

CONTENT VALIDITY AS A WINDOW INTO LEADERSHIP PRACTICE

Mark Blitz
The University of Wisconsin-Madison
Educational Leadership and Policy Analysis
Doctoral Candidate
mblitz@wisc.edu

Matthew Clifford, PhD
Senior Research Associate
American Institutes for Research
Mclifford@air.org

The research reported in this chapter was supported by the US Department of Education Institute of Education Sciences (Award R305A090265) and by the Wisconsin Center for Education Research, School of Education, University of Wisconsin–Madison. Any opinions, findings, or conclusions expressed in this chapter are those of the authors and do not necessarily reflect the views of the funding agencies, WCER, or cooperating institutions.

Please do not cite without permission.

Abstract

The CALL (Comprehensive Assessment of Leadership for Learning) survey will offer school leaders opportunities to assess the school's leadership practice through research-supported domains comprised of leadership tasks. The CALL research team's work to validate the content of the survey involved eliciting feedback from practitioner groups who scrutinized the survey's items. The research team compared the practitioner's feedback to the existing survey items. The subsequent changes made to the survey reflect not only the constructive iterative process of content validation and survey construction, but also the nature of leadership practice and the potential gaps between it and research. The process of content validation with CALL is aligned with design experimentation, as is discussed in this paper. The authors then describe the process of design experimentation including the formation and work of the practitioner groups. By examining terminology, definition choice, and relevancy to practice, the CALL research team gained insight into the reality of leadership practice and how it has informed the construction of the CALL survey. The paper concludes with a discussion of findings and implications for researchers and practitioners.

Introduction

Design research is a form of naturalistic inquiry that provides mechanisms for building and testing theory on how students and adults learn, and how social systems are organized and change through an iterative process of creating and testing innovations intended to improve local conditions (Edelson, 2002). During design research, innovations (e.g., curriculum, assessments) are constructed through systematic testing to ensure their effectiveness and practicality in the context of practice (Brown, 1992). Through the design process, researchers gain an understanding of practice, as it occurs in everyday life, and this understanding opens researcher access to multiple, interconnected variables that may not be available in laboratory settings (Kelly, 2004).

In this paper, we describe how CALL development involves design research, and discusses results of our first design research effort, which focuses on content and construct validation of a formative assessment of school leadership systems. While the work of content validation informed the CALL survey, it also provided learning opportunities for the CALL research team to develop insight into leadership practice as compared to the theory of action espoused through the survey. This paper focuses on the findings from the content validation work.

This first design research effort engaged 20 educational leaders from multiple school districts in co-constructing CALL survey items, and this work ran in parallel with focus group studies and Rausch modeling of version one of the CALL instrument. Like many design research studies, CALL's design research work compliments traditional psychometric validation of CALL (see Condon, Milanowski & Clifford, 2010) by providing thick description of constructs and reasons for item development. The design

research also provides insight into the nature of school leadership, because it surfaces tensions and concerns in formative assessment of leaders.

Overview of Design Research

Lewin (1946) is recognized as bringing design research to the social sciences, and articulated the dual social benefit of combining action (i.e., design) with systematic study of effects. On the one hand, social innovations (e.g., social programs, curricula, assessments), Lewin said, that are designed through "... a spiral of steps, each of which is composed of a circle of planning, action, and fact-finding about the use of action..." are more likely attuned to influence thinking and behavior of individuals in a context. Additionally, design research opened access to daily practice and the change process, which Lewin believed, would heighten researcher understanding: "If you want to understand something, try to change it."

Nearly fifty years after Lewin's initial work, Brown (1992) and Collins (1992) reintroduced design research as an approach to studying the learning in context through the design of educational innovations. Since this re-introduction, researchers have sought to define and legitimate design research as a rigorous approach to knowledge development (Bell, 2002; McCanliss, Kalchman & Bryant, 2002). The Design-Based Research Collective (2002) and Barab & Squire (2004) provide the following precepts, which define design research for the field. In design research:

- Theory development and intervention design occur in learning environments and/or through engagement with practitioners;
- Research and design occur in a continuous cycle;

- Research accounts must clearly document processes and outcomes, both theoretical outcomes and impact improvements.

