Measuring School Capacity for Continuous Improvement

Elizabeth Leisy Stosich
Harvard Graduate School of Education

Suggested citation: Elizabeth Leisy Stosich (in press). Measuring School Capacity for Improvement: Piloting the Internal Coherence Survey. Accepted for publication in A. Bowers, B. Barnett, & A. Shoho (Eds.), Using Data in Schools to Inform Leadership and Decision Making. Charlotte, NC: Information Age Publishing.

Abstract

Although there is substantial data on student performance available to inform decision making in schools, school leaders have limited information about the leadership practices and organizational conditions they can establish to support continuous improvement. The Internal Coherence (IC) Survey is a tool for bringing information on research-based factors related to schoolwide improvement to bear on data driven decision making in schools (Elmore & Forman, 2011). The usefulness of survey tools depends on their ability to measure research-based constructs and their appropriateness for a given context and purpose. In this paper, I describe a process for bringing together researchers and practitioners as partners in survey development. I conducted a series of studies, including expert review, cognitive interviews, and principal components analysis (N=9; N=6; N=2,550), to assess and improve the construct validity and reliability of the survey. I conclude that the revised Internal Coherence Survey is a powerful tool for providing school leaders with information about specific practices, processes, and beliefs they can foster to enhance school capacity for improving instruction and student learning.

Introduction

Recent accountability policies have heightened pressure for school leaders to use data to improve student achievement schoolwide (e.g. NCLB, 2001). Similarly, professional licensure standards call on school leaders to "collect and use data to identify goals, assess organization effectiveness, and promote organizational learning" (Council of Chief State School Officers, 2008, p.14). While accountability policies have led to substantial data about student performance (Colyvas, 2012), little data is collected on the organizational conditions that school leaders would need to change to foster continuous improvement (Anderson, Leithwood, & Strauss, 2010; Halverson, 2010). Continuous improvement requires school leaders to articulate a vision for improved instruction and student learning and to develop systems for collecting, analyzing, and acting upon information that supports them in realizing this vision (Fullan, 2007; Hawley & Sykes, 2007). This information must go beyond data on student achievement and include data on factors that are both related to enhanced student performance and can be influenced by school leaders (Leithwood, Patten, & Jantzi, 2010).

To address the need for information about the factors that school leaders can influence to support improvement, Elmore and Forman (2011) developed the Internal Coherence (IC) Framework and corresponding teacher survey. The IC Framework describes "a school's capacity to engage in deliberate improvements in instructional practice and student learning across classrooms, over time" (Elmore & Forman, 2011). The IC Framework and Survey draw on decades of research on school leadership and continuous improvement to identify specific leadership practices, organizational processes, and efficacy beliefs that school leaders can foster to enhance school capacity

for improvement (Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010; Garvin, Edmondson, & Gino, 2008; Goddard, 2002; Goddard, Goddard, & Tschannen-Moran, 2007; Ross & Gray, 2006; Tschannen-Moran & Hoy, 2001). School leaders can use the IC Survey to collect data about school capacity for improvement and identify specific leverage points that they can act on to bring about change (Mandinach & Gummer, 2013).

The usefulness of this tool depends on how well it measures research-based factors related to school improvement and can be used by school leaders and teachers to inform their improvement efforts. Thus, I brought together researchers and practitioners as partners in furthering the development of this tool for supporting data driven decision making. I conducted a series of studies to enhance the validity of the IC Survey, the degree to which it measures the theoretical constructs it is designed to measure, and the degree to which the survey items are clear and reliable measures of these constructs (Messick, 1995). This series of studies addressed four questions:

- *Do the scales exhibit content validity?*
- *Do respondents understand the meaning of each survey item?*
- *Are respondents able to answer appropriately, given the response anchors?*
- *What is the factor structure of each of the survey domains?*

The significance of this work lies in developing a valid and reliable tool for collecting data on school IC to inform continuous improvement efforts. I begin by describing the relationship between IC and continuous improvement. I discuss the need for advancing the development of the IC Survey, comparing it to existing surveys in the field and explaining the need for thorough pretesting to enhance the validity of the IC Survey. I present the findings from my collaborative survey development process, which

brought together researchers and practitioners to advance our understanding of how to assess IC in schools, and present preliminary analyses of the factor structure of the instrument. I conclude by describing the implications of my findings, which highlight the usefulness of the survey for researchers and supporting school leaders in using data to foster continuous improvement.

Background: IC and continuous improvement

The IC Framework and Survey bring together research on leadership practices that promote learning, organizational processes that support collaboration, and teacher efficacy beliefs that support improvements in instruction and student learning to form a developmental model designed to explain how schools support improve (see Figure 1). The goal of this model is to identify specific practices, processes, and beliefs that are related to improvement and describe them in actionable ways, making it possible for school leaders to use data about school IC to inform their continuous improvement efforts. School leaders play an important role in fostering improvement, but their influence on the core work of schools—teaching and learning—is largely indirect. Scholars suggest that school leaders influence teaching and learning indirectly by developing a clear vision for instruction, working collaboratively with teachers to realize this vision, and developing supportive conditions for teacher learning (Bryk et al., 2010; Goddard, Goddard, Kim, & Miller, in press; Hallinger & Heck, 1996; Leithwood, Seashore Louis, Anderson, & Wahlstrom, 2004; Robinson, Lloyd, & Rowe, 2008). The IC Framework and Survey build on this research to identify specific factors that school leaders can act on to foster continuous improvement.

The factors related to IC are organized into 3 domains: 1) Leadership Practices for Instructional Improvement, 2) Organizational Processes, and 3) Efficacy Beliefs. The connections among these domains are supported by recent quantitative research. Goddard and associates (in press) found that principals' instructional leadership was a significant, positive predictor of teacher collaboration, higher levels of teacher collaboration predicted stronger teacher collective efficacy beliefs, and these efficacy beliefs were a significant, positive predictor of student achievement. I describe each of the domains related to IC and their connections in detail below.

Leadership Practices for Instructional Improvement

School leaders can drive improvements in instruction and student performance by communicating a vision for teaching and learning, creating a safe environment for teachers to talk openly about instruction, and fostering opportunities for professional learning (Bryk et al., 2010; Hallinger & Heck, 1996; Heck & Hallinger, 2009; Leithwood et al., 2004; Robinson et al., 2008). When school leaders develop a compelling vision for teaching and learning and provide the support needed for teachers to realize this vision, teachers are more motivated to contribute to schoolwide improvement efforts (Thoonen, Sleegers, Oort, Peetsma, & Geijsel, 2011). Although school leaders play an important role in initiating improvement efforts, researchers find that school leaders are more likely to share leadership with teachers as their schools develop higher levels of capacity for improvement (Day et al., 2010; Hallinger & Heck, 2011). Schools with principals who work collaboratively with teachers toward shared improvement goals and support the teacher professional growth needed to meet these goals are associated with higher quality

instruction and higher levels of student achievement than schools with principals who do not engage in these practices (Leithwood & Mascall, 2008; Marks & Printy, 2003; Printy, Marks, & Bowers, 2009). The collaborative nature of these leadership practices demands a "psychologically safe" environment, one in which people acknowledge the dynamic nature of knowledge, encourage experimentation, and consider failure a part of the learning process (Edmondson, 2008). School leaders can actively cultivate a psychologically safe learning environment by acknowledging the limits of their own knowledge and encouraging teachers to speak openly about their instructional practice.

