Journal of Applied Research on Children: Informing Policy for Children at Risk

Volume 3
Issue 2 *Measuring Success in Public Education*

Article 4

2012

The Comprehensive Assessment of Leadership for Learning: A Next Generation Formative Evaluation and Feedback System

Carolyn Kelley
University of Wisconsin - Madison, kelley@education.wisc.edu

Richard Halverson
University of Wisconsin - Madison, halverson@education.wisc.edu

The Comprehensive Assessment of Leadership for Learning: A Next Generation Formative Evaluation and Feedback System

Acknowledgements

The authors would like to thank the many educational leaders who contributed to the development and testing of the CALL Survey Instrument, as well as our research collaborators: Mark Blitz, Eric Camburn, Matthew Clifford, Seann Dikkers, Shree Durga, Steve Kimball, Marsha Modeste, Tara Piche, and Jason Salisbury.

Background and Purpose

It would be difficult to overstate the impact of the 2001 Federal No Child Left Behind Act (NCLB) on education and schools in the United States. NCLB created a system of strong accountability for continuous improvement in student achievement test scores in public schools. Under NCLB, schools and districts were considered successful and could avoid sanctions only if academic yearly performance improved toward proficiency for all student subgroups according to a timeline negotiated between each state and the federal government. All students in all subgroups were expected to achieve proficiency on the state tests by 2014.

A recent study[1] of the impact of NCLB on teachers and schools found improvements in elementary school math performance, particularly for traditionally disadvantaged populations, increased expenditures per pupil and the number of teachers with master's degrees, and shifts in school curricula to focus more narrowly on tested subjects. In addition, NCLB promoted data collection and analysis, formative and summative assessment of students, and test preparation activities. The law significantly increased pressure on teachers, schools, and districts to produce results in terms of improvements in student test scores.

NCLB has also promoted increased research attention to the characteristics of schools that succeed in closing achievement gaps and raising achievement for all students.[2-4] This research highlights the importance of *school leadership* in promoting student learning.

While NCLB has been successful in shifting attention in schools toward accountability for state test scores, there is broad agreement that the goal of 100% proficiency in student performance for all student subgroups is not achievable. Efforts to revise the law through the reauthorization process have not been successful, but in September 2011, the federal government invited all states to submit waiver applications to be relieved of the law's 2014 proficiency requirement. To date, 11 states have been granted waivers, and an additional 26 states have requested waivers from NCLB provisions. To be granted the waiver, states must propose an alternative system of accountability for performance improvement that must include: (1) *state curriculum standards and assessments* designed to ensure that students graduate from high school ready for college or careers; (2) *systems of differentiated recognition, accountability, and support* that include "rigorous state interventions" for the lowest performing schools or districts and rewards and recognition for the highest performing schools and districts; and (3) *state guidelines for teacher and principal evaluation and support systems* that include

measures of student learning and professional practice.[5]

This article focuses on principal evaluation and support in the current federal and state accountability environment. While much attention has been paid to how to measure success in public education through student test scores, limited attention has been paid to assessing the leadership practices that promote student learning. We attempt to address this gap in the literature by describing current principal evaluation practices and by profiling the development of a next-generation formative school leadership assessment and feedback system, the Comprehensive Assessment of Leadership for Learning (CALL).

Current Principal Evaluation Practices

The NCLB Waiver Provision has promoted a flurry of activity in the development of state principal evaluation frameworks. Established based on guidance from the U.S. Department of Education, these frameworks are required to focus on holding principals accountable for school-wide student learning gains and for effective principal practices. For example, the Wisconsin Framework for Educator Effectiveness[6] defines a framework for principal evaluation that includes 50% of the evaluation based on student outcomes and 50% based on professional practices. Professional practices are assessed through a state evaluation system or an equivalent process adopted by the school district. Student outcomes are assessed through a formula developed by the state, to include the state assessment results (15%), district assessment results (15%), classroom-level student learning objectives and school-wide reading for elementary and middle schools (15%), graduation rates for high schools (2.5%), and a locally determined optional criterion (2.5%).

A statewide 2011 study[7] of evaluation practices in Wisconsin found that prior to the adoption of the state framework, principal evaluation systems were locally defined and inconsistently administered across districts. Despite significant changes in the practice of leadership in the last decade, over half of the systems in Wisconsin are more than 10 years old. Few districts have defined what it means to be an effective principal, and where these definitions exist, they are not aligned with the evaluation system. In most districts, evaluation design is at the discretion of the evaluator, as comprehensive policies, procedures, or guidelines for principal evaluation do not exist. According to the study, the highest quality principal evaluation systems link the principles of effective evaluation to a clear definition of effective leadership and tie evaluation results to principal professional development.

To an even greater degree, formative performance evaluation of principals tends to be random, unaligned to widely accepted standards of professional practice, and relatively inconsequential to practice.[8] Federal and state-level policy changes heighten the importance of developing systematic, standards-driven principal performance assessments and have heightened interest in performance feedback design.