These general precepts define design research as a methodological approach for careful conduct of design research studies (Cobb et al, 2004). These precepts have raised concern for researcher objectivity and study validity that require researchers to engage with the field in remarkably different ways than standard approaches to inquiry (Barab & Kirschner, 2001).

The application of design research has contributed to our understanding of educational issues, and the explanatory power of other studies. Researchers have noted that design research can be fruitful for theory generation, because it occurs in situ and examines learning environments as they are lived (Cobb, McClain, de Silva Lamberg & Dean, 2003), improving the clarity of causal links between interventions and impact (Levin & O'Donnell, 1999), or raising challenges to existing theory about learning (Brown, 2002) and instructional change (Barab & Luehmann, 2004).

Examples of design research now dot the research literature, particularly in the field of the learning sciences. Available studies have tended to focus on explaining teaching and learning contexts by focusing on teacher, student, and designer interactions with software, curriculum, or other artifacts. And design research studies have tended to employ qualitative methods to systematically document teaching and learning. For example, Bouillon (2005), Brown (2002), Barab (2004) and Fishman, Marx, Blumenfeld, Krajcik and Saloway (2004) each employ design research as a method of curriculum development and a means of challenging existing theory on factors associated with curriculum adaptation. Sandoval (2003) also uses curriculum development as an occasion

for increasing understanding of student argumentation, and Joseph (2002) uses design research as a method for defining new approaches to curriculum design.

While design research practice continues to evolve, the approach has received some criticism as lacking methodological rigor, or standing as a distinct methodology at all (Kelly, 2004). In part, methodological ambiguity stems from the novelty of approach and the application of techniques associated from various research paradigms (The Design-Research Collaborative, 2003). Many design research studies, for example, employ case study or multi-case study to explore and explain phenomena, and attend to requisite techniques and standards for conducting these types of studies. Discerning design research from other forms of inquiry, and providing clarity on appropriate approaches to conducting design research with rigor are clearly necessary (Barab & Kirschner, 2001; Collins, Joseph & Bielaczyc, 2004; Kelly 2004).

These lines of research and design have been fruitful in improving our understanding of teaching and learning environments, providing quality interventions to the field, and defining a replicable design process for educational materials development. The majority of design research studies have, however, been applied to questions about teaching and learning, rather than school leadership. Applying design research to educational leadership makes sense, given that leadership aims to improve teaching and learning conditions through the design of policies, programs, and other artifacts (Spillane, Halverson & Diamond, 2004), and tools development (e.g., teacher performance assessment, instructional quality monitoring systems, data monitoring).

Design Research in CALL

The CALL project has been funded by the U.S. Department of Education, Institutes for Education Sciences to develop a formative feedback system that school leaders can use to identify areas of improvement and chart progress. As Halverson & Dikkers (2010) and Kelley, Kimball & Clifford (2010) describe, CALL takes a systems perspective on school leadership, which means that CALL is intended to measure and provide feedback on the quality of a school leadership system (as opposed to the performance of an individual leader). For example, CALL gathers data from multiple stakeholders on systems intended to support safety in schools. CALL's design builds on existing standards and summative assessment tools, and draws upon Halverson and Kelley & Shaw's practical experiences in leading criterion-based, formative assessment conversations with district- and school-level leaders. Once fully developed, CALL will be delivered as a Web-based system.

The CALL development plan should result in a validated, reliable, and practical formative assessment system that represents a marked improvement in the quality and quantity of data provided to school leaders. Validated and reliable formative assessment instruments of school leadership are essential because such assessments are intended to result in actionable knowledge that have consequences for school policies and programs (for discussion, see Condon & Clifford, 2009). Practicality is important to ensure CALL's use in the field. Assessments, curricula, or other tools developed for educators are often not taken up because they do not represent a marked improvement to current practice or are impractical given restrictions in practice (Rogers, 1995; Ball & Cohen, 1999). Toward these ends, the design process will demonstrate that self-assessors can and do use the CALL assessment that are based on actual features of the school and not

simply on the assessor's idiosyncratic interpretation of the evidence and the rating scales, and (2) that the CALL instrument is related to the desired leadership outcomes, such as enhanced leadership processes, school climate, and student learning.