Organizational Processes

School leaders can establish organizational processes that foster teacher involvement in instructional decisions at the whole-school and team level. Whole-school processes can serve as opportunities for negotiating beliefs about effective instruction and aligning work with schoolwide improvement goals (Argyris & Schon, 1996). Team processes that focus teachers' work on the specific issues of instructional practice and student learning are also essential for improvement (Edmondson, 2002; Vescio, Ross, & Adams, 2008). In schools with high levels of IC, whole-school structures, such as faculty meetings, and team structures, such as grade-level or content meetings, can support the process of learning for improvement; however, in schools that lack coherence, processes for collaborative work may be disconnected from goals for improving teaching and learning.

Efficacy Beliefs

While school leaders cannot directly influence teachers' efficacy beliefs, they can create a psychologically safe environment for learning and put in place processes for collaboration. Teacher collaboration can encourage teachers to experiment with new approaches, reflect on their instruction, enhance teachers' efficacy beliefs, and, ultimately, lead to improved student performance (Goddard et al., in press; Goddard et al., 2007; Moolenar, Sleegers, & Daly, 2012; Thoonen et al., 2011). Teachers' individual and collective efficacy beliefs measure their perceptions of their ability and the ability of the faculty as a whole, respectively, to plan and execute effective instruction. Positive efficacy beliefs are fostered through experiences of mastery—trying a new instructional strategy and witnessing improvements in student learning—and encouragement from colleagues to experiment with new practices (Bandura, 1997). Teachers with high levels of individual efficacy are more likely to exert sufficient effort to improve (Stajkovic & Luthans, 1998), implement effective teaching practices, and foster higher levels of student achievement (Goddard, Hoy, & Hoy, 2004). While individual efficacy is an individual trait, collective teacher efficacy resides in the school, which makes it essential for continuous improvement (Bandura, 2000). Schools with high levels of collective efficacy have greater potential for meeting improvement goals because these beliefs establish shared expectations for success that make teachers more likely to experiment with new practices and persevere in the face of challenges (Goddard et al., in press; Goddard et al., 2004).

In sum, assessing current levels of IC and identifying the specific practices, processes, and beliefs related to IC provides school leaders with data about specific

factors that they can act on to develop school capacity for improvement. School leaders face a wide range of challenges in building capacity for continuous improvement. The IC Framework and Survey are tools for bringing together information on key factors related to improvement to inform decision making in schools.

Existing instruments: Why further the development of the IC Survey?

Research on the leadership practices, organizational processes, and efficacy beliefs that support school improvement, all factors related to school IC, can serve as data to inform decision making in schools. However, encouraging school leaders to draw on research in their day-to-day practice takes more than simply disseminating findings (Anderson et al., 2010). The IC Survey builds on existing surveys designed for research that measure leadership practices and organizational conditions related to continuous improvement (e.g. Bryk et al., 2010; Garvin et al., 2008; Goddard et al., in press; Leithwood et al., 2010) to develop a clinical tool for use in schools. Data from research-oriented surveys may be challenging to apply to decision making in schools if they measure broad constructs related to improvement. The IC Survey describes key factors related to improvement in actionable ways, so school leaders can apply this information directly to their practice. The survey can act a tool for bringing research-based understandings to bear on improvement efforts in schools (Coburn & Stein, 2010; Donovan, 2011).

The IC Survey provides information about the extent to which teachers perceive that key practices and processes that have been found to foster higher levels of teacher efficacy and improved student learning are in place in a school. School leaders can use

this information to inform decision making in their unique school context. Data from principal evaluation tools (e.g. Goldring, Porter, Murphy, Elliott, & Cravens, 2009; Kelley & Halverson, 2012) can also be used by school leaders to improve their practice; however, these surveys are evaluative. When used for accountability purposes, leadership surveys can operate as a tool for assessing leadership quality rather than encouraging organizational learning (Firestone & Gonzalez, 2007). The IC Survey is designed to provide diagnostic data on schoolwide capacity for continuous improvement that school leaders can use to inform decision making.

Developing research-based tools: Enhancing validity and reliability

The usefulness of the IC Survey depends on the extent to which it accurately and reliably measures factors related to school capacity for improvement. Although experts were consulted during the original survey development, no systematic analysis of the validity or reliability of the survey was conducted. Given the intended use of the survey by school leaders in making key decisions about improvement, an in-depth analysis of the survey instrument was essential. If school leaders use data from an invalid or unreliable tool to inform their improvement efforts, they may make decisions that negatively affect school performance (Porter et al., 2010). Construct validity is related to the degree to which the survey measures the concepts that it is designed to measure—including how it describes these concepts and how this is interpreted by survey participants—and reliability is related to the consistency with which it measures these concepts (Messick, 1995; Marsden & Wright, 2010). Developing valid and reliable surveys for measuring factors related to school improvement requires both thorough pretesting to develop sound

instruments and examination of the factor structure of the scales after administering the survey to participants (Gehlbach & Brinkworth, 2011).

I conducted a series of studies that evaluated the content validity, reliability, and structure of the twelve scales of the most recent version of the IC Survey, which consisted of three key analyses: expert validation, cognitive interviewing, and principal components analysis (PCA). Experts in research and practice provided insights into whether the items in each scale measured the construct they were intended to measure and whether they did so in a comprehensive manner (Dillman, Smyth, & Christianson, 2009; Gehlbach & Brinkworth, 2011). Cognitive interviews with survey participants, asking participants to "think out loud" as they answered the survey, revealed whether participants consistently interpreted survey items as researchers intended (Collins, 2003). Finally, I conducted PCA to examine whether the items in each scale appeared to measure the construct they were designed to measure. Together these analyses revealed information about the construct validity and reliability of the scales, highlighted areas for improvement, and provided evidence to support the use of the revised survey as a tool for assessing and developing school capacity for improvement.

Researchers have endeavored to understand organizational improvement by designing surveys to assess the leadership practices and organizational conditions associated with improved performance (e.g. Bryk et al., 2010; Garvin et al., 2008; Goddard et al., 2007); however, limited information is available about the validity of previous scales. The literature on measures of school improvement seldom emphasizes the survey development process. Instead, researchers often focus on testing psychometric properties of scales, with little emphasis on the process of establishing construct validity.

For example, Schechter (2008) provided detailed data from factor analysis but limited information from qualitative pretesting. Analyzing the factor structure is an important part of testing the relationships among items in a scale, but it cannot provide information about construct validity. As Light, Singer, and Willett (1990) remind us, "*Statistical analysis cannot retrospectively repair a flawed instrument* [original emphasis]" (165). To address this concern, I conducted thorough qualitative pretesting before conducting PCA with data from the pilot test of the survey, so I could minimize the influence of measurement error and maximize learning from the initial administration (Gehlbach & Brinkworth, 2011).

The validity of a survey instrument depends on not only the extent to which survey items measure the intended constructs but also the purpose for and context in which it is used (Messick, 1995). Thus, thorough pretesting of the survey instrument is not only a critical practice for assessing the validity of the instrument but also an ongoing process (Porter et al., 2010). When researchers and practitioners adopt surveys for a new purpose or context, they can consider assessing the validity and reliability using the approach I describe below.