Existing Systems of School Leadership Assessment

Prior to the development of state evaluation frameworks under the NCLB waiver provision, a limited number of a number of 360-degree and other instruments have been developed to evaluate leadership practice in schools. Condon and Clifford[9] reviewed 20 commonly used leadership assessments for evidence of reliability and validity in their development and use. They found that the instruments vary on important constructs and are based on different conceptions or dimensions of school leadership. Building on the work of Condon and Clifford, next we summarize selected commercial school leader assessments for the type of feedback they provide. These products focus on assessing the leadership practice of the principal to support leadership development. [10]

Vanderbilt Assessment of Leadership in Education (VAL-ED)

VAL-ED is a 360-degree principal assessment tool developed by researchers at Vanderbilt University from research on "learning-centered leadership"[8] and is licensed by Discovery Education. The assessment focuses on principal performance on perceptions of 6 leadership components (high standards for student learning, rigorous curriculum, quality instruction, culture of learning and professional behavior, connections to external communities, and performance accountability) and 6 processes (planning, implementing, supporting, advocating, communicating, and monitoring).

Respondents can include teachers, the principal, and principal supervisor. Each respondent indicates how effective the leader is on each of 72 items. In addition to the effectiveness rating (from "ineffective" to "outstandingly effective"), raters check from a list of 5 sources of evidence (reports from others, personal observations, school documents, school projects or activities, other sources, or no evidence) on which their claims are based.

Feedback reports are provided based on who responded (numbers and distribution), what evidence was used to evaluate the principal, and what the results say about the principal's leadership behaviors. Results can be interpreted against a norm (percentile ranking vs. national sample)

and standards-referenced criteria to highlight strengths and areas for improvement. The assessment was designed to have sufficiently high reliability and validity for use as a summative assessment instrument, but results are intended for both formative and summative purposes.

Balanced Leadership Profile, McREL

The Balanced Leadership Profile, developed by McREL,[11] is based on the 21 leadership responsibilities and dimensions of change identified in meta-analysis research covering over 1000 studies of educational leadership by Marzano et al.[12] The online survey format provides 360-feedback on principal behaviors from respondents including the principal and the option of teachers (a minimum of 5) working with the principal and the principal's supervisor.

The assessment is designed for formative purposes, and school or district leaders receive a report on how the principal scored on the 21 responsibilities and change dimensions. The report also includes questions for the leader to consider relating to areas of needed focus and suggested actions to address the areas. In addition, the leader is provided with online professional development tools and strategies that are linked to the 21 leadership responsibility and change dimensions.

NASSP Leadership Skills Assessment

The National Association of Secondary School Principals (NASSP) Leadership Skills Assessment is based on 10 skill dimensions in 4 main areas: instructional leadership, resolution of complex problems, communication, and development of self and others. It is designed for formative purposes, including professional development of current or prospective principals.

Interested principals are encouraged to review the 10 NASSP skill dimensions and then complete the self-assessment. Up to 15 colleagues may also assess practice in a 360-degree assessment. An in-basket activity is also included, with the option of 2 colleagues assessing principals. When activities are completed, participants can view and print a report for each activity. A summary report can also be printed; this report pulls together data from all 3 sources (self-assessment, 360-degree assessment, and in-based activity). The report ranks skills in terms of principal developmental interests and level of skill demonstrated. Principals can then identify which areas to develop and match with professional development activities. Based on that information, a final report is provided with suggested professional development.[13]

Summary of Instruments

Each of these instruments is based on an interpretation of the literature on leadership effectiveness or on the Interstate School Leadership Licensure Consortium (ISLLC) standards.[14] All include 360-degree evaluations of leadership, and all but one (VAL-ED) are solely intended for formative purposes. Although all of the systems reviewed by Condon and Clifford[9] provide some evidence of psychometric testing, few meet minimum bars for validity and reliability. Moreover, the majority of assessments were developed 10 to 15 years ago, when school leaders operated in a significantly different policy and accountability context with different expectations for leadership focus and performance. None of the reviewed assessments displayed evidence of consequential validity, defined as the power of the assessment to promote changes in practice. Further, these systems appear to focus almost entirely on the leadership of an individual, primarily the school principal.

The Comprehensive Assessment of Leadership for Learning

This article describes the design and validation of the Comprehensive Assessment of Leadership for Learning (CALL), a 360-degree, online, formative assessment and feedback system for middle and high school leadership. CALL is designed to assess specific leadership practices or tasks characteristic of high-performing middle and high schools.[2,15.16] The survey captures leadership practices and school cultures across 5 domains of leadership practice:

1. Focus on Learning
2. Monitoring Teaching and Learning
3. Building Nested Learning Communities
4. Acquiring and Allocating Resources
5. Maintaining a Safe and Effective Learning Environment

CALL is unique among school leadership assessments in 3 ways. First, it focuses on distributed leadership, rather than the leadership of the principal. Second, it captures leadership practices rather than opinions about leadership. Finally, it is designed to address the specific accountability context of NCLB and the changes in school leadership that have resulted from this law. The CALL assessment is designed to measure the presence of formal and informal leadership practices distributed throughout the school that promote student learning and advance learning equity for children at risk. A brief description of each of the 5 CALL domains follows.

The CALL *Focus on Learning* domain contains 4 sub-domains: maintaining a school-wide focus on learning; formal leaders are recognized as instructional leaders; focusing on a collaborative design of an integrated learning plan; and providing appropriate services for students who traditionally struggle.