As outlined in the proposal, CALL will be designed through design research. CALL development is organized into two phases. Phase 1 (years 1 and 2) is the developmental phase, when the CALL survey, feedback mechanism and Web-interface are developed. Phase two (years 3 and 4) focuses on validating CALL and stabilizing the system for scaled administration. We are currently in Phase 1, and are continuing to refine the CALL survey and online interface.

During Phase 1, CALL designers and researchers will increase their engagement with educational contexts of use and will refine the survey, feedback mechanism, and interface. In Year 1, CALL design researchers engaged 20 middle and high school leaders (hereafter, the "practitioner group") in validating constructs, co-constructing items, and designing the on-line survey. All 20 middle and high school leaders were publicly recognized for their competence (e.g., awards for excellence of service, participation on state-level leadership committees), and had a strong working knowledge of leadership research. Over the course of several months, practitioners were brought together to discuss CALL design, and in the process they explained design choices. The explanations of design choices provided insight into school leadership practices. This paper focuses on learning arising from these interactions.

While the practitioner group met with CALL designers, approximately 80 other practitioners completed CALL version 1.0, and participated in focus groups about their experiences. This work occurred in parallel to the practitioner group's work, and learning

was shared across research teams in Summer 2010. Survey data were used for Rausch modeling (see Condon, Milanowski & Clifford, 2010), and focus group data triangulated practitioner group findings.

In Year 2, the CALL version 1.0 will be fully developed and ready for pilot testing in the “messy” situation of school practice. CALL researchers will develop case studies of practice and will continue to conduct validation testing to inform production of CALL version 2.0. Case study will provide detailed information about CALL use and the consequences of its use in selected schools.

In Phase 2 (years 3 & 4), a stable CALL platform will be implemented in three medium-sized public school systems. Phase two studies involve validity (concurrent, validation, and consequential) and reliability analyses will be conducted. The approach to quantitative analyses will generally be to collect validity evidence within each district, treating such evidence as the result of a single validity study. Then, results will be combined across districts to get a more reliable estimate of the relationship using meta-analytic procedures. The approach will accommodate the use of the different alternative indicators of school or leader performance (e.g., student achievement tests) that are likely to be used by the participating districts. Because we will be working with four districts in this effort, we expect to have data from a fairly large sample of schools and school leadership teams.

Methods

This section provides a detailed overview of CALL practitioner group operations, and describes techniques for surfacing and validating CALL design issues and our

understanding of the use of CALL as a formative assessment in schools. In addition to the design considerations, this inquiry answers the following question: How does content validation through design experimentation inform our understanding of educational leadership and help to connect research and practice?

The initial task to set up a design experiment for content validity was to transform the set of rubrics into survey form. The rubrics contained five domains, and each domain contained four to five tasks. Within each task, each subtask provided three narrative descriptions of leadership practice that *Needs Attention*, is *Proficient*, and is *Exemplary*. Within each of those narrative descriptions, the research team referred to each individual content strand as an item. Figure 1 provides an example of a task within a domain, and the items within those tasks.

By looking at the corresponding items within each evaluative subtask box within the rubrics, the research team constructed questions based on the items across the subtasks that shared a common construct. As a result, based on the number of tasks and items within those tasks, each domain contained

Domain One: Focus on Learning			
Component Tasks	Needs attention	Proficient	Exemplary
1.1 Maintaining a school-wide focus on learning	School leaders have not engaged the school community and staff in collaborative conversations about student learning linked to teaching. The school does not have a clear, shared vision for learning; if there is a vision statement, it is regarded by community members as irrelevant to the daily practices of teaching and learning. Leaders have not established a common language for instruction. School leaders rarely arrange meetings to discuss either student achievement data or concrete examples of instructional practice with teachers	School leaders have engaged the school community and staff in conversations about student learning that serve as the foundation of a shared vision. The school has a collaboratively developed vision statement reflects the actual thinking and practice of teachers. Leaders have made an effort to establish a common language for instruction, but all teachers do not actively use the shared language to discuss their work. School leaders arrange meetings to discuss student achievement data at least twice per year and examples of instructional practice with teachers at least once per year.	School leaders regularly engage the school community and staff in on-going conversations, 4-6 times per year, that serve as the foundation of a collective understanding of student learning. The school has collaboratively developed, and annually revisits, a vision statement of learning that reflects the actual thinking and practices of teachers. Leaders have established a common language for instruction that all teachers actively use to discuss their work. School leaders arrange and conduct monthly meetings to discuss either student achievement data or concrete examples of instructional practice with teachers.

approximately 20 questions. With the rubrics now in survey form, and after some editing for clarity, the tool for experimentation was ready for use by the practitioner groups and participating focus groups.