Methods

In the following section, I provide evidence of the validity and reliability of the survey instrument by detailing my extensive, collaborative survey development process and then present preliminary data on the structure of the survey scales. In Study 1, I conducted an expert review with practitioners and researchers to examine the validity of the survey scales. In Study 2, I conducted cognitive interviews with six teachers to learn

whether the individual items were valid and reliable indicators of the respective scales. After modifying the survey based on data from expert review and cognitive interviews, I administered the IC Survey to a sample of 2,550 teachers and teaching assistants. Finally, I examined the structure of the scales by conducting PCA in Study 3. I describe the procedures and findings from these studies below.

Study 1: Expert Review

I conducted an expert review of the IC Survey to examine the degree to which the IC Survey scales measured the constructs they were intended to measure. I found that there were three threats to the validity of the survey scales: unclear focus on instruction and student learning, emphasis on outcomes rather than processes or practices, and multidimensionality. I provide a detailed description of these threats to validity and how I addressed them.

Participants

The sample for Study 1 consisted of nine experts in school improvement and organizational learning. I purposefully selected these experts based on their research expertise and/or practical experience. Four experts had developed surveys measuring constructs related to improvement that informed the development of the IC Survey. Three experts were district leaders with experience using the IC Survey to inform school improvement efforts.

Procedures

I administered a survey to experts to learn whether the items were a representative sample of the construct measured by each scale (McKenzie, Wood, Kotecki, Clark, & Brey, 1999). I provided a detailed definition of the construct that each scale measured and asked experts to rate the each item on a scale from 1-5, from "not at all" to "completely" representative or clear (Dillman et al., 2009; Rubio, Berg-Weger, Tebb, Lee, & Rauch, 2003). I asked about item representativeness to learn to what extent experts thought each item measured the construct. I asked about item clarity to learn whether the item wording was understandable. Complicated wording could lead to misinterpretations by participants, which could result in misresponse or nonresponse (Dillman et al., 2009). In addition, I included open-ended items for experts to provide suggestions for rewording and identify additional items to measure each concept.

Not all experts rated all items (Rubio et al., 2003); therefore, I analyzed the quantitative data together with expert comments for each scale to identify threats to validity. All items that were rated as "somewhat representative" or "somewhat clear" or lower by two or more experts and all items identified as unrepresentative or unclear in experts' comments were identified as possible threats to content validity. I analyzed these items collaboratively with members of the IC research team[1], including those who created the IC Survey to identify determine whether they should be revised or removed.

Findings and Discussion

I found that experts rated items relatively highly in terms of representativeness and clarity. However, by examining expert ratings and comments as a whole, I found that

[1] The IC research team members include Richard Elmore, Michelle Forman, and Candice Bocala.

there were three main threats to the content validity of the scales: unclear focus on instruction and student learning, emphasis on outcomes rather than processes or practices, and multidimensionality. In the following sections I offer examples that highlight these three threats to validity and explain how they have been addressed.

Focusing on instruction and student learning

The IC Survey was intended to serve as a tool for understanding school capacity for improving teaching and learning. However, experts found that the connection to teaching and learning was not always clear. For example, the Teachers' Collective Work scale was designed to measure the extent to which teachers are involved in collaborative processes related to improvement in instruction and student learning (Goddard et al. 2007). However, one expert commented that the items that asked about parent communication and student discipline policy were less representative of the construct and were only weakly correlated with the rest of the items in the scale in previous research (Goddard et al. 2007). I removed these two items to strengthen the scale's focus on instruction and student learning. I retained all other items in the scale since they asked about practices closely related to teaching and learning, such as selecting instructional methods.

Emphasizing processes and practices rather than outcomes

Each scale in the IC Survey was designed to measure practices, processes, or beliefs that could be fostered to enhance school improvement capacity. Experts found that some items did not reflect this emphasis on process over outcome. For instance, one

item asked about the consideration of student learning outcomes as part of teacher evaluations. This item focused on performance outcomes rather than encouraging ongoing monitoring and feedback for improvement. I replaced it with the following item: "The principal uses data from classroom observations to give teachers feedback about instruction." The goal of this revision was to better align the description of data use with the construct's focus on principal practices that support ongoing improvement in instruction and student learning.

Multidimensionality

Creating valid measures of each dimension related to IC requires that all indicators in each scale measure the intended construct. Experts raised concerns about whether scales measured more than one construct. For example, the Creation of a Learning Environment scale was intended to measure the extent to which principals create a supportive learning environment by fostering psychological safety (Edmondson, 2002). Although most of the items ask about psychological safety, as it has been measured in previous scales (Garvin et al., 2008; Higgins, Ishimaru, Holcombe, & Fowler, 2011), others focused on support. The items asking about support were included in the scale because the ability to seek support when struggling with instruction was seen as a condition of a psychological safety. However, two experts questioned whether the items related to support measured a different construct. Since this question could best be answered by empirical data from the pilot test, I retained the items and made note to examine their relationship to the rest of the items in the scale when conducting PCA.

Conclusion

As described above, evidence from expert review revealed three key changes that had to be made to improve the content validity of the IC Survey scales: 1) strengthen the focus on instruction and student learning, 2) maintain the focus on processes, practices, and beliefs related to improvement, and 3) address multidimensionality. After analyzing data from expert review, I revised each of the survey scales to better measure the underlying construct. These revisions served to enhance the validity of the scales, making the survey a more useful tool for school leaders and researchers.

Study 2: Cognitive Pretesting

If respondents do not interpret the items and response anchors as intended, the substantive validity of the instrument is threatened (Messick, 1995). Respondent misinterpretation of the items could diminish the reliability of the survey since all items in a scale were intended to measure the same construct (Marsden & Wright, 2010). To address these issues, I conducted cognitive interviews with teachers to determine whether respondents understood the meaning of each survey item and were able to respond appropriately, given the response anchors. I detected four threats to the validity of the scales: lack of clarity regarding the focus on instruction and student learning, people, and context and inappropriate response anchors. I describe these threats to validity and how I addressed them below.

Participants

The sample for Study 2 was taken from a population of elementary and K-8

teachers and teaching assistants in a large urban public school district in the Southern US. These teachers were selected because the district was going to use the IC Survey as part of an improvement initiative focused on elementary and K-8 schools. I used purposive selection to choose a sample of six elementary teachers from one school from across grade levels and content areas (Seidman, 2006). I interviewed four classroom teachers, each at a different grade level, a music teacher, and a special education teacher to examine whether the survey was understandable and applicable to teachers from a relatively wide range of subjects and specializations. All teachers invited chose to participate.

Procedures

I conducted cognitive interviews with teachers, asking each participant to "think aloud" while answering each survey item (Garcia, 2011). The purpose of cognitive interviewing was to learn whether respondents understood the meaning of each item, could answer appropriately given the response anchors, and were motivated to think carefully through their response choice (Barge & Gehlbach, 2012; Collins, 2003; Dillman et al., 2009). All interviews were transcribed and coded for emerging patterns and recurring problems. All items that two or more teachers found confusing or misinterpreted were revised or removed. Together, information from cognitive interviews and expert review was used to inform survey revisions for improved clarity and construct validity.

Findings and Discussion

In examining data from cognitive interviews, I detected four threats to validity. Participants were less likely to respond accurately when the focus on instruction and student learning, the people (i.e. teachers, administrators), or context (i.e. school, district) was unclear, or the response anchors were inappropriate. I identified 14 survey items that participants found difficult to understand, difficult to answer, or interpreted in a different way than intended by the survey designers. Response anchors were generally clear and motivating, with the exception of those that asked participants to assess the frequency with which they engage in specific actions. In the following sections, I present examples that illustrate these key findings and discuss survey revisions made based on these findings.