The *Monitoring Teaching and Learning* domain contains 4 sub-domains: formative evaluation of student learning; summative evaluation of student learning; formative evaluation of teaching; and summative evaluation of teaching.

The *Building Nested Learning Communities* domain contains 4 sub-domains: collaborative school-wide focus on problems of teaching and learning; professional learning; socially distributed leadership; and collegial relationships.

The *Acquiring and Allocating Resources* domain contains 5 sub-domains: personnel practices; structure and maintenance of time; focus of school resources on student learning; integration of external expertise into the school instructional program; and coordination and supervision of relations with families and the external communities.

The *Ensuring a Safe and Effective Learning Environment* domain contains 4 sub-domains: clear, consistent, and enforced expectations for student behavior; safe learning environment; student support services that provide a safe haven for students who traditionally struggle; and buffering of the teaching environment.

Assessing Distributed Leadership

Consistent with research on distributed leadership,[17,188-19] the CALL survey defines leadership as distributed across the entire school organization, rather than as the actions or behaviors of a single person. Thus, the CALL survey examines the set of leadership practices carried out by formal and informal leaders distributed throughout the school. This is unique to CALL and its design as a formative organizational assessment instrument, rather than as a formative or summative assessment of the leadership of the principal. The focus on leadership is consistent with our theoretical framing of leadership as distributed across the school organization[19] and with the principles of effective feedback, which suggest that to motivate and direct improvements in performance,

feedback should focus on the task and task performance, not on the individual person or the person's self-concept.[20]

While the principal plays a central role in laying the groundwork for advancing student learning, particularly at the middle and high school levels, issues such as large school size, complexity of organizational cultures, and norms of teacher autonomy and isolation highlight the importance of assessing and developing *leadership throughout the school* rather than simply focusing on a single *school leader*. Distributed leadership is critical because it is ultimately change in practice at the classroom level that determines whether school improvement plans will have a direct impact on student learning.

Assessing Leadership Tasks

Leadership is not merely a generic feature of organizations. Rather, leaders across the school engage in a series of tasks that establish the conditions for teaching and learning in schools.[21] Improving organizational leadership requires tools that help researchers and practitioners identify these tasks, determine who performs them, and then measure the degree to which the tasks actually improve teaching and learning.[15] Tools that provide information on the key tasks of leadership practice can provide principals with valuable feedback to aid in the improvement of leadership across the school organization and support to guide the ongoing development of instructional leadership throughout the continuum from novice to expert practice.

CALL measures distributed leadership tasks by asking specific questions about the practices carried out within the classroom, in interactions between teachers and other staff members, and across the school organization. For example, the survey assesses the frequency of teacher conversations with other teachers about student work, test scores, and instruction as a measure of instructional leadership practiced by teachers in small peer groups. The extent to which these interactions are formally structured is a measure of the principal's instructional leadership to structure time to facilitate professional teacher collaboration and to create an expectation that teachers will use the time to talk about teaching and learning.

Providing Formative Feedback

The recent proliferation of benchmark assessment systems in public schools[22] demonstrates the felt need for educators to receive timely,

standardized feedback on the progress of student learning. School leaders also need timely information on the progress of local initiatives in professional development, resource allocation, assessments, and school safety to improve teaching and learning across the school.

The CALL assessment system is designed to provide formative feedback to strengthen instructional capacity by providing timely information on *leadership task enactment* across the school organization at the middle and high school levels. The potential for feedback on distributed leadership tasks is an important alternative to individual leader performance feedback because leadership is distributed differently in different schools depending on the organizational context, individual leader skills and expectations of the role, and the distribution of expertise across the school. For example, if a school has an assistant principal who is a very strong instructional leader, the principal may choose to delegate some of these responsibilities to the assistant principal and focus his or her own work on other aspects of school leadership or management. Thus, assessing the quality of leadership by examining an individual leader may misrepresent the practice of leadership in the school organization.

Furthermore, focusing on formative assessment of leadership practices, rather than on the characteristics of an individual leader, provides clearer guidance to the school on how to improve. For instance, if the assessment system identifies the leader as a weak communicator, the path toward strengthening communication skills may be unclear. In contrast, if the system identifies the school as having weak communication practices, such as the lack of a clear system for communicating student progress to parents and families, the path toward addressing this gap is much clearer.

The CALL feedback system includes 3 levels of feedback. First, school leaders receive a report showing summary results of leadership practices by domain and sub-domain and an item-level distribution report that provides information about the range of responses as well as the average response. This information can help school leaders identify professional development needs, local expertise, and the distribution of leadership practices across the school. Second, school leaders receive information about effective practices for each domain and sub-domain. This information draws from the research literature to define effective practices. Third, school leaders receive guidance on specific steps they could take and tools they could use to strengthen distributed instructional leadership in their school. The guidance and tools are based on our prior research in schools that have successfully closed achievement gaps and

advanced learning for all students.[2,16] Together, the feedback provides a roadmap for school leaders in identifying paths to move their schools forward in addressing the learning needs of *all* students, including children at risk of failure.