For the UW-Madison team, the goal for the recruitment of practitioners was to pull from middle schools and high schools, as well as the various leadership positions (both formal and informal) within those schools. As a result, the middle school practitioner group consisted of two principals, one 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 member of an organization for school administrators. All practitioners were drawn from different schools, although the schools were located within the same region if not the same district.

The practitioner groups were divided into middle school and high school groups, and met seven times over the course of four months. For the first meeting, the group members met together were given the survey in its entirety and were asked to take the survey, recording any preliminary observations. For each subsequent meeting, the middle school and high school groups met separately in order to capture how a school

setting factors into perspective on practice, with the research team taking the assumption that the culture and goals of each school type would contribute to the practitioners' perspectives. In each meeting, the practitioners examined a specific domain closely.

They were given the following instructions:

Take section [domain number] of the assessment (revised version). Identify 5 questions that you think we need to work on the most. (We'll be talking about all of the questions, but this will help us prioritize the discussion.) Please keep the following checklist in mind as you reflect and take notes on the section:

- Language
- Clarity (Do any terms need to be defined?)
- Appropriateness to school level
- Importance of the question (is there anything left out? Is the question missing the mark?)
- Who should answer this question? Should there be a different form of the question for subgroups in the school?
- Should the question be in a different format?

During each meeting, the practitioner groups would scrutinize each section, or domain, of the survey and identify three to five survey questions they found most problematic and needing review. The research team would record the groups' feedback directly on the survey document through "tracked changes," and by taking notes and identifying quotes. Through the groups' feedback, the research team would identify the questions that needed the most attention and begin to discuss each survey item, spending

the most time on the most identified questions. As a group, the practitioners would provide their feedback on the specific questions mostly based on terminology, relevance to practice, and universal definitions of roles and labels.

The work of each meeting took approximately two hours to complete. After each meeting, the research teams would convene and share the practitioner feedback. In examining the feedback and making decisions concerning the purpose and direction of the survey, the research team would make the necessary revisions to the survey. The process of doing so allowed the research teams to gain insight into the practitioner realm, and consequently worked to develop the survey to more accurately reflect current school leadership practice. The discussions, which brought to light gaps between the rubrics and the practitioner groups' perspective, were tools used in revising the survey, but also offered general insight into practice.

Findings

The practitioner group feedback provided invaluable data to inform the development of the CALL survey. Furthermore, during this process, the research team gained insight into leadership practice and the realities of day-to-day tasks taken on by leadership teams. This section provides examples of the evolution of certain items within the survey and engine behind the significant iterations. Additionally, this section provides candid practitioner feedback in general that speaks to and supports the changes made in the survey.

Example # 1 (Relevance):

Domain three, task three focuses on *Socially Distributed Leadership*: the practice of involving teachers and faculty in leadership tasks and decision-making. Based on the items within the subtasks, the research team created the following survey stem and responses:

In our school, informal leaders (e.g. people who do not have administrative or other formally recognized leadership position):

- a. Often act to thwart or undermine the agenda of formal leaders for improving teaching and learning.*
- b. Passively support the agenda of formal leaders for improving teaching and learning.*
- c. Work with formal leaders to create the agenda for improving teaching and learning; constructive criticism “allowed”.*