Unclear focus on instruction and student learning

When the focus on instruction and student learning was unclear, teachers found it difficult to respond to survey items. For example, three teachers found the following item confusing: "The principal at this school asks probing questions." One teacher was unsure about whether the question asked about all instances of questioning or was focused on a particular context. The item was designed to measure practices that foster public learning related to instruction and student learning. Thus, I reworded the item to ask specifically about probing questions about teaching and learning.

Unclear focus on people

Teachers also found it difficult to answer items when it was unclear whom the item referenced. For example, one item stated: "My professional development

experiences this year have included follow-up support from leaders as we implement what we have learned." One teacher was unsure about who was included in the term "leaders." This item was intended to measure the extent to which follow-up support was available to promote ongoing learning. However, the role of the people who provide this support was irrelevant; thus, I removed the phrase "from leaders" from the item.

Unclear context

As noted previously, the survey was intended to focus on factors that school leaders can influence. Two teachers commented that they were uncertain whether items that asked about their professional development (PD) experiences were related to school or district PD and that they would respond to the items differently for each kind of PD. Given that principals have greater control over PD at their school than PD offered by the district, I reworded the introduction to these items to include instructions to consider only PD at the school when answering these items. In making this revision, I aimed to focus the scale on PD opportunities that principals could develop to meet the unique needs of their teachers.

Inappropriate response anchors

When survey items asked about the frequency with which participants engage in specific practices, teachers' responses seemed less accurate. The accuracy of their

responses appeared to be related to the wording of the response anchors. For example, one item asked: "How often have you worked with members of your team to discuss teaching decisions based on student work?" Two teachers chose the response anchor "almost daily" and explained that they regularly have informal conversations with colleagues during lunch about individual students. Although both informal and conversations are related to teacher learning, these teachers' responses seemed to indicate that they were responding based on how often they discussed students but not necessarily how often they engaged in the practice of discussing "teaching decisions based on student work."

The response anchors may have encouraged less than optimal responses. This item asked about a time-intensive practice, discussing teaching decisions based on student work. The response anchor may have signaled to teachers that they *should* have these discussions "almost daily," possibly leading teachers to report that they engage in these discussions more frequently than they actually do. Thus, I modified the frequencies described in the response anchors to reflect the fact that they would not have adequate time to engage in these practices every day.

Conclusion

Based on information from cognitive interviews, I revised items identified as difficult to understand or respond to accurately. I aimed to improve the validity and reliability of survey items by more clearly specifying the focus on teaching and learning, the people referenced by the item, and the school context. In addition, I revised the wording of response anchors that asked about the frequency with which teachers engaged

in specific practices by making them more appropriate given the time needed to engage in the practices referenced.

Study 3: Examining the Factor Structure

The goal of Study 3 was to examine the factor structure of the Leadership Practices for Instructional Improvement, Organizational Processes, and Efficacy Beliefs domains. Given the desire to provide school leaders with maximum information about school capacity for improvement, it was important to examine whether each of these domains appeared to be comprised of multiple scales or whether they appeared to represent only one dimension of information. As I describe below, I found evidence to suggest that each domains was comprised of multiple dimensions of information, though not necessarily the same dimensions as originally hypothesized. These findings supported further revisions of the survey, including combining scales that measured the same dimension of information and retaining only the items that best measured each construct.

Participants

The sample for Study 3 included 2,550 elementary and K-8 teachers and teaching assistants from a Southern US district that were responsible for classroom instruction.[2] The district required survey participation but allowed individuals to decide whether their responses could be used in research. The sample used in Study 3 consisted of participants who allowed their responses to be used in research and completed at least half of the survey, 80% of individuals who received the survey. Participants in the sample worked in schools that varied greatly in size, student achievement, and student socio-economic

status. Since IC is a school-level construct, I surveyed all faculty members responsible for instruction to learn about their collective perceptions of school capacity for improvement. While other faculty members, such as librarians, contribute to school culture and student learning, they were excluded from the survey because of its focus on instruction.

Procedures

I conducted PCA to identify the most efficient measures for retaining maximum information. In other words, the goal of PCA was to identify and retain the survey items that together explain maximum variance in participant responses. I identified the number of dimensions of information present in each domain and retained only the items that were the best measures of each dimension of information. I used Kaiser's criterion (Kaiser, 1960) to identify the number of components with eigenvalues of 1.0 or greater to retain for further investigation. Since PCA may result in too many components, I used a scree test to plot the eigenvalues of each component in descending order of magnitude and visually determine the number of factors that appeared to provide meaningful information about response variance (Kline 2000).

Given the large number of variables included in each domain, interpreting the loadings of the principal components was challenging; therefore, I engaged in an iterative process to determine the items that best measured each of the principal components. I first considered factor loadings and conceptual connections between items to determine the items that appeared to best measure each principal component. Then I removed items that seemed to be the strongest measures of one component and re-examined the factor

loadings on the other principal components. Data from PCA provided evidence to improve the efficiency of the survey scales by combining scales that appeared to measure the same dimension of information and removing indicators that were not highly correlated with the construct they were designed to measure.

After identifying the items that seemed to best measure each component, I computed Cronbach's alpha to analyze the estimated internal-consistency reliability of each of these scales. In other words, I estimated the degree to which respondents answered all items in a scale similarly since these items were designed to measure the same construct. This could confirm or disconfirm whether items that appeared to be related to the same construct during expert validation were highly related in pilot survey responses (Gehlbach & Brinkworth, 2011). Since the goal of item removal was to enhance not only efficiency but also the precision with which items measured the underlying construct, I recommended removing items based on a combination of the conceptual basis of the scale and empirical factor loadings (Schechter, 2008).

Findings and Discussion

The IC Survey was designed to measure 12 distinct dimensions of information related to IC. By conducting PCA, I found evidence that there were 10 dimensions of information present. I found strong evidence to support the original hypothesis that the survey measured two constructs related to the Efficacy Beliefs domain. However, my analysis led me to combine some of the original scales in the Leadership Practices for Instructional Improvement and Organizational Processes domains that appeared to measure the same construct. In addition, I was able to improve the efficiency of the scales

by retaining only those items that best measured each construct conceptually and quantitatively. I present a detailed analysis of the factor structure of each domain below.

Domain 1: Leadership Practices for Instructional Improvement

I conducted PCA to identify the smallest number of components that could be used to best explain the interrelations among the items in the Leadership Practices for Instructional Improvement domain and identify the indicators that best measured these factors. The IC Survey was designed to measure four dimensions of information about leadership, as represented by the following four scales: Modeling Public Learning, Instructional Leadership, Creation of a Learning Environment, and Professional Development. However, I identified only three components with eigenvalues greater or equal to one, suggesting that there may be three orthogonal dimensions of information present in the original 26 indicators (Willett, 2011).

Although the first component explained most of the variance in responses (64.55%), examination of the scree plot and the factor loadings on each of the first three components provided evidence that there were three unique dimensions of information present in the domain. I found that all items of the Professional Development scale and six items of the Creation of a Learning Environment scale clustered together. As noted previously, two experts thought that the Creation of a Learning Environment scale measured two constructs: psychological safety and support. The items that appeared to hold together as one dimension of information were all measures of psychological safety. This evidence supported experts' assertions that instructional support was not a core

element of psychological safety; thus, I renamed the construct "Psychological Safety" and retained only the six items that clustered together during PCA.

