District Capacity Assessment (DCA) Technical Manual

Developed in collaboration between NIRN and MIBLSI

July, 2016

Suggested Citation: Russell, C., Ward, C., Harms, A., St. Martin, K., Cusumano, D., Fixsen, D. Levy, R. & LeVesseur, C. (2016). District Capacity Assessment Technical Manual. National Implementation Research Network, University of North Carolina at Chapel Hill

Table of Contents

Preface

Purpose of This Manual

The purpose of the District Capacity Assessment (DCA) Technical Manual is to provide background information on the technical adequacy of the DCA (Ward et al., 2015). This current version draws upon a rich history and background of previous work assessing district capacity. Notably, the current version includes significant modifications from earlier iterations including revised items, a rubric for scoring and a glossary of terms. This version of the DCA was released in the spring of 2015 following a thorough development process and early validation work resulting in a high quality assessment of district capacity for implementation of effective innovations. Validity evidence collected during the assessment development process is rarely obtained and when it is obtained it is not often presented in detail (Carretero-Dios & Perez, 2007). This technical manual details the development process to date, the validity work that has been completed, usability testing efforts accomplished and an outline of next steps to continue the work to a fully established assessment for local education agencies (LEAs).

Audience

This manual was written for state, regional and local agencies that are considering or already using the DCA to assess district capacity for implementation of effective innovations. This manual can help with the selection process an agency may engage in when choosing an assessment of capacity. Additionally, DCA Administrators, facilitators and respondents may use this manual to deepen their background knowledge on the development and validation of the DCA.

Overview of the District Capacity Assessment (DCA)

Description of the District Capacity Assessment

The District Capacity Assessment (DCA) is a 26-item team-based self-assessment developed to assist school Local Education Agencies (e.g., school districts) in the implementation of effective innovations that benefit students (Ward et al., 2015). A District Implementation Team, inclusive of District Leadership, uses the DCA to assist with the development of an action plan to improve capacity for implementation of an effective innovation (EI), help with monitoring of the action plan's effectiveness in improving overall capacity, and support the development of a district-wide consistent structure for supporting initiatives and practices across schools. Additionally, the DCA can be used as part of a feedback structure to a state or regional education body to improve and focus the work of individuals who support districts.

The district team works through items with a specific Effective Innovation in mind. An effective innovation is "anything that is new to a district and that is intended for use to improve effectiveness or efficiency. The innovation was developed based on the best available evidence (e.g., evaluation results, research findings)" (Ward et al., p. 29). Consequently, a team is able to utilize the DCA with any/all innovations that are occurring within the system or use the assessment with their most prominent initiative.

The DCA is grounded in the understanding that districts must develop capacity in the Active Implementation Frameworks (Fixsen, et al. 2005) to reach desired outcomes from an innovation. Ward et al. (2015) define district capacity as the development of "systems, activities, and resources that are necessary for schools to successfully adopt and sustain Effective Innovations"(p.5). Key organizational activities required for strong implementation and sustainability of efforts are organized into three critical Implementation Drivers that include Leadership, Competency, and Organization.

> **Leadership -** Active involvement in facilitating and sustaining systems change to support implementation of the effective innovation through strategic communication, decisions, guidance, and resource allocation. *Leadership includes: Leadership and Planning*
>
> **Competency** - Strategies to develop, improve, and sustain educators' abilities to implement an Effective Innovation as intended in order to achieve desired outcomes. *Competency Drivers include: Performance Assessment, Selection, Training, and* Coaching
>
> **Organization** – Strategies for analyzing, communicating, and responding to data in ways that result in continuous improvement of systems and supports for educators to implement an effective innovation. *Organization Drivers include: Decision Support Data System, Facilitative Administration, and Systems Intervention*

The suggested schedule for conducting a DCA is twice a year, about every six months. An administration in February/March timeframe has been found to be strategic in informing district budgeting process for the upcoming school year followed by a repeated administration

six months later. Throughout the administration of the 26-item self-assessment, a rubric is utilized to anchor current functioning with a score of 0, 1 or 2.

The DCA requires specific roles, including a DCA Administrator, Facilitator, Note Taker and Respondents. Preparation for the administration of assessment includes commitment to the time for the DCA administration, identification of roles, and securing leadership support for the administration and use of the results for action planning (please see DCA Administrating training course for more training on how to administer the DCA at http://implementation.fpg.unc.edu/resources/district-capacity-assessment-dca). During administration the team uses a simultaneous and public voting process where respondents simultaneously hold up either a finger or a response card to indicate their vote of a 0, 1, or 2 for each item. Voting is guided by requirements included in the DCA scoring guide (rubric). The facilitator contributes to the process by providing contextualization for any items and rubric requirements. While scoring is important, discussions occurring throughout the administration process serve as critical links to action planning. Upon completion of the DCA, the team enters their results into a web-based application provided by SISEP.org, which stores data from all administrations. It is important to note that the DCA does not end when the last item is scored. In contrast, the team then moves into developing an action plan that includes assigning activities to improve the district's capacity to support the identified EI.