Success in the Current NCLB Accountability Context

School leaders undertake a variety of leadership and management tasks to promote desirable outcomes for schools. The CALL survey is designed to support leadership tasks that promote *student learning*. Feedback associated with the survey provides information for the principal or leadership team on how to align school structures and cultures to high expectations for learning for all students and to promote improved outcomes on high-stakes tests. The tasks assessed in CALL were identified through research on schools that have closed achievement gaps and improved learning for all students as measured by multiple learning goals, including achievement on state tests (which is the accountability measure used in NCLB).

Furthermore, an analysis of the relationship between the CALL survey and the ISLLC standards, on which many of the NCLB waiver principal evaluation frameworks are based, showed a strong relationship between CALL and ISLLC. Of the 6 ISLLC standards, the first 5 standards were the most strongly related to CALL. The last standard, which focuses on leadership in the larger political environment of schools, was not strongly related to CALL, because CALL's focus is on the teaching and learning environment of the school, not on the principal's leadership activities outside of the school.

Research on leadership development in schools suggests that development activities focused on building distributed leadership can be a highly effective means of strengthening the leadership of the individual principal as well as the leadership of teachers and others in the school. The focus on distributed leadership diverts the leader from a defensive posture to critical feedback and instead helps him or her to focus on how to model effective leadership behavior to teach others to become stronger instructional leaders.[23]

Thus, CALL is potentially an important tool for principals seeking to improve leadership performance and student outcomes within the current and evolving NCLB accountability environment.

Development of the CALL Survey

Development of the survey instrument began in 2009 with the support of a 4-year grant from the U.S. Department of Education to design and validate

the survey instrument. To ensure that the survey met what we viewed as 4 essential design criteria, the process of designing survey items involved attention to design issues consistent with the formative nature of the assessment. Specifically, each item was designed to be:

- aligned with *research on effective middle and high schools;*
- grounded in *leadership practices rather than opinions* about the leader; and
- framed to *communicate transparently the underlying theory of action*, so that the process of taking the survey would serve as a developmental experience for school leaders and instructional staff; and
- consistent with *best practices in survey design.*

To ensure that these criteria were met, an initial draft of the survey was developed by the research team based on rubrics created by Richard Halverson in conjunction with the University of Pittsburgh Institute for Learning (IFL).[24] These rubrics were consistent with research conducted by Carolyn Kelley and James Shaw[2] on leadership in schools that had consistently closed achievement gaps and improved overall student learning. We also conducted extensive reviews of research on effective middle and high school leadership and on each of the domains of practice to ensure that item development was consistent with the research literature on effective leadership for learning and more specifically on the practices represented by specific survey domains, sub-domains, and individual items.

Distributed leadership analyses propose that practice is composed of macro- and micro-tasks.[21] Macro-tasks refer to the general organizational tasks, such as providing adequate resources, planning, and designing professional development, that organize much of the work of school leadership. Micro-tasks articulate these general responsibilities into the day-to-day activities of school leaders. Our survey design work translated micro-tasks into items that described practices that could be observed by teachers, leaders, and staff in a typical school context. Our focus on middle and high school leadership contexts led us to describe micro-tasks to reflect the appropriate departmental, grade level, and instructional staff (e.g., special education, counseling, instructional coaches, and mentor) contexts. The CALL survey articulated the work of school from 5 leadership macro-tasks into 115 items that described micro-tasks relevant for improving learning. The tasks described in the 5 domains include: 1) focus on learning; 2) monitoring teaching and learning; 3) building nested learning communities; 4) acquiring and allocating resources; and 5) maintaining safe and effective learning

environments. Each macro-task, or domain, is organized into 4 to 5 sub-domains, which contain the specific items.

Practices Rather Than Perceptions

CALL is designed to capture levels of leadership for learning by measuring existing leadership and learning practices from the perspective of school leaders and staff and providing feedback to strengthen leadership. CALL is designed to provide feedback in 3 ways:

- through *transparency in the design of assessment items* so that learning occurs as educators take the assessment;
- by providing *assessment results* that identify established levels of expertise, patterns in response items, and more traditional statistical summaries of results; and
- by providing *leveled guidance on next steps* for strengthening and building principal and distributed leadership for learning.

A major goal of the survey design process was to ground survey items in choice options that reflect actual practices, rather than framing responses in terms of perceptions of leadership practice (e.g., “strongly agree” to “strongly disagree” or “not at all” to “to a great extent”).

To the extent possible, the survey relies on prevalence of practices (e.g., what is the number of times per week teachers meet to talk about instruction?) rather than perceptions (e.g., to what extent do you think your principal is an effective instructional leader?) to gather data on leadership practices. By being explicit about a choice set ranging from low to high levels of practice, the survey provides clearer information about best practices underlying the assessment items and attempts to contextualize item response choices. The resulting survey has a relatively high cognitive demand, but items reflect actual practices in schools, consistent with a clearly specified model of leadership.

In addition, the leadership domains and rubrics that underlie survey design are available to participating schools and districts and provide a clear identification of critical elements of effective leadership for learning, specified in the 5 CALL domains of leadership practice.

A consistent comment we have received from practitioners who have taken the survey is that it is comprehensive and that taking the survey provided them with an opportunity to think about the things that they should do, that they do well, and that they need to work on in their leadership practice. Because CALL reflects a model of distributed leadership, broad participation in the survey helps build awareness of leadership practices and challenges across the school community.