Finally, I combined the Modeling Public Learning and Instructional Leadership scales since they appeared to represent only one dimension of information. I found that this construct measured leadership practices that foster learning, such asking questions and supporting teachers in solving instructional issues; thus, I renamed the combined scale Leadership for Learning. Reviewing the factor loadings together with the conceptual basis of each item, I identified five items for removal. These items appeared to be weaker measures of the construct because they focused on monitoring teaching and learning rather than encourage ongoing learning.

After thorough examination of the three dimensions of information in the domain identified during PCA, I separated the indicators into three scales: 1) Leadership for Learning, 2) Psychological Safety, and 3) Professional Development. As I show in Table 1, the estimated internal-consistency reliability for each of the three revised scales was relatively high. The high Cronbach's alpha estimates indicated that respondents answered quite similarly across all items in each scale, providing additional evidence of the strong relationships among indicators in each scale.

Domain 2: Organizational Processes

Similar to the Leadership Practices for Improvement domain, I found evidence to support the presence of most but not all dimensions of information the Organizational Processes domain was intended to measure. After revisions based on cognitive interviews and expert review, the domain was designed to measure six constructs: Collaboration

Around an Improvement Strategy, Teachers' Collective Work, Shared Understanding of Effective Practice (Whole-School), Shared Understanding of Effective Practice (Team), Support for Team, and Team Processes. However, by conducing PCA, I identified only five components with eigenvalues equal or greater than 1.0. Although the first component explained the majority of the variance (53.38%) in responses, evidence from PCA and the scree plot suggested that there were five dimensions of information present. Examination of the loadings of the 41 original indicators on each of the five components provided evidence to suggest that the Teachers' Collective Work and Shared Understanding of Effective Practice (Whole-School) scales represented only one component, but all other scales in the domain represented unique dimensions of information.

As noted previously, one of the threats to validity identified during expert review was multidimensionality, when scales measure more than one construct. Examining the factor loadings of the items in each the scale together with the conceptual meaning of the scale, I identified items that appeared to measure a different construct than the one intended. For example, the Share Understanding of Effective Practice (Team) scale was designed to measure the practices that principals could encourage to *foster* a shared understanding of effective practice; however, half of the items asked about the frequency with which teachers engaged in practices that foster this shared understanding and half asked about the extent to which team members were committed to developing this shared understanding. The items that asked about specific practices better represented the construct and clustered together during PCA, which led me to retain only these items in the scale. I enhanced the construct validity and efficiency of the scales by removing items that appeared to be weakly related to each scale conceptually and quantitatively.

I found that the Teachers' Collective Work and Shared Understanding of Effective Practice (Whole-School) scales appeared to measure the same dimension of information, which led me to combine them. The items that were most strongly related to the construct asked about teachers' collective involvement in decisions about instructional practice; thus, I renamed the combined scale, Teachers' Involvement in Instructional Decisions, and only retained those items that best measured this construct conceptually and quantitatively.

As I show in Table 2, I found that the estimated internal-consistency reliability of each of the revised scales was quite high. Thus, I concluded that the Organizational Processes domain was best measured by the following five constructs: Collaboration Around an Improvement Strategy, Teachers' Involvement in Instructional Decisions, Shared Understanding of Effective Practice (Team), Support for Team, and Team Processes.

Domain 3: Efficacy Beliefs

By conducting PCA and examining the scree plot, I found strong evidence to suggest that there were two dimensions of information present in the Efficacy Beliefs domain, which reflected the Collective Efficacy and Individual Efficacy scales. These two dimensions explained 62.13% of the total variance in responses, with each dimension explaining a somewhat large proportion of the variance (42.98% and 19.15%, respectively).

By examining item factor loadings together with the relation of each item to the conceptual basis of each scale, I was able to identify several items for removal. For

instance, I found that items that asked about behavior management, rather than academic instruction seemed to be weaker measures of teachers' individual and collective efficacy beliefs. After removing items that were relatively weak measures of each scale, I found that the estimated internal-consistency reliability of both the Individual Efficacy and Collective Efficacy scales was quite high, as I show in Table 3.

Conclusion

Evidence from PCA provides preliminary evidence to support the hypothesis that each domain measures multiple dimensions of information about the practices, processes, and beliefs related to IC. Revisions made based on this evidence served to enhance the efficiency and precision of the survey scales. Thus, these revisions improved the usefulness of the survey for practitioners and researchers by more accurately measuring each component related to the underlying construct, IC.

General Discussion

The findings from this series of studies suggest that the revised IC Survey is a valid and reliable tool for collecting and using diagnostic data about school IC in decision making in schools. Evidence from expert review, cognitive interviews, and PCA led to three key improvements in the survey. First, survey items are more focused on issues related to instructional practice and student learning. Second, the survey items are more focused on the processes, practices, and beliefs that principals can actively foster to enhance school capacity rather than outcomes related to teaching and learning. Third, the revised IC Survey is a more efficient measure of the constructs related to IC. As I show in

Figure 2, findings from expert review, cognitive interviews, and PCA led me to reduce the number of scales in the survey from 12 to 10 and the total items included in the survey from 91 to 57. In addition, findings from PCA suggested that the survey measured multiple dimensions of information about school IC. These findings suggest that the revised IC Survey is an effective tool for providing unique information about specific practices and processes that principals can put in place to enhance school capacity for improvement.

The usefulness of the survey depends on how well it measures the factors related to IC. Validity of the survey depends on both the accuracy with which it measures the theoretical constructs it is intended to measure and the appropriateness of the instrument for its intended use. Thus, "validity is an evolving property and validation a continuing process" (Messick, 1995, 741). In this study, I brought together researchers and practitioners who are knowledgeable about continuous improvement and the district context in which the survey will be used. This partnership between researchers and practitioners may serve as a model for future analyses of validity when the survey is used in new contexts or for new purposes.

Additional research would be needed to understand the relationships among the factors related to IC. This study provided evidence to support the presence of multiple dimensions of information related to IC; however, it was unclear how these factors were related to the hypothesized domains and the larger construct of IC. Conducting CFA with data from a future administration of the IC Survey could confirm whether or not the survey measures the factors I have identified in this study and provide insights into the relationships among these factors.

One of the limitations of this study is the use of a sample of elementary and K-8 teachers for survey development and testing. Given this sample, the language used to describe leadership and teachers' work in teams may be inappropriate for use in secondary schools. It would be valuable for future research to examine how to measure leadership practices in schools with multiple leaders, such as grade-level or departmental leaders.

Summary and Conclusions

Accountability policies and professional standards encourage school leaders to use data to inform decision making in schools. Although substantial data is available on student performance, data on research-based practices and organizational processes that school leaders can establish to foster schoolwide improvement is more limited (Anderson et al., 2010; Halverson, 2010). Given the challenge of leading for continuous improvement, concrete tools are needed to support school leaders in applying learning from research to the unique challenges they face in improving instruction and student learning schoolwide. Findings from this series of studies suggest that the IC Survey is a valid and reliable tool for collecting and using information about practices, processes, and beliefs related to school capacity for improvement. Further research is needed to understand how school leaders apply this information to their school context.