History of DCA

The State Implementation and Scaling up of Evidence-based Programs (SISEP) Center began in October 2007. One of the goals was to find or develop implementation capacity assessments. By June 2008, a review of available measures had been completed by Sandra Naoom and Michelle Duda (SISEP) and Amanda Fixsen (Portland State University graduate student). Content derived from this effort was then reviewed by Rob Horner (University of Oregon) and George Sugai (University of Connecticut) who were co-developers of SISEP. In December 2008 the first draft of State Capacity Assessment (SCA) items was circulated among SISEP Evaluation Advisory Board members David Mank, Mark Greenberg, and Mitchell Yell.

In 2009, the SISEP staff began to develop items related to state capacity, regional capacity, and district capacity. By January 2010, a first draft of district capacity items was produced.

In March 2010, Carol Sadler (State Transformation Specialist in Oregon; developer of Effective Behavior & Instructional Support Systems [EBISS] in the Tigard-Tualatin school district) conducted a crosswalk of the EBISS District System Support Plan (DSSP) assessment tool and the draft of SISEP district capacity items. This led to a broader discussion of related district assessment work taking place in Oregon.

In late 2010, Rob Horner convened a group of individuals who had established and were using some form of an assessment of district capacity. The group included SISEP (Michelle Duda), EBISS (Carol Sadler and Erin Chaparro), Response to Intervention (David Putnam), and Positive Behavior Interventions and Supports (Rob Horner). Concept areas, the logical relationships

between items and concept areas, the wording of items, scaling rubrics, and so on were major topics for discussion in the periodic meetings that occurred over the next several months.

Subsequently, another program, Michigan's Integrated Behavior Learning Support Initiative (MIBLSI), led by Steve Goodman, created the District Multi-Tiered Systems of Support Capacity Assessment (DMCA) in 2011. The assessment was highly informed by the DCA but tailored for Michigan teams and specific to the effective innovation of Multi-Tiered Systems of Support (MTSS). In an effort to use feedback to shape the items, language and format of the assessment, the DMCA was put through a content validation study and usability testing.

In March 2012 the first complete draft of the SISEP District Capacity Assessment (DCA) was distributed to SISEP staff for its first use in districts in active scaling states.

In June 2014, Rob Horner (University of Oregon) convened the SISEP team and MIBLSI team to refine, reconcile, and produce a nationally agreed upon District Capacity Assessment using quantitative and qualitative data collected on the usability on both the original DCA used by SISEP and the DMCA used by MIBLSI. As a result of the reconciliation, content validity, and usability testing processes described in this manual, the revised District Capacity Assessment (v 6.0) was developed and is currently in use by SISEP active scaling states, National Technical Assistance Centers, and other SEA funded implementation work.

Role of Implementation Science within Education

Increased attention is being paid to how innovations are implemented because students cannot benefit from educational practices they do not experience. While this seems obvious (and it is), education systems are working to develop the implementation capacity to help all teachers make good use of evidence-based practices that enhance the quality of education and outcomes for all students. Strong pressure to implement solutions to overcome challenges or problems in social systems such as education are not new; however, pressure to draw solutions from a growing portfolio of strategies that have documented outcomes narrows the pool of innovations from which we can choose. In this quest to affect meaningful changes in educational outcomes, we must dually direct our attention to **what** effective innovations are selected and **how** they are implemented. In short, efforts to improve socially significant outcomes for students and families require strong collaborative systems supporting the implementation of practices selected to address targeted challenges. "How" practices are implemented is as important as "what" strategies are sought to fix the problem.