Consistent with Best Practices in Survey Design

The CALL survey design process began with initial construct identification and survey development based on the Halverson rubrics in Fall 2009. Beginning in Spring 2010, research to support item selection and construct validation was undertaken at the University of Wisconsin-Madison and through the North Central Regional Education Laboratory.

Item Selection and Construct Validation

In Spring 2010, 2 practitioner focus groups reviewed the draft survey design. The middle school practitioner group consisted of 2 principals, an assistant principal, a school psychologist, a former Title 1 reading specialist, a special education teacher, and a language arts coordinator for the district. The high school practitioner group consisted of a principal, a department chair, a special education teacher, an assistant principal, and a former high school principal. All practitioners were drawn from different schools, although some of the schools were located in the same district.

The groups met 7 times over the course of 4 months. In each meeting, the practitioners examined a specific domain closely, with the goal of providing feedback on the appropriate use and clarity of language, appropriateness to school level, importance of the question (including items that were not focusing on critical features of the construct or missing items), advice on who in the school should answer the question, and whether there should be any format changes to the questions. Individuals were also asked to determine whether the response options reflected the appropriate range of practice in middle and high schools.

Changes included adjusting language and defining terms, adding new items, revising items to address core issues more effectively, creating multiple items out of a single item to eliminate double-barreled questions, and changing response options to reflect gradations of practice more accurately.[25]

School Leadership Team Focus Groups

Upon completion of the initial survey revision from the practitioner groups, 78 school-level leaders from 11 middle and high schools in Illinois and Wisconsin took the survey and provided feedback on its design and usefulness for leadership development and school improvement. Leadership team members were selected because they were likely to use CALL data for decision making. The focus groups completed the online CALL survey, rated the clarity of CALL survey items, and provided feedback on the utility of CALL data for application to school-level decision making.

The focus group data indicated that this initial draft of the survey was comprehensive and reflected major school leadership systems and actions and that taking the survey prompted self-reflection on leadership quality. Of concern to the leadership teams was the length of the survey, which required in excess of 35 minutes for respondents to complete. Suggestions included splitting the survey into pieces or allowing individuals to leave the survey and return later to complete remaining items.

In addition, the focus groups provided similar feedback as the practitioner groups regarding terminology, double-barreled questions, the reorganization of items to speed response time, the need to make questions and responses more concise, and the elimination of question redundancy.

From a utility perspective, the focus groups indicated that they did not have access to these data from other sources. They believed that these data would assist them in improvement planning, particularly if they could be combined with some demographic data that could show response differences among different groups (e.g., departments, leadership team members, etc.) within the school. They suggested that we consider providing access to research associated with constructs to reinforce the importance of leadership system quality for school improvement as well as access to other, similar schools and high-scoring schools so leaders can connect with others about how to improve practice.

Year 2: Pilot Testing

The survey was pilot tested in 2010-11 with 1784 educators in 11 school districts in the midwestern and southern United States. In addition, 3 rounds of interviews were conducted around school context, survey administration, and utility of feedback with the principals and other survey users (i.e., leadership team members, teachers, and other staff) in 6 schools.

In addition to using the pilot testing as an opportunity to explore the utility and practicality of the survey, the pilots provided an important opportunity to test the Web-based survey platform and identify any particular challenges associated with large-scale survey administration. Six of the pilot schools were involved in 3 rounds of interviews regarding school context, survey administration, and design and utility of feedback. Round 1 was an interview with the school principal; these interviews focused on understanding the organizational and leadership context of the school. These data were designed to triangulate survey data to enable us to check on the ability of the survey to capture critical context and

leadership factors that shape the leadership challenge and impact student learning.

Round 2 involved interviews with the principal and teachers at each school to discuss survey administration and to capture principal and teacher experiences in taking the survey. From this round, interviewees indicated that taking the survey was an important professional development opportunity for themselves and other staff members. The survey was designed to communicate a clear theory of action to survey respondents through the focus of questions and the response options, which were ordered to be transparent in terms of levels of practice. Teachers reported that the survey promoted informal conversations about professional practice among teachers and created expectations for future action in the school. In addition, responding to the survey was perceived as a means of communication with the principal about the school's strengths and weaknesses.

Round 3 of data collection involved an interview with the principal to discuss survey results. In addition, the CALL instrument provides formative feedback to school leaders through summary results by domain, sub-domain, and item. Based on these results, school leaders also receive suggestions on leadership practices they could engage in to strengthen distributed instructional leadership. The feedback designs reflect research on the principles of effective feedback,[20,26] professional development,[8,27,28] and adult learning.[29-31]

In the interview with the principal, we specifically asked for information on how the principal would use the results of the CALL survey and what form of feedback would be most useful to principals. The survey is designed as a standards-based rather than norm-referenced survey, and principals were more interested in data communicating levels of practice (a frequency distribution across item responses) rather than summary data that would likely be used for comparative purposes (e.g., my school received an average of "3" on this response and the other schools in the district received a "2.5").