Researchers can use the survey to assess the level of IC in schools and examine the relationships among factors related to school capacity for improvement. Scholars suggest that the actions school leaders take to support improvement may differ depending on the level of school capacity (Day et al., 2010; Hallinger & Heck, 2011). Exploring the

relationships among the factors on the IC Survey could further our understanding how schools improve and the actions school leaders can take to support improvement in teaching and learning at different levels of capacity.

The IC Survey was designed to provide school leaders with research-based evidence about practices and processes for schoolwide improvement based on their distinctive school context. Policymakers seeking to foster continuous improvement can consider encouraging the use of diagnostic tools for assessing organizational conditions related to improvement as part of data driven decision making. Performance data, such as student achievement data and teacher evaluations, can support school leaders in setting goals for improvement but provides little guidance on how to improve. The IC Survey can act as a powerful tool for providing school leaders with information about specific practices, processes, and beliefs they can foster to enhance school capacity for improving instruction and student learning.

References

Anderson, S., Leithwood, K., & Strauss, T. (2010) Leading data use in schools: Organizational conditions and practices at the school and district levels. *Leadership and Policy in Schools, 9*(3), 292-327.

Argyris, C., & Schon, D. A. (1996). *Organizational learning II: Theory, methods and practice*. Reading, MA: Addison-Wesley.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman and Company.

Bandura, A. (2000). Exercise of human agency through collective efficacy. *Current Directions in Psychological Science, 9*(3), 75-78.

Barge, S., & Gehlbach, H. (2012). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education*, *53*(2), 182-200.

Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: The University of Chicago Press.

Coburn, C. E., & Stein, M. K. (Eds.). (2010). *Research and practice in education: Building alliances, bridging the divide.* Lanham, MD: Rowman & Littlefield.

Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research*, *12*, 229-238.

Colyvas, J.A. (2012) Performance metrics as formal structures and through the lens of social mechanisms: When do they work and how do they influence? *American Journal of Education, 118*(2), 167-197.

Council of Chief State School Officers. (2008). *Educational leadership policy standards: ISLLC 2008*. Washington, DC: Author.

Day, C., Sammons, P., Leithwood, K., Hopkins, D., Harris, A., Gu, Q., & Brown, E. (2010). *Ten strong claims about successful school leadership*. Nottingham, UK: The National College for School Leadership.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.), Hoboken, NJ: J. Wiley.

Donovan, M. S. (2011). *The SERP approach to research, design, and development: A different role for research and researchers.* Paper accepted for presentation at the annual conference of the American Education Research Association, New Orleans, LA.

Edmondson, A. (2002). The local and variegated nature of learning in organizations: A group-level perspective. *Organization Science, 13*(2), 128-146.

Edmondson, A. C. (2008). The competitive imperative of learning. *Harvard Business Review*, *86*(7), 60–67.

Elmore, R. F. & Forman, M. L. (2011). Building Coherence within Schools. Retrieved December 18, 2012 from the Strategic Education Research Partnership Website: http://ic.serpmedia.org

Firestone, W.A., & Gonzalez, R.A. (2007). Culture and processes affecting data use in school districts. In Moss, P.A. (Ed.), *Evidence and decision making; 2007*

Yearbook of the National Society for the Study of Education (pp. 132-154). Chicago: NSSE.

Fullan, M. (2007). *The new meaning of educational change.* New York: Teachers College Press.

Garcia, A. A. (2011). Cognitive interviews to test and refine questionnaires. *Public Health Nursing*, *28*(5), 444-450.

Garvin, D. A., Edmondson, A. C., & Gino, F. (2008). Is yours a learning organization? *Harvard Business Review*, *86*(3), 109–116.

Gehlbach, H., & Brinkworth, M. E. (2011). Measure twice, cut down error: A process for enhancing the validity of survey scales. *Review of General Psychology, 15*(4), 380-387.

Goddard, R. D. (2002). A theoretical and empirical analysis of the measurement of collective efficacy: The development of a short form. *Educational and Psychological Measurement*, *62*(1), 97-110.

Goddard, R. D. & Goddard, Y. L. (2001). A multilevel analysis of the relationship between teacher and collective efficacy in urban schools. *Teaching and Teacher Education, 17,* 807-818.

Goddard, R. D., Goddard, Y. L., Kim, E. S., & Miller, R. J. (in press). Coherent evidence: The roles of instructional leadership, teacher collaboration, and collective efficacy beliefs. *American Educational Research Journal.*

Goddard, Y. L., Goddard, R. D., & Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teachers College Record*,

109(4), 877-896.

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2004). Collective efficacy beliefs: Theoretical developments, empirical evidence, and future directions. *Educational Researcher, 33*(3), 3-13.

Goldring, E., Porter, A., Murphy, J., Elliott, S.N., & Cravens, X. (2009). Assessing learning-centered leadership: Connections to research, professional standards, and current practices. *Leadership and Policy in Schools, 8*(1), 1-36.

Hallinger, P., & Heck, R.H. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980–1995. *Educational Administration Quarterly*, *32*, 5–44.

Hallinger, P., & Heck, R.H. (2011). Exploring the journey of school improvement: Classifying and analyzing patterns of change in school improvement processes and learning outcomes. School *Effectiveness and School Improvement*, *22*, 1–27.

Halverson, R. R. (2010). School formative feedback systems. *Peabody Journal of Education*, *85*, 130-146.

Hawley, W., & Sykes, G. (2007). Continuous school improvement. In W. Hawley (Ed.), *The keys to effective schools: Educational reform as continuous improvement* (pp. 153–172). Thousand Oaks, CA: Corwin.

Heck, R. H., & Hallinger, P. (2009). Assessing the contribution of distributed leadership to school improvement and growth in math achievement. *American Educational Research Journal*, *46*(3), 659-689.

Higgins, M., Ishimaru, A., Holcombe, R., & Fowler, A. (2011). Examining the building blocks of learning in schools. *Journal of Educational Change*.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurements, 20,* 141–151.

Kelley, C., & Halverson, R. (2012). The comprehensive assessment of leadership for learning: A next generation formative evaluation and feedback system, *Journal of Applied Research on Children: Informing Policy for Children at Risk*, *3*(2), Article 4.

Kline, P. (2000). *The handbook of psychological testing* (2nd ed.). New York: Routledge.

Leithwood, K., & Mascall, B., (2008). Collective leadership effects on student achievement. *Educational Administration Quarterly, 4*, 529-561.

Leithwood, K., Seashore Louis, K., Anderson, S. & Wahlstrom, K. (2004). *How leadership influences student learning*. New York: Wallace Foundation.

Leithwood, K., Patten, S., & Jantzi, D. (2010). Testing a conception of how school leadership influences student learning. *Administration Quarterly*, *46*(5), 671-706.

Light, R.J., Singer, J.D., & Willett, J.B. (1990). By design: Planning research on higher education. Cambridge, MA: Harvard University Press.

Mandinach, E. B., & Gummer, E. S. (2013). A Systemic View of Implementing Data Literacy in Educator Preparation. *Educational Researcher*, *42*(1), 30-37.

Marks, H. M. & Printy, S. M. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly, 39*(3), 370-397.

Marsden, P.V. & Wright, J.D. (Eds.). (2010). *Handbook of survey research* (2nd ed.). Bingley, UK: Emerald Group Publishing Limited.