This question attracted a great deal of scrutiny from the practitioners. One group member discussed the rationale behind the “thwarting” of the leadership agenda. On the surface, to thwart connotes an act of disrupting a positive initiative. However, this is assuming the leadership is effective, thus making the act of thwarting counterproductive. If the leadership is widely recognized as ineffective, the thwarting may be justified and perhaps even necessary and encouraged by the faculty. Conversely, a principal in the practitioner group discussed the equally harmful *inaction* of the part of informal leaders: those that passively comply without providing input, whether their input is elicited or not. The research team was faced with a decision with how to possibly reconstruct the question. Paramount in this process was maintaining fidelity to the rubrics as a whole

and what data the subsequent survey would elicit. Therefore, combined with the other aspects of the survey, the leadership practice could be measured and analyzed, thus placing different meaning to the “thwarting” issue. For example, if a particular school were to reveal a culture of closed-door leadership practice and top-down management practice, an act of thwarting the leadership practice would further support this culture of ineffective practice. Furthermore, along with editing the original responses, the research team inserted a fourth option that captured the practitioner groups’ feedback about compliance and overall inactivity: (informal leaders): *Are typically not engaged with the instructional agenda of formal leaders.* This option, capturing a culture of indifference, would further reveal a school’s lack of collaborative leadership practice.

Example # 2 (Relevance):

Domain two, task one focuses on the *formative evaluation of student learning*: how schools use data to inform their practice and assess student learning. A question within this task concerns the impact of standardized tests and how classroom teaching practice and student performance within the class predicts student achievement according to these external summative evaluation tools. The original stem and responses read as follows:

The formative assessment program in our school:

- a. *Does not exist.*
- b. *Exists but inaccurately predicts student performance on standardized tests.*
- c. *Allows leaders and teachers to accurately predict student*

performance on standardized tests.

d. Allows leaders and teachers to accurately predict student performance on standardized tests, and informs the process of tailoring differentiated instruction to students.

The original question assumes an impact of standardized tests. Given the current education policy climate, many schools have been likely focusing on how to meet standards and score well on state standardized tests. Therefore, this question focuses on the formative assessment programs and how they are used in concert with standardized tests. In choosing responses *a* or *b*, it is assumed that the school does not correctly prioritize assessment of student learning. However, a middle school principal in the practitioner group questioned the practice of focusing on predicting performance on standardized tests, and instead took the stance that formative assessment should be predictive of student learning. This difference in perspective informed the construct of the survey question, but it also forced the research team to look more deeply into the content of the survey questions and how user demographics would impact and possibly change the meaning and level of item significance. For some schools, the notion of student learning and standardized test performance may be considered the same. However, according to these test scores, these schools are likely struggling, and therefore their formal assessment programs are designed to focus on improving standardized test scores. The CALL research team found that the practitioner group members were leaders and members of relatively successful schools that did not need to spend time and energy focusing on the standardized tests directly. As a result, the survey question content

focused more specifically on *student learning* rather than *student achievement* as dictated by standardized testing. This is an important distinction that is not just applicable to the construction of the survey, but to educational policy and theory as well, which is a desired outcome of design experimentation.

Example # 3 (Terminology):

In creating an online survey for self-assessment, usability is critical in that users' cognitive energy needs to be spent reflecting on practice rather than deciphering the intent and meaning of the survey questions. Therefore, the terminology used within the survey needs to be fluid and somewhat vague, since the more specific a term, the more likely that it is not used or valued within a school. For example, in the following question in domain four, which focuses on *Acquiring and Allocating Resources*, the survey seeks to examine teacher incentives, something that can come in various forms:

Regarding teacher incentive and performance rewards:

- a. There are few incentives available to reward teachers for excellent performance.*
- b. Leaders find ways to reward individual teachers for excellent performance.*
- c. Leaders have developed an incentive system to reward teachers for collaborative work that results in documented progress toward school-wide learning goals.*

The practitioner groups responded to two terms specifically within this question: *incentives* and *excellent performance*. For *incentives*, the group members speculated on the forms of incentives. For some, depending on the district and/or type of school, incentives may be monetary rewards, tangible rewards, or any method of acknowledging success. One group member considered incentives to mean “public acknowledgement or recognition” whether within the school community or the surrounding community. Furthermore, another practitioner group member questioned the optimal response to the question: the idea of an *incentive system*, in which incentives are consistent and balanced. Option *c* introduces the idea of rewarding teachers for collaboration, thus making the claim that collaboration is integral in enhancing teaching and learning, an aspect of the research team’s theory of action that resonates throughout the survey. However, the group members were curious if the idea of an incentive system the way it is currently portrayed in the survey question was aligned with research. As a result, the research team edited the question in removing the term *system* while maintaining the idea of collaboration. The main construct of the question, however, the dispensing of incentives, was maintained in the survey. By not specifying the type of incentives, the research team allowed for the survey user’s perception and experience with the term *incentive* to contribute to their response. Whatever the incentive may be, the focus on the use of them; whether they are effective or not may be gathered from the response to this question in concert with other related survey items.