Principals also identified specific design features that promote effective communication of results and support mobilization for improvement, including:

- the transparency of the survey, which communicates effective research-based practices to survey participants and gives the school a sense of areas it may need to work on;
- presentations of survey results that promote clear identification of current and desired practices;

- item-by-item guidance on steps the school could take to improve practice for that particular item, connected to a larger vision or theory of action for effective leadership practice;
- the survey's focus on distributed leadership rather than the leadership of a particular individual. Principals felt that this design communicated the important role that all teachers play in taking ownership of leadership for learning in the school.

Statistical Analyses of Year 2 Survey Pilot Data

In addition to the qualitative information collected through interviews and focus groups, Year 2 survey pilot data were analyzed to assess the statistical properties of the survey to inform further survey refinement. Some features of the CALL survey include intentionally designing the survey instrument itself as a meaningful professional growth activity. At times, designing the survey instrument to provide formative information for school leadership development was in tension with other goals of survey design, such as designing to optimize statistical properties of the survey. While weighting one set of goals for survey design at times interfered with addressing other goals of survey design, throughout the process we prioritized item design that reinforced formative features of the survey as well as maximizing statistical properties.

Examples of the trade-offs in designing items for statistical properties versus formative feedback occur in the design of item response sets that are ordered so the respondents can clearly identify desirable and undesirable practices, that focus on practices rather than perceptions, and that include items that have response choices that are theoretically desirable but occur rarely in schools. These design features increase the risk of socially desirable responses, increase the cognitive demand of the survey, and include items that may reduce scale reliability. For example, a survey item focusing on the use of technology to support student learning includes high-end options that are atypical in most schools. However, because the survey is intended to communicate best practices to encourage schools to strive for more effective practice, we chose to leave in high-end options even though few schools choose those options. These decisions make the survey results robust for communicating a theory of action to survey respondents and provide valuable data for discussion in staff meetings, but they slightly reduce the robustness of statistical properties of the survey. Further research is needed to explore these trade-offs, but we believe prioritizing the formative goals of the instrument enhances its utility as a vehicle for formative conversations to improve leadership practice.

Reliability Analysis

Reliability is a basic measure of the validity of a scale. Conceptually, reliability is defined as the degree to which a scale is free from errors of measurement. Measurement errors will be higher to the extent that different measurements of the same person vary. Reliability is operationalized as a measure of the degree of consistency between multiple, equivalent measurements of the same construct. Reliability is higher when multiple measurements are more consistent with each other and lower when measurements are less consistent. An important property of reliability statistics is that they tend to increase with a greater number of measurements.

In the CALL survey, multiple survey items that measure a particular sub-domain can be viewed as multiple measurements of a construct. For example, the CALL survey includes 6 items that are intended to measure sub-domain 1.2, "Formal Leaders are Recognized as Instructional Leaders." Our goal in instrument design was to achieve a reliability of at least .7 for each of the sub-domains. Reliability analysis based on the CALL Survey Version 1.0 provided mixed results in achievement of that goal, with initial Chronbach's Alpha reliability scores of .7 or above for 11 of the 16 sub-domain scales. For each scale with a reliability score below .7, we have reviewed items in that scale and have added items or revised items to improve reliability. The reliability analysis is being repeated for CALL Version 2.0 following administration of the revised survey in Spring 2012.

Rasch Analysis

We applied the Rasch model to CALL survey items to better understand scale reliability and the degree to which items within a sub-domain tapped a unitary dimension. The Rasch model is perhaps the simplest item response theory model that considers individual persons' responses relative to the response frequencies of all people. Item "fit" statistics from these models provided a useful diagnostic of how well particular items measured intended sub-domains. Item "difficulties" estimated from Rasch models provide evidence about whether there are sufficient items at all levels of the distribution of people on the scale to provide valid measures of the full range of sample members.

Scale reliabilities from the Rasch were similar to the standard Chronbach's Alpha statistics presented earlier. Items with poor fit statistics were identified as candidates for deletion or significant revision. Analysis of item difficulties identified sub-domains that would benefit from adding

"harder" or "easier" questions. Reliabilities on some sub-domains were identified as low due to the limited number of questions in that domain (the survey had 1 to 3 items in some of the domains).

Not surprisingly, several of the items identified for revision and sub-domains identified for addition of items had already been flagged through the qualitative reviews of the survey (i.e., reviews based on findings from the initial focus groups, pilot interviews, and data analysis). The CALL survey was further revised as needed to address issues identified through the Rasch analysis. These changes included adding items to the sub-domain, moving items from one sub-domain to another that had a better conceptual fit, recalibrating response options in survey items with skewed distributions, and refocusing items that were reducing scale reliability.

Variance Decomposition

A variance decomposition was conducted to assess the within-school versus across-school variance of survey items. Similar to other research in education, the decomposition of variance for most sub-domain scales showed more variance within schools than across schools. Typically, approximately 10% to 20% of the variance in sub-domain scale scores lay between schools. For formative feedback purposes, we believe that within-school variance can be as important as between-school variance for promoting discussions of differences in practice across classrooms or departments within a school.

The between-school variance analysis provides an opportunity to recognize important contextual and performance differences between schools as well. A challenge throughout the process has been to interpret and respond appropriately to recognized survey diagnostic procedures within a formative assessment context.

Item Frequency Distributions

Item frequency distributions were produced to provide an opportunity to use analysis of frequency distribution to inform survey refinement and to explore initial results of the CALL survey. Frequency distributions were produced for the teacher and principal versions of the CALL survey.