McKenzie, J. F., Wood, M. L., Kotecki, J. E., Clark, J. K., & Brey, R. A. (1999).

Establishing content validity: Using qualitative and quantitative steps. *American Journal of Health Behavior, 23*(4), 311-318.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749. doi: 10.1037/0003-066X.50.9.741

Moolenaar, N.M., Sleegers, P.J.C., & Daly, A. J. (2012).Teaming up: Linking collaboration networks, collective efficacy, and student achievement. *Teaching and Teacher Education*, *28*(2), 251–262.

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

Porter, A. C., Polikoff, M. S., Goldring, E., Murphy, J., Elliott, S. N., & May, H. (2010). Developing a psychometrically sound assessment of school leadership: The VAL-ED as a case study. *Educational Administration Quarterly, 46*(2), 135–173.

Printy, S. M., Marks, H. M., & Bowers, A. J. (2009). Integrated Leadership: How Principals and Teachers Share Transformational and Instructional Influence. *Journal of School Leadership*, *19*(5), 504-532.

Schechter, C. (2008). Organizational learning mechanisms: The meaning, measure, and implications for school improvement. *Educational Administration Quarterly*, *44*(2), 155–186.

Seidman, I. (2006). Interviewing as qualitative research: A guide for researchers in education and the social sciences. New York: Teachers College.

Stajkovic, A. D. & Luthans, F. (1998). Self-efficacy and work-related performance: A meta- analysis. *Psychological Bulletin*, *124*(2), 240-261.

Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed Likert items. *Journal of Marketing Research, 45*(1), 116-131.

Thoonen, E.E.J, Sleegers, P.J.C., Oort, F.J., Peetsma, T.T.D, & Geijsel, F.P. (2011). How to improve teaching practices: The role of teacher motivation, organizational factors, and leadership practices. *Educational Administration Quarterly, 47*(3), 496-536.

Tschannen-Moran, M & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17,* 783-805.

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education, 24*, 80-91.

Willett, J. (2011). *S052/III.1(bc): Using PCA to form an "ideal" composite from multiple indicators* [PowerPoint slides].

Notes

[1] Richard Elmore, Michelle Forman, and Candice Bocala, collaborated in analyzing data from expert review and cognitive interviews.

[2] The district hired the IC research team to assess IC in all elementary and K-8 schools and provide professional development to support school leaders in building capacity for improvement in these schools.

Figure 1. Internal Coherence Framework

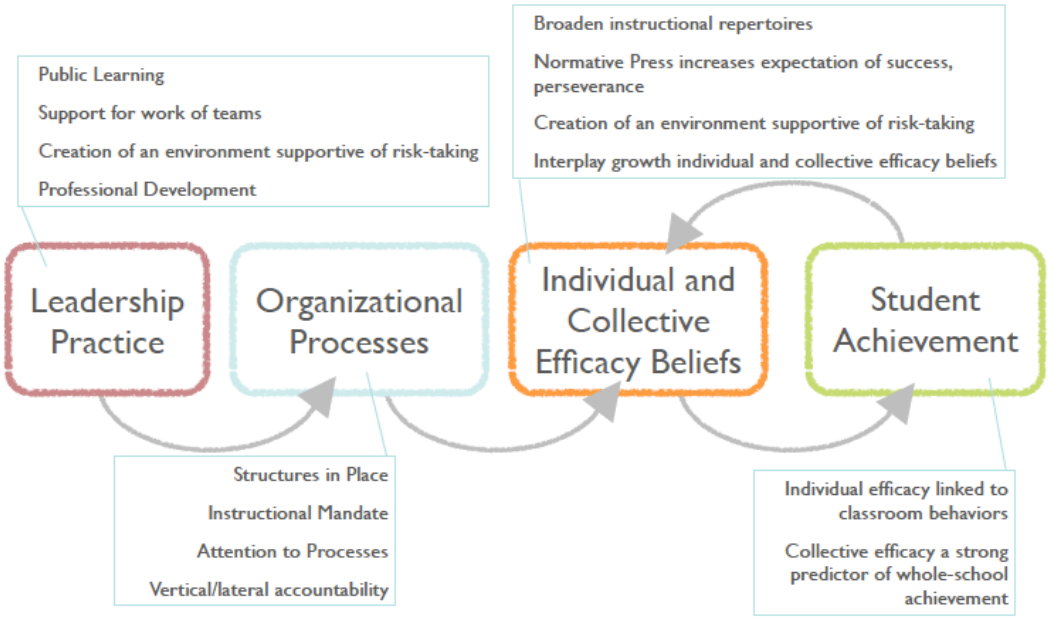

Figure 1. Internal Coherence Framework displaying the relationships between the Leadership Practices for Instructional Improvement, Organizational Processes, and Efficacy Beliefs domains and their connection to student achievement. From "Building Coherence within Schools," by R. F. Elmore & M. L. Forman, 2011, retrieved March 18, 2013 from the Strategic Education Research Partnership Website: http://ic.serpmedia.org.

Table 1. Revised Scales in Leadership Practices for Instructional Improvement Domain

Table 1. Indicators, observations and Cronbach's coefficient alpha of each revised scale in the Leadership Practices for Instructional Improvement domain.

Scale (Items)	Indicators	Factor loadings	Cronbach coefficient alpha
Leadership for Learning (8)	The principal at this school invites input from faculty in discussions about teaching and learning.	0.3546	0.9700
	The principal at this school asks probing questions about teaching and learning.	0.3540	
	The principal at this school listens attentively.	0.3578	
	The principal at this school encourages multiple points of view.	0.3606	
	The principal at this school acknowledges his / her own limitations with respect to knowledge or expertise.	0.3550	
	The principal at this school is knowledgeable about effective instructional practices.	0.3482	
	The principal communicates a clear vision for teaching and learning at our school.	0.3507	
	The principal at this school is directly involved in helping teachers address instructional issues in their classrooms.	0.3473	
Psychological Safety (6)	People in this school are eager to share information about what does and does not work.	0.3523	0.9250
	Making mistakes is considered part of the learning process in our school.	0.4159	
	If I make a mistake at this school, it will not be held against me.	0.4047	
	In this school, teachers feel comfortable experimenting with untried teaching approaches, even if they may not work.	0.4095	
	In this school, it is easy to speak up about what is on your mind.	0.4330	
	People in this school are usually comfortable talking about problems and disagreements about teaching and learning.	0.4289	
Professional Development (5)	My professional development experiences this year have been closely connected to my school's improvement plan.	0.4376	0.9406
	My professional development experiences this year have included enough time to think carefully about, try, and evaluate new ideas.	0.4489	
	My professional development experiences this year have been valuable to my practice as a teacher.	0.4480	
	My professional development experiences this year have been designed in response to the learning needs of the faculty, as they emerge.	0.4387	
	My professional development experiences this year have included follow-up support as we implement what we have learned.	0.4426	

Table 2. Revised Scales in Organizational Processes Domain

Table 2. Indicators, factor loadings and Cronbach's coefficient alpha of each revised scale in the Organizational Processes domain.