The other somewhat problematic term within this survey question, *excellent performance*, brings issues to light beyond the scope of this survey, which is a benefit of design experimentation. One practitioner group member offered the following advice

and questions: “You may need to define excellent performance. Do you mean student achievement? Hard work? New ideas?” Excellence in teaching may take on various forms, as this response indicates. This may depend on the school, and how the school community defines success. If teaching excellence is measured by student achievement, as the group member asked, then some schools would find this problematic while others find this to be sensible and commonplace. In the current educational policy climate, the term *student achievement* is often thought of in terms of student test scores on standardized tests. However, while some schools directly focus their teaching strategies and curriculum on raising these scores, other schools are not as concerned with this data. As a result, the research team was forced to reconsider the intent of the question and to find terminology that reflected the theory of action of the survey work. While teachers can excel in different aspects of the teaching position, the primary goal in their work should be to enhance *student learning*. This term is broad enough to include the various ways to measure student learning, whether it is local system or a largely universal and quantifiable system. As a result, the current iteration of the survey does not mention *excellent performance*, but rather each response item maintains a focus on *student learning*. Therefore, school type would not be as significant in ascertaining teacher effectiveness toward increasing student learning.

Example #4 (Definition Choice):

Terms that are seemingly commonplace in a school, district, or region may not be as commonplace in more distant regions, based on culture, philosophy, and/or perspective. As a feature of the online CALL survey, the survey offers highlighted terms

in hypertext that provide a “mouse-over” definition and examples if necessary. These definitions include but are certainly not limited to: formative feedback, support services, strategy, community, paraprofessionals, and instructional coaches. It may be that these terms were overly vague, and it may also be that schools may incorporate these ideas but with different terms. Therefore, offering definitions was thought to be critical. In focusing on appropriate use of definitions, one survey question in particular gave the research team an opportunity to reflect more critically on the intent of the question and the larger ramifications on school culture and practice. The question, in domain five, read as follows:

In regards to student identification, student support services at our school:

- a. Often mis-categorizes or over-identifies students.*
- b. Inconsistently identifies students.*
- c. Consistently identifies students.*
- d. Consistently and effectively identifies students.*

Initially, the practitioner group called for defining *student identification* and *student support services*. While this task would be relatively simple, it in turn led to a discussion of other pressing issues associated with this question such as the variety of categories under which to identify students, the types of services offered after identification, and the impact or effectiveness of these services. Clearly, this would be too much to address in one survey question, and as was the case with other questions, the research team created additional questions in order to avoid one question containing only one construct, which would otherwise make answering that type of question difficult and often not possible. Based on the practitioner group feedback, the research team

understood that some categories may be identified differently. A workable solution for this common type of problem was to create a table with multiple categories on the y axis and quantifiers or qualifiers across the x axis. The research group identified four categories with which to identify students and offered the survey user a range of *over-identified to accurately identified* with the options *under-identified* and *mis-identified* in the middle (note: for scoring purposes, the first three qualifiers are given weights of 1 or 2, while *accurately identified* is weighted as 5).

