factor in the Missouri Educator Evaluation System, and the system is structured such that a teacher cannot receive a proficient or distinguished rating if student growth is low. The plans for including student growth are currently under development.

Each state's model will be examined separately using a two-step modeling process. First, researchers will examine the fit of the measurement model. Researchers will examine multiple fit indices (χ2, goodness-of-fit index [GFI], parsimony goodness-of-fit index [PGFI], root mean square error of approximation [RMSEA], and comparative fit index [CFI]) and whether or not factor loadings are significant to determine the degree to which the model fits the data. If it is determined that the measurement model fits the data, analysis will proceed with the structural portion of the model. If it is determined that the model does not fit the data, researchers will examine the modification indices, residuals, factor loadings, factor correlations, and theoretical considerations to potentially re-specify the model. The model would only be re-specified only if the re-specification is in line with the theory that supports the model. Re-specification may include allowing indicators to load on more than one factor and allowing measurement errors to covary.

To determine the relationship among the variables proposed in each model of performance feedback, the fit of the structural model will be tested and the path coefficients within the model will be examined. Information about the model for each state is presented in Table 10. Maximum likelihood will be used as the estimation method. In maximum likelihood, "the estimates are the ones that maximize the likelihood (the continuous generalization) that the data (the observed covariances) were drawn from this population"(Kline, 2005, p. 112). Once the model is tested, the path coefficients will be examined, and several goodness-of-fit indices (χ2, goodness-of-fit index [GFI], parsimony goodness-of-fit index

[PGFI], root mean square error of approximation [RMSEA], and comparative fit index [CFI]) will be examined to determine the degree to which the model fits the data. Multiple-model fit indices are needed to accurately examine model fit (Tabachnick & Fidell, 2001). Modification indices, which are reported in most SEM statistical output, will be examined to determine the extent to which the overall model fit would increase if particular paths are freely estimated. In particular, we will examine the path coefficients and modification indices of the characteristics of feedback and modify the model to test whether responsiveness to feedback mediates the relationship between perceptions of feedback and performance.

Table 10: Information about each state's structural equation model[2]

	Colorado	Kansas	Missouri
Parameters	69	65	63
Variances of Measurement Error	30	28	27
Factor Loadings	30	28	27
Factor Correlations	6	6	6
Factor Covariances	3	3	3
Indicators	30	28	27
Data Points	435	378	351
Degrees of Freedom	366	313	288

The assumptions of multivariate normality and linearity will be evaluated by analyzing the multivariate and univariate skewness and kurtosis. The distribution and shape of all survey data and evaluation ratings will be examined to screen for outliers, skewness, and kurtosis. If significant skewness or kurtosis is found (alpha ≤ .01), estimation techniques that address non-normality, such as robust maximum likelihood (RML), will be used. RML gives standard errors with unspecified distributional assumptions, yielding the least biased standard errors when multivariate normality assumptions are false (Bentler & Dijkstra, 1985; Chou & Bentler, 1995).

Researchers will synthesize the findings across the states using meta-analytic structural equation modeling (MASEM) techniques such as the two-stage structural equation modeling (TSSEM) approach (Cheung and Chan, 2005). Using the TSSEM approach, researchers will first use multiple-group confirmatory factor analysis approaches to test the homogeneity of the correlation matrices across the states. Next, the pooled correlation matrix and its asymptotic covariance matrix (ACM) will be calculated from the state matrices. The pooled correlation matrix will be analyzed and its ACM will be used as the

[2] These numbers were calculated using methods proposed by Tabachnick & Fidell (2001). The number of parameters equals the total number of regression coefficients, variances, and covariances to be estimated in the model. Each of the indicators in the models have an associated variance of measurement error and factor loading. The six straight arrows between the factors represent factor correlations and the three curved arrows in the model represent factor covariances. Data points equals $n(n-1)/2$, where n is the number of indicators. Degrees of freedom equals the number of data points minus the number of parameters.

weight matrix using the asymptotic distribution-free method as the estimation method. Fit indices from this analysis will be examined to determine the extent to which model fits the pooled data.

Publication Plans

All results for REL Central's rigorous studies will be made available to the public through peer-reviewed evaluation reports that are published by IES on the IES website. The datasets from these rigorous studies will be turned over to the REL's IES project officer. These data will become IES restricted use datasets requiring a user's license that is applied for through the same process as NCES restricted use datasets. Even the REL contractor would be required to obtain a restricted use license to conduct any work with the data beyond the original evaluation.

Timeline

The timeline for data collection, analysis and reporting is shown in Table 11.

Table 11: Schedule of activities

Study	Activity	Schedule
Pilot Study	Hold in-person recruitment meeting	November 2013–February 2014
	Administer survey instrument	April 2014–May 2014
	Conduct interviews	May 2014
	Analyze and report	May 2014–July 2014
Full Study	Hold in-person recruitment meeting	August 2014–January 2015
	Administer survey instrument	April 2015–May 2015
	Collect existing teacher evaluation rating data	May 2015–June 2015
	Analyze and report	May 2015–November 2015

A17. Approval to Not Display the Expiration Date for OMB Approval

No request is being made for exemption from displaying the expiration date.

A18. Exception to the Certification Statement

No exceptions to the certification statement are requested or required.

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public schools. *Journal of Labor Economics*, *25*(1), 95–135.

Bentler, P. M., & Dijkstra, T. (1985). Efficient estimation via linearization in structural models. In P. R. Krishnaiah (Ed.), *Multivariate analysis VI* (pp. 9–42). Amsterdam: North Holland.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.

Central States Education Laboratory (2011). Technical Proposal – Final Proposal Revisions, Regional Educational Laboratories 2012-2017 Central Region, RFP ED-IES-11-R-0036. Centennial, CO: Marzano Research Laboratory.

Cheung, M. W.-L., Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods*, *10*(1), 40–64.

Chou, C. P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling* (pp. 37–55). Thousand Oaks, CA: Sage.

Colorado Association of School Executives. (n. d.). *Colorado School Leaders Speak Out on Teacher Evaluation*. Retrieved July 18, 2012 from http://www.co-case.org/associations/779/files/2%200%20Colorado%20School%20Leaders%20Speak%20Out%20On%20Teacher%20Evaluation.pdf.

Geijsel, F., Sleegers, P., Stoel, R., & Krüger, M. (2009). The effect of teacher psychological, school organizational and leadership factors on teachers' professional learning in Dutch schools. *The Elementary School Journal*, *109*, 406–427.

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.

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, *64*(4), 349–371.Kane, T. J., & Staiger, D. O. (2012). Gathering feedback for teaching: combining high-quality observations with student surveys and achievement gains. Methods of Effective Teaching project, Bill & Melinda Gates Foundation, Seattle, WA. Retrieved July 18, 2012 from http://www.metproject.org/downloads/MET_Gathering_Feedback_Practioner_Brief.pdf.

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

Konstantopoulos, S., & Chung, V. (2010). The persistence of teacher effects in elementary grades. *American Educational Research Journal 48*(2), 361–386.

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

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*(3), 237–257.

Papay, J. P. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review*, *82*(1), 123–141.

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

Tuytens, M., & Devos, G. (2011). Stimulating professional learning through teacher evaluation: An impossible task for the school leader? *Teaching and Teacher Education 27*, 891–899.

Web Accessibility Center. (2008). *A survey of survey tools.* Author: Retrieved from http://wac.osu.edu/workshops/survey_of_surveys/.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: implications for teacher evaluation. *Journal of Personnel Evaluation in Education 11*: 57–67.