Scale (Items)	Indicators	Factor loadings	Cronbach coefficient alpha
Collaboration Around an Improvement Strategy (4)	Our school has an improvement plan, of which we are all aware.	0.4717	0.9392
	We focus our whole-school improvement efforts on clear, concrete steps.	0.5089	
	We coordinate curriculum, instruction and learning materials with our school improvement plan.	0.5054	
	The programs or initiatives we implement connect clearly to our school improvement plan.	0.5133	
Teachers' Involvement in Instructional Decisions (7)	Teachers in this school work collectively to plan school improvement.	0.3744	0.9522
	Teachers in this school work collectively to select instructional methods and activities.	0.3906	
	Teachers in this school work collectively to evaluate curriculum an programs.	0.3922	
	Teachers in this school work collectively to determine professional development needs and goals.	0.3839	
	Teachers in this school work collectively to plan professional development activities.	0.3723	
	As a full faculty, we work toward developing a shared understanding of effective instructional practices.	0.3682	
	As a full faculty, we regularly revisit and revise our thinking about the most effective instructional practices we can use with our students.	0.3631	
Shared Understanding of Effective Practice (Team) (4)	How often have you worked with members of your team to discuss teaching decisions based on student work?	0.4990	0.9138
	How often have you worked with members of your team to discuss teaching decisions based on student assessment data?	0.5131	
	How often have you worked with members of your team to evaluate curricular or assessment materials?	0.5058	
	How often have you worked with members of your team to discuss lesson plans or specific instructional practices?	0.4815	
Support for Team (5)	The principal at this school provides teacher teams with the right balance of direction and independence.	0.4545	0.9372
	The principal at this school gives teacher teams a clear and meaningful purpose for their time together.	0.4636	
	The principal at this school provides adequate time for teacher teams to meet.	0.4297	
	The principal at this school ensures that teacher meeting time is protected and maintained consistently throughout the year.	0.4296	
	The principal at this school supports teacher teams in following through on instructional decisions made by the group.	0.4575	
Team Processes (6)	Our team meetings have an agenda, which we do our best to follow.	0.3985	0.9056
	There is always someone who has the responsibility of guiding or facilitating our team discussions.	0.3888	
	When our team makes a decision, all teachers on the team take responsibility for following through.	0.4218	
	Our team meetings include productive debate.	0.4036	
	All members of the team are actively involved in our collective	0.4077	

	learning.	
	Team meetings connect to each other and the overarching purpose for team work.	0.4278

Table 3. Revised Scales in Efficacy Beliefs Domain

Table 3. Indicators, factor loadings and Cronbach's coefficient alpha of each revised scale in the Efficacy Beliefs domain.

Scale (Items)	Indicators	Factor loadings	Cronbach coefficient alpha
Individual Efficacy (5)	How confident are you that you can craft good questions for your students?	0.4194	0.8698
	How confident are you that you can use a variety of assessment strategies?	0.4500	
	How confident are you that you can provide an alternative explanation or example when students are confused?	0.4516	
	How confident are you that you can provide appropriate challenges for very capable students?	0.4580	
	How confident are you that you can differentiate instruction for individual students?	0.4559	
Collective Efficacy (6)	Teachers in this school are confident they will be able to motivate their students.	0.3861	0.9234
	Teachers in this school have the skills needed to produce meaningful student learning.	0.4119	
	If a child doesn't learn something the first time, teachers will try another way.	0.4065	
	Teachers in this school believe that every child can learn.	0.4034	
	Teachers in this school are skilled in various methods of teaching.	0.4189	
	Teachers in this school have what it takes explore new instructional approaches to help underperforming students meet standards.	0.4216	

Figure 2. Original and Revised Internal Coherence Survey scales and items

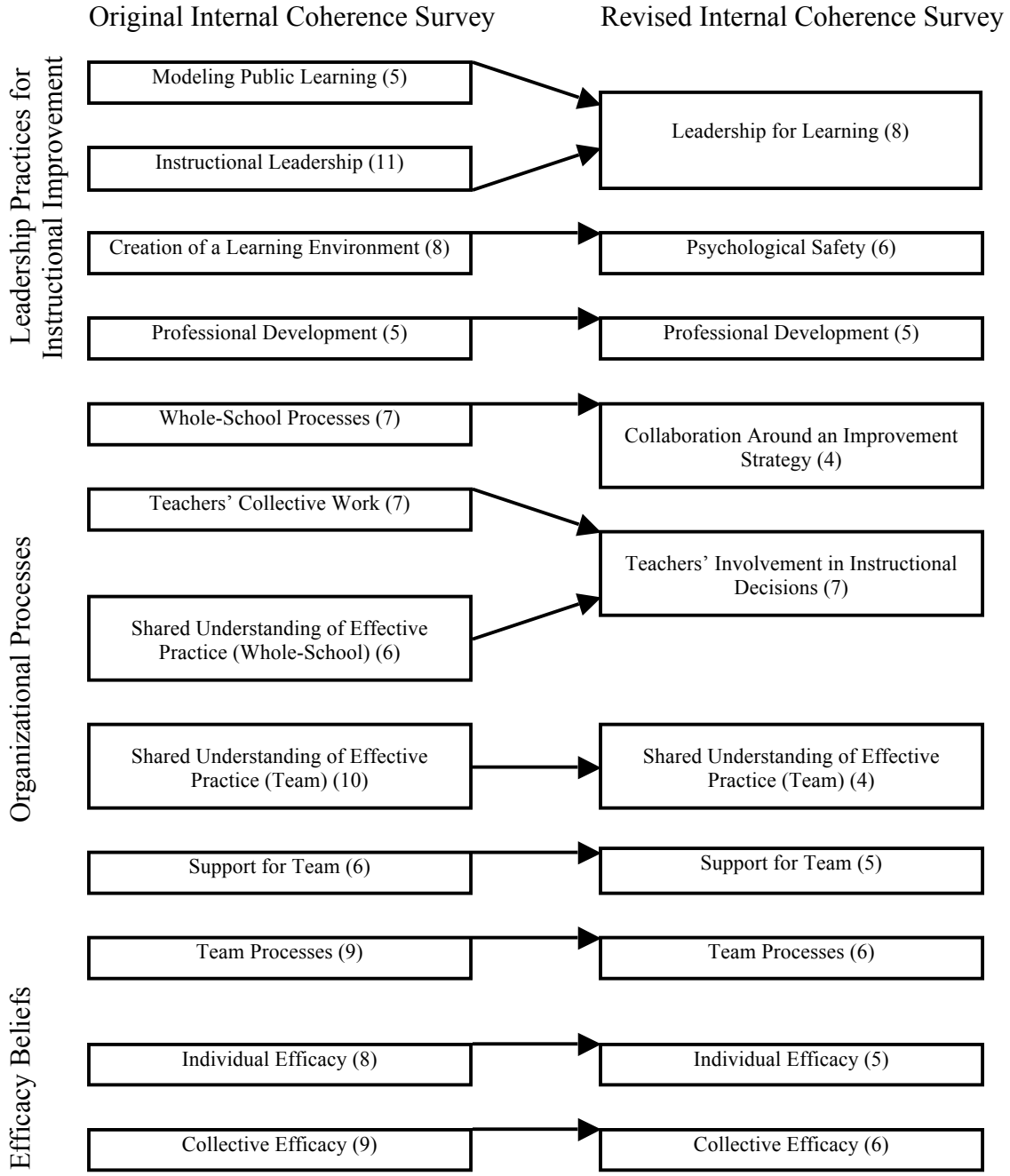

Figure 2. Original and Revised Internal Coherence Survey displaying the relationships between the scales and number of items.