In 2005, the National Implementation Research Network (NIRN) released a monograph synthesizing implementation research findings across a range of fields. Based on these findings, NIRN developed five overarching frameworks referred to as the Active Implementation Frameworks. The Active Implementation frameworks help define what needs to be done, how to establish what needs to be done, who will do the work and when, and establish the hospitable environment for the work to accomplish the positive outcomes (Blase, Fixsen, Naoom & Wallace, 2005). The Active Implementation Frameworks (AIFs) are universal and apply to attempts to use any innovation. For more information and resources on the Active

Implementation Frameworks, visit the Active Implementation Hub: http://implementation.fpg.unc.edu/modules-and-lessons

Table 1. Active Implementation Frameworks

Framework	Definition
Usable Innovations	To be usable, an innovation must not only demonstrate the feasibility of improving outcomes, but also must be well operationalized so that it is teachable, learnable, doable, and readily assessable.
Implementation Stages	Stages of implementation require thinking through the right activities for each stage to increase the likelihood of successful use of the AIFs and the practice. Stages are exploration, installation, initial implementation, and full implementation
Implementation Drivers	Key components of the infrastructure and capacity that influence the successful use of an innovation. There are three driver domains: Competency (selection, training, coaching, fidelity), Organization (decision support data systems, facilitative administration, systems intervention), and Leadership (adaptive, technical)
Improvement Cycles	Iterative processes by which improvements are made and problems solved based on the Plan-Do-Study –Act Cycle (3 types of cycles: Rapid Cycle problem solving, Usability Testing, and Practice-Policy Communication cycles)
Implementation Teams	Teams are accountable for planning and seeing the implementation process through to full implementation.

Developing the skills, knowledge, and abilities of Local Education Agencies (LEAs) to use the Active Implementation Frameworks is imperative for the sustained and effective use of evidence-based practices so that socially significant and meaningful outcomes are obtained. Without these cooperative and aligned supports, the result is often inequities in outcomes for staff and students. (Skiba, Middelberg & McClain, 2013; Fuchs & Deshler, 2007).

Need for a Measure of District Capacity for Implementation

Attempts to analyze components of implementation have taken several approaches such as: very general measures that do not specifically address core implementation components (e.g. Landenberger & Lipsey, 2005; Mihalic & Irwin, 2003); measures specific to a given innovation that may lack generality across programs (e.g. Olds, Hill, O'Brien, Racine, & Moritz, 2003; Schoenwald, Sheidow, & Letourneau, 2004); or measures that only indirectly assess the influences of some of the core implementation components (e.g. Klein, Conn, Smith, Speer, & Sorra, 2001; Aarons, Cafri, Lugo, & Sawitzky, 2012). In order for LEAs to support schools to

successfully use and sustain the use of evidence-based practices, it is essential to have reliable and valid measures of implementation components. This information will inform the district's planning for effective supports to school staff and will assist them in assessing progress towards implementation capacity. Additionally, these data can be used to conduct rigorous research on effective and efficient implementation supports.

Despite these earlier efforts, the need for a measure that addresses core implementation components that is generalizable across innovations remains. In response to this void, a series of Implementation Capacity Assessments have been developed that span across the educational system from the State Education Agency (SEA) to the school level. These measures target "implementation capacity" with a focus on the systems, activities, and resources that are necessary to successfully adopt, use, and sustain effective innovations. Included in this series is the State Capacity Assessment (SCA), Regional Capacity Assessment (RCA), **District Capacity Assessment (DCA),** and Drivers Best Practices Assessment (DBPA). Importantly, these capacity assessments are "action assessments." That is, they promote actions to support implementation of best practices through rich discussions that occur during the administration process.

Validation of Assessments

Approaches to Validity

Validity is considered the most important issue in assessment. Establishing validity significantly influences the accuracy of assessments and ability for an assessor to assign meaning to its results (Popham, 2008). In education, assessments are routinely used within a cycle of school improvement. These data have the power to sway resource allocation and determine priorities for action planning within a district or school. In light of this, it is essential that assessments be developed in a technically sound manner with appropriate attention paid to psychometric properties such as reliability and validity. Evidence must show that the assessment captures what it was intended to measure and that the meaning and interpretation of test scores are consistent with each intended use. The American Psychological Association (APA) recommends the use of strong psychometric procedures in the design of assessments as a way to reduce or eliminate bias within the assessment (APA, 2010, p. 13).

Historically, approaches to establishing validity have focused on three areas: 1) content validity, 2) criterion validity and 3) construct validity. Typically each of these three areas is conceptualized in isolation and reported separately. While attending to validity in this way can lead to a better understanding of how well an assessment is measuring a construct, Messick (1995) proposed an alternative method where validity is considered one large concept with a number validity sub-areas that should be investigated to validate an assessment tool fully. The *Standards for Educational and Psychological Testing* (American educational Research Association, 2014) reinforce Messick's alternative method stating that best practice is to report findings as five sources of evidence to determine the overall validity of an assessment. These

five sources of validity are; 1) test content, 2) response process, 3) internal structure, 4) relationship to other variables, and 5) consequence of testing.