The construction of this table addressed the issue of needing to effectively define *student identification*. However, other issues arose from the discussion of this survey question. In order to examine the services provided, the research team has added the following question the current CALL survey iteration:

At our school, support services improve learning for identified students:

- a. *Rarely.*
- b. *Successfully for some students.*
- c. *Successfully for most students.*

The research moved this subtask a step further in identifying a specific type of strategy to address student support services: Response to Intervention (RtI). The research team constructed a question focusing directly on RtI, believing this method to work effectively, and yet it can also take on various forms. The team discussed the universal application and usage of RTI, and was faced with a decision on how to define the term in a mouse-over. The research team made an important decision that presents implications in assessing leadership practice and school culture: the survey question is currently posed as *Describe your school's use of **Response to Intervention** (RTI)*. The first response

option (to be scored with a weight of 1) is *What is RTI?* By offering this response, the survey can determine a school's (and district's) alignment with research and "best practices." While other areas of the survey offer mouse-over definitions of terms in order for the user to accurately reflect on practice, this unique example does not offer this tool, with the rationale that a lack of understanding of this particular term is data in itself. Therefore, the decisions on what to define and what not to define have been purposeful and take into account usability, survey output data, and aligning with research. What might appear to be an otherwise trivial action in term definition, led to additional survey questions and methods of eliciting important information through survey question and response construction.

Lastly, in reviewing the survey on a holistic level, the practitioner group members found questions to be problematic that required the users to offer their perception on the practices and views of other school community members. For example, an original question asked the users to estimate the level of involvement the school principal has in professional organizations and professional development opportunities. One focus group member responded: "If you think about leading by example, yeah, I would like to know that my principal is a member of some professional organizations, that she's not just getting everything off Wikipedia...but you could easily communicate that..." The focus group member was interrupted (at first he drew laughter, and then someone else started a new conversation thread). Through conjecture, one might ascertain that he felt this information could be gleaned in some way other than by direct questioning, so as not to put the faculty in a position to guess an answer about the principal's latent practice. As a result, the research team has closely examined and discussed "perception" questions,

identifying which questions would offer fruitful outputs and which ones would put a user in a difficult position to provide an accurate response. Some “perception” questions were proven to be valuable, and could offer a window into the culture of the school.

Moreover, the research team is in the process of developing a social network component of the survey that would value perception questions that target expertise, informal leadership, and consistency of perspectives of school community.

To be sure, the practitioner groups identified other problematic areas during this first iteration of the CALL survey, yet for the purpose of this article, it is not worthwhile to cover all aspects of their feedback. Rather, by focusing specifically on relevance to practice, terminology, and definitions, the research team has not only worked to achieve its goal of refining the survey for enhanced usability and validating content, but the team also widened the window into school leadership practice. This encouraging byproduct of content validity is discussed in the next section, where we also return to the application of design experimentation and the work to bridge the gap between research and practice.

Discussion and Conclusion

The relationship between research and practice is has the potential to be strengthened through design experimentation work, in which (in this case) practitioners grapple with actual cases and their responses are recorded as data for reporting in research. Still, consumers of the research findings may still need to reconcile the differences between their own perception of best practices and best practices put forth by research. In effect, this reconciliation is what often transpired during the content validity process, and the research team would in turn play the role of architect in determining the

most appropriate components for the CALL survey. The original rubrics that served as the foundation for the survey were based on research itself, rather than having been materialized from conjecture or opinion. Therefore, when the practitioner groups responded to the survey and questioned the relevance of certain aspects of the survey, their responses were matched against previous research findings. While the research team contemplated the differences in data, the process also afforded the research team to gain a tacit understanding into practice and the role context plays in school leadership. This opportunity was apparent on the occasion of the practitioner team responding to the survey's assertion that state mandated standardized test scores should be taken into account when evaluating certain aspects of teaching and learning. While urban schools, which have been more likely to ascribe to test scores as indicators of student learning, other schools located in more suburban areas do not spend as much time and energy on test preparation or on the significance of the test scores. It was with mostly successful schools that the practitioner group members mostly associated. Therefore, the research team learned how the context and profile of the school contributed to the perception of relevance to practice. While this conclusion aided in the revision of the survey, it also brings to light a point of interest for researchers and practitioners in this field, which is an affordance of design experimentation.