The results were examined by CALL researchers in a collaborative meaning-making session and compared to results of the reliability, Rasch, and variance decomposition analyses. Three primary patterns emerged as important for informing survey refinement.

First, items that clustered around a single response were identified for possible refinement in terms of adjusting response options to capture

variations in practice that were currently being grouped into a single response category.

Second, items that resulted in unexpected results were identified as needing additional refinement to clarify the question or response options. For example, a very high percentage of respondents in the pilot indicated that they participated in or experienced learning walks in the school. For this item, we determined a need to more clearly define the term "learning walks" to ensure that we were capturing actual practice and not a misinterpretation of the meaning of the question and response options.

Third, items that did not successfully distinguish between schools were identified for refinement or elimination to ensure that the item was successfully distinguishing between schools in important leadership dimensions.

Consistency with Research on Effective Survey Design

The statistical analyses and qualitative studies described above helped to inform significant modifications to the survey instrument. In Spring 2012, CALL researchers contracted with the University of Wisconsin Survey Center to assess survey design and identify further refinements in the survey instrument, site recruitment, and survey administration and to develop a robust Web-based platform for administration of the survey at a broader scale.

The survey instrument was reviewed for its conformity with best practice in survey design, including rules about question wording, question structure, response format, reference period, definitions, and instructions. The Survey Center made recommendations as follow:

- Simplify wording and sentence constructions to promote cognitive processing by respondents that is more accurate and reliable.[32,33]
- Use "native" terms and phrases instead of "analytic" terms and phrases. Although the Year 1 practitioner focus group and leadership team reviews focused on this issue, the Survey Center review revealed additional analytic terms to consider for revision. Research demonstrates that the respondents' ability to comprehend questions and retrieve information is better when the words and phrases used in the question match those used by respondents.[34]
- Use parallel question wording and question structures. This was a particular challenge in the CALL survey, since the items describe actual practices, reducing the ability to common response choice patterns across questions.

- Avoid double-barreled questions. Research demonstrates that double-barreled questions slow respondents' cognitive processing.[35]
- Include a clear reference period when asking about events or behaviors that occurred in the past. A reference period is the period of time you want the respondent to consider when answering a question about an event or behavior that occurred in the past. Reference periods should appear as the first element in a question and be explicit, specific, and an appropriate length for the item you are asking about.[36]
- Incorporate definitions and instructions into the body of questions to ensure that all respondents have the same information when they answer a question.

Recommended revisions to question design were reviewed and incorporated as appropriate into the survey design.

Discussion and Future Research Directions

We present our research on the development of the CALL survey within the broader context of federal efforts to promote stronger systems of principal and teacher evaluation and support because we believe that CALL has the potential to support leadership development and school improvement and provide school leaders with clearer information about how to improve their own leadership practice, as well as teaching and learning in their schools.

In 2012, CALL is being administered in 120 middle and high schools across the country. In this phase of development, CALL survey results will be compared with value-added test scores, climate survey data, and other measures of leadership to validate the instrument and will, to our knowledge, make CALL the first validated formative leadership assessment instrument. The CALL feedback system is under development as well to provide specific feedback to school leaders on how to interpret CALL results and to help them identify what steps they can take to strengthen distributed leadership for learning in their schools.[37]

CALL researchers are also currently developing an elementary school and district version of the survey. While the school-level specific instrument provides important context-specific data about leadership for learning in middle and high schools, many districts have expressed interest in administering CALL to all of their schools in order to provide data to support district-wide leadership development initiatives. We plan to pilot an initial version of the elementary and district surveys in 2013 and to continue development and validation of these instruments moving forward.

References

1. Dee TS, Jacob B. The impact of No Child Left Behind on student achievement. *J Policy Analysis Manage*. 2011;30:418-446.
2. Kelley C, Shaw JJ. *Learning First! A School Leader's Guide to Closing Achievement Gaps*. Thousand Oaks, CA: Corwin; 2009.
3. Chenoweth K. *How It's Being Done: Urgent Lessons from Unexpected Schools.* Cambridge, MA: Harvard Education Press; 2009.
4. Odden AR, Archibald SJ. *Doubling Student Performance: . . . And Finding the Resources to Do It.* Thousand Oaks, CA: Corwin; 2009.
5. Office of the Press Secretary, White House. Fact sheet: Bringing flexibility and focus to education law. http://www.whitehouse.gov/sites/default/files/fact_sheet_bringing_flexibility_and_focus_to_education_law_0.pdf. Published September 23, 2011. Accessed March 31, 2012.
6. Wisconsin Department of Public Instruction. *Wisconsin Framework for Educator Effectiveness Design Team Report and Recommendations*. http://dpi.wi.gov/tepdl/pdf/ee_report_prelim.pdf. Published November 2011. Accessed March 31, 2012.
7. Masterson SA. *The Role of Evaluation and Professional Development Practices in Advancing Mastery of Principal Practice* [dissertation]. Madison: University of Wisconsin; 2011.
8. Goldring E, Porter A, Murphy J, Elliott SN, Cravens X. Assessing learning-centered leadership: connections to research, professional standards, and current practices. *Leadership Policy Sch*. 2009;8:1-36.
9. Condon C, Clifford M. *Measuring Principal Performance: How Rigorous Are Commonly Used Performance Assessment Instruments?* Naperville, IL: Learning Point Associates; 2010.
10. Kelley C, Kimball S, Clifford M. Building effective formative feedback. Paper presented at: the annual meeting of the University Council for Educational Administration Annual Convention; November 2010; New Orleans, LA.
11. McREL. Balanced Leadership Profile®. http://blp.changetheodds.org/LearnMore. Accessed August 29, 2012.
12. Marzano RJ, Waters T, McNulty BA. *School Leadership That Works: From Research to Results.* Alexandria, VA: Association for Supervision and Curriculum Development; 2005.
13. National Association of Secondary School Principals. *NASSP Leadership Skills Assessment.* http://www.nassp.org/professional-development/nassp-leadership-skills-assessment-. Accessed April 1, 2012.