Table 2. Sources of Validity

Sources of Validity	Description	Example Methodologies
Test Content	Instrument characteristics such as themes, wording, format of items, tasks, questions, instructions, guidelines and procedures for administration and scoring	• Basis for items/literature review • Qualification of authors and reviews • Item writing process • Review by panel of experts • Vetting and editing process
Response Process	Fit between the items and process engaged in by individuals using the assessment	Think Aloud Protocols
Internal Structure	Analysis of patterns and trends among items that allow items to be reduced to larger constructs based on relationships between them	Factor Analysis
Relationship to Other Variables	Relationship of test scores to variables external to the text	Relationship between a test score and an outcome • Predictive evidence • Concurrent evidence • Convergent evidence • Divergent
Consequence of Testing	Intended and unintended consequences of test use	Purpose, use, and outcomes of test administration including arguments for and against

Technical adequacy in the area of validity relies on integrating multiple sources of evidence, but no source of evidence is considered inherently better. It is the relevance and quality of the evidence that matters. While reporting multiple sources of validity is the expectation for an assessment to be considered valid, gathering evidence across all five areas is a lengthy process, not a discrete activity. Therefore, validity evolves over time as additional sources of evidence become relevant to collect and report at varying stages of the assessment development and use process.

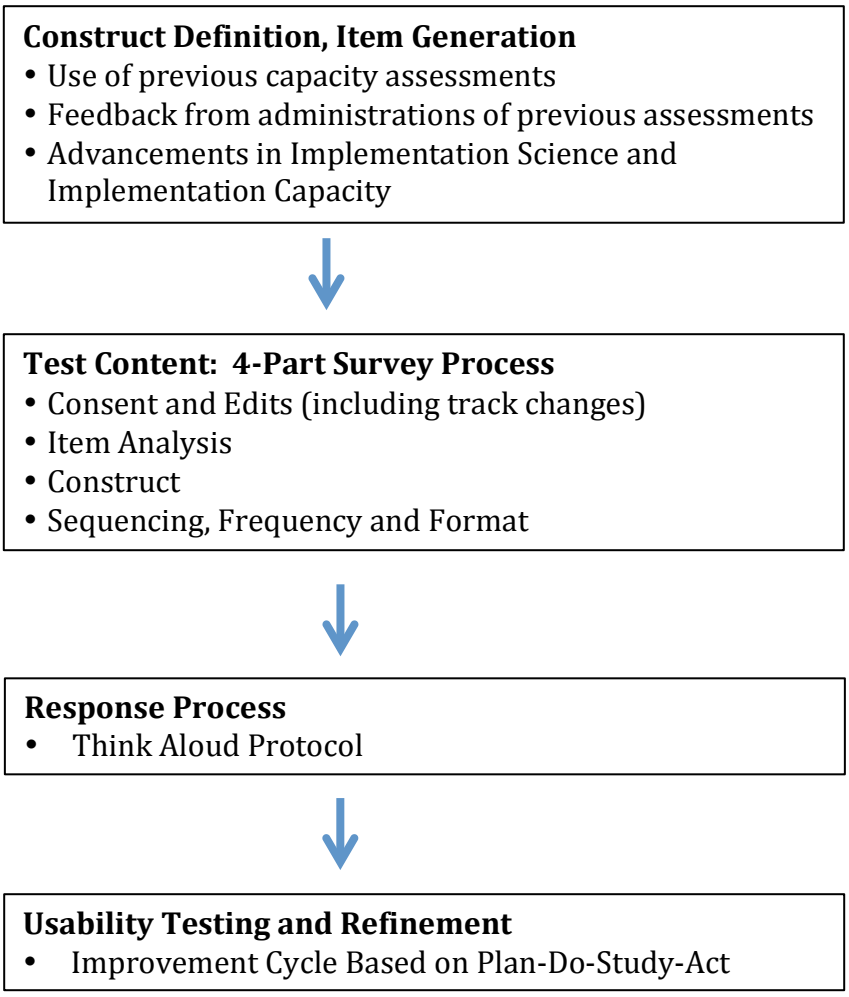
A critical first step in instrument development is gathering evidence of how soundly the test content measures the construct. This should be gathered as a part of the test development process. Gathering evidence of test content establishes the appropriateness of the conceptual framework and how well items represent the construct (Sireci & Faulkner-Bond, 2014). Test content validity is considered an important feature when developing an instrument because it represents the extent to which items adequately sample the construct (Gable & Wolf, 1993, Beck & Gable, 2001). Other forms of validity and reliability do not carry as much weight without first establishing strong test content validity. Scales and items that are poorly developed can have an impact on whether the assessment is biased, flawed, or otherwise not designed in a way to elicit quality responses leading to a sound measure of the construct at hand. Consequently, the quality of the construction of the DCA hinges on important content factors, such as how well the instructions are written, how clearly items are phrased, and the format and appropriateness of the scale that is used.