The byproducts gained through the content validity process in creating the CALL survey offer further support to the value of design research. What the research team has gleaned from the content validity process has sparked interesting and important conversations on teaching and leadership practice, and yet the findings from those conversations are summarily discussed in this paper to demonstrate the importance of

design experimentation as a means to validate content. However, when taking the opportunity to explore ideas of best practice, practitioners may directly apply findings to their own practice, thus building more than a bridge between research and practice, but rather overlapping the two areas. In this case of developing an instructional leadership self-assessment survey, the research team is comparing various perspectives on and sources of evidence of effective practice. The end result is the development of a useful tool for practitioners. In a case where practitioners take on the same process, and supply their own practical experience for comparison, the same findings used by the CALL research team could be put to use by practitioners engaged in action research.

Overall, the content validity process offers a window into the current realm of leadership practice in which terminology, definition choice and relevance is informed by contextual factors including policy, demographics, and both individual and shared notions of “best practice.” Current school leaders apply their experience to their perception of what practical and relevant. The challenge that lies ahead for the CALL research team, as it continues to develop this instructional leadership tool, is to identify a presentation of survey content that cuts across various contextual factors of school leadership so that school leaders will find the survey to be informative and beneficial regardless of region, political climate, and demographics. The work of the CALL research team is to continue to strengthen its theory of action, and the content validity informs that theory through the work of revising and honing the survey tool.

Bibliography

- Ball, D. & Cohen, D. (1996). Reform by the book: What is or might be the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, 25(9), 6-14. Barab 2004
- Barab, S.A. & Kirschner, D. (2001). Guest editors' introduction: rethinking methodology in the learning sciences. *The Journal of The Learning Sciences*, vol. 10 (1&2), pp. 5-15.
- Barab, S. & Luehmann, A. (2003). Building sustainable science curriculum: Acknowledging and accommodating local adaptation. *Science Education* 87 (4), 454-467.
- Barab, S. & Squire, K. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences*, 13(1), 1-14.
- Bell, P. (2004). On the theoretical breadth of design-based research in education. *Educational Psychologist*, 39(4), 243-253.
- Bouillon, L. & Gomez, L. (2001). The case for considering cultural entailments and genres of attachment in the design of educational technologies. *Smart machines in education: The coming revolution in educational technology*. Cambridge: MIT Press.
- Brown, A.L. (1992). Design experiments: theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of The Learning Sciences*, vol. 2. no, 2, pp. 141-178.
- Brown, M. (2002). Toward a theory of curriculum design and use. Paper presented at the Annual Meeting of the American Educational Research Association.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R. & Schauble, L. (2004). Design experiments in educational research. *Educational Researcher*, 32(1), 9-13.
- Cobb, P., McClain, K., de Silva Lamberg, T. & Dean, C. (2003). Situating teachers' instructional practices in the institutional setting of the school and district. *Educational Researcher*, 32(6), 13-24.
- Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.) *New Directions in Educational Technology* (15-20). Berlin: Springer-Verlag.
- Collins, A., Joseph, D. & Bielaczyc, K. (2004). Design researcher: Theoretical and methodological issues. *Journal of the Learning Sciences*, 13(1), 15-42.
- Condon, C. & Clifford, M. (2009). Measuring principal performance: How rigorous are publicly available principal performance assessment instruments? Naperville, IL: Learning Point Associates

Condon Milanowski Clifford 2010

Design-Based Research Collaborative (2002). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5-8.

Edelson, D. (2002). Design research: What we learn when we engage in design. *Journal of the Learning Sciences*, 11(1), 105-121.

Fishman, B., Marx, R., Blumenfeld, P., Krajcik, J. & Saloway, E. (2004). Creating a framework for research on systemic technology innovations. *Journal of the Learning Sciences*, 13(1), 43-76.

Kelly, A. (2004). Design research in education: Yes, but is it methodological? *Journal of the Learning Sciences*, 13(1), 115-128.

Kelley Kimball Clifford 2010

Lewin, K. (1946) Action research and minority problems. *Journal of Sociology*, 2(4), 34-46.

Levin, J. & O'Donnell, A. (1999). What to do about educational research's credibility gaps? *Journal of Cognitive Education and Psychology*, 1(2), 201-256.

McCanliss, B., Kalchman, M. & Bryant, P. (2002). Design experiments and laboratory approaches to learning: Steps toward collaborative exchange. *Educational Researcher*, 32(1), 14-16.

Rogers, E. (1983). *Diffusion of Innovations*. New York: Free Press.