14. Council of Chief State School Officers. *Educational Leadership Policy Standards: ISLLC 2008 as Adopted by the National Policy Board for Educational Administration.* http://www.ccsso.org/Documents/2008/Educational_Leadership_Policy_Standards_2008.pdf. Published April 2008. Accessed April 1, 2012.
15. Halverson R. A distributed leadership perspective on how leaders use artifacts to create professional community in schools. In: Stoll L, Louis KS, eds. *Professional Learning Communities: Divergence, Detail, and Difficulties*. Maidenhead, UK: Open University Press; 2007:93-105.
16. Kelley C. *Advancing Student Learning Through Distributed Instructional Leadership: A Toolkit for High School Leadership Teams.* Madison, WI: Wisconsin Department of Public Instruction; 2010.
17. Leithwood K, Mascall B. Collective leadership effects on student achievement. *Educ Adm Q.* 2008;44:529-561.
18. Leithwood K, Seashore-Louis K. *Linking Leadership to Student Learning.* San Francisco: Jossey-Bass; 2012.
19. Spillane JP, Halverson R, Diamond JB. Investigating school leadership practice: a distributed perspective. *Educ Res.* 2001;30(3):23-28.
20. DeNisi AS, Kluger AN. Feedback effectiveness: can 360-degree appraisals be improved? *Acad Manage Executive.* 2000;14:129-139.
21. Spillane JP, Halverson R, Diamond JB. Towards a theory of leadership practice: a distributed perspective. *J Curriculum Stud.* 2004;36:3-34.
22. Means B, Padilla C, Gallagher L. *Use of Education Data at the Local Level: From Accountability to Instructional Improvement*. Washington, DC: US Dept of Education, Office of Planning, Evaluation and Policy Development; 2010.
23. Kelley C, Salisbury J. Leadership for learning at the high school level: leveraging the role of department chair as instructional leader. *J School Leadership.* In press.
24. Halverson R. Rubrics. http://www.callsurvey.org/resources/rubrics. Published 2005. Accessed March 20, 2012.
25. Blitz M, Clifford M. Content validity as a window into leadership practice. Paper presented at: the annual meeting of the University Council for Educational Administration Annual Convention; 2010; New Orleans, LA.
26. Wimer S. The dark side of 360-degree feedback: the popular HR intervention has an ugly side. *Training Dev.* 2002;36(9):37-44.
27. Desimone L, Smith TM, Phillips KJR. Does policy influence

mathematics and science teachers' participation in professional development? *Teachers Coll Rec.* 2007;109:1086-1122.
28. Smylie MA, Bennett A, Konkol P, Fendt CR. What do we know about developing school leaders? a look at existing research and next steps for new study. In: Firestone WA, Riehl C, eds. *A New Agenda for Research in Educational Leadership.* New York, NY: Teachers College Press;2005:138-155.
29. Knowles MS. *The Modern Practice of Adult Education: From Pedagogy to Andragogy.* New York, NY: Cambridge Books; 1980.
30. Merriam SB, Cafferella RS, Baumgartner, LM. *Learning in Adulthood: A Comprehensive Guide.* 3rd ed. San Francisco, CA: Jossey-Bass; 2007
31. Wood FH, Thompson SR. Guidelines for better staff development. *Educ Leadership*. 1980;37:374-378.
32. Dillman DA. *Mail and Internet Surveys: The Tailored Design Method 2007 Update with New Internet, Visual, and Mixed-Mode Guide.* 2nd ed. Hoboken, NJ: John Wiley & Sons; 2007.
33. Tourangeau R, Rips LJ, Rasinski K. *The Psychology of Survey Response.* New York, NY: Cambridge University Press; 2000
34. Belson WA. *Validity in Survey Research.* Brookfield, VT: Gower Publishing; 1986.
35. Bassili JN, Scott BS. Response latency as a signal to question problems in survey research. *Public Opinion Q.* 1996;60:390-399.
36. Schaeffer NC, Presser S. The science of asking questions. *Ann Rev Sociol.* 2003;29:65-88.
37. Kelley C, Kimball S, Clifford M, Dikkers S, Modeste M. Design to engage: features of the CALL Formative Feedback System that promote leadership for learning in middle and high schools. Paper presented at: the annual meeting of the American Educational Research Association; April 2012; Vancouver, Canada.