Following test content evaluation and subsequent editing of the assessment based on the results, it is beneficial to ensure that participants interpret the instrument as expected. Evidence of the response process is determined by the extent to which participant responses are aligned with the intended interpretation of scores (Smith & Smith, 2007). The purpose is to observe participant performance strategies and responses, such as how participants approach and analyze particular items. This enables investigators to rethink or format items that have been misinterpreted, thus removing any items that do not represent the construct (Standards, 2014, p.12).

Focus and Process of Current Validity Work

Test content and response process elements of validity, along with usability testing, were accomplished through a multi-phase process using a multi-method approach collecting both qualitative and quantitative responses. Following the initial development of the DCA items and scoring rubric in the fall of 2014, a 4-part survey was developed to collect feedback from experts and practitioners regarding the assessment during November of 2014. In December of 2014, think aloud protocols were completed, and through the winter of 2015 usability testing was completed. Following each phase of work with the DCA, the assessment was refined based on the feedback and information gathered. The results of each stage and the modifications made are discussed in further sections of this technical report.

Figure 1. Phases of Initial Development, Validity, and Usability



Initial Development of the District Capacity Assessment

Construct Definitions

The DCA is designed to measure practices as operationalized within Active Implementation Frameworks, meaning that the school district works to provide specific supports for a program both to benefit end users and sustain practices over time. Supports assessed are those discussed within implementation science research. Terminology such as effective innovation, capacity for implementation, and implementation drivers are all concepts embedded within the assessment and must be well understood by those interacting with the tool. Martinez, Lewis and Weiner (2014) point out that current language and definitions used within implementation

science are not consistent, leading to variance in how constructs are described within research articles and instruments assessing implementation. In an area of research, such as implementation science, where the lack of maturity of the content area leads to variance in the uses of critical terms, it is essential that constructs presented are well stated and visible within the assessment. Constructs used within the DCA were defined to align with the National Implementation Research Network's (NIRN) Active Implementation Frameworks and definitions (see Table 1).

Items and Rubric

Items from previous versions of the DCA and the DMCA were used to set the stage for initial item generation. Factors influencing item generation included: a process of comparing similar items between assessments; careful consideration of items found on only one assessment; use of feedback collected from administrators and facilitators of previous DCA administrations, and recent advances in the field of implementation science. Items included in previous measures were deleted when deemed inappropriate or ineffective and new items were created to fill gaps within the assessment. Careful consideration was given to features outlined by Haynes (1995) and DeVellis (2012) with attention to how well each item reflected the scale's purpose, decreasing redundancy within the assessment, reading difficulty level, length of an item, avoiding multiple negatives, double barreled items, confusing language, and negatively versus positively stated wording.

Item generation concluded with 28 items in the assessment including a scoring rubric for each item reflecting "Fully in Place", "Partially in Place" or "Not in Place." Each item was categorized within one of the Implementation Drivers of Leadership, Organization and Competency. Accompanying introductory sections, instructions, and tools for administration and scoring were developed to support the appropriate use of the tool. The sections include: Introduction and Purpose, DCA Administration Fidelity Checklist, DCA Scoring Form, Action Planning, and Glossary.

Content Validation Process: 4-Part Survey Protocol

General Survey Development

A content validity survey was developed to gather feedback on four components: the importance/relevance of each DCA item; the attainability of each item; definitions of terms and constructs, sequencing, frequency; and format. Designing the survey into four components provided shorter and more manageable segments of work for participants; which minimized the risk of participant fatigue. Separation of critical aspects of the validation process also aided in the analysis of results.

Test Content Participants

The number of participants suggested for a content validation survey varies from 2-20 (Gable & Wolf, 1994; Grant & Davis, 1997; Lynn, 1986; Tilden, Nelson, & May, 1990; Waltz, Strickland, & Lenz, 1991). What is important is that the end group of participants is representative of the range of experience, background and expertise that is desired for a full review of the assessment. The DCA content validity survey results included feedback from 34 participants. Initially, 56 individuals received the request for survey participation resulting (57% response rate). Individuals approached to participate met one of the following criteria:

1. A researcher with at least one publication in the area of implementation science;
2. Staff member with NIRN who provided national technical assistance related to implementation science;
3. Staff from Michigan's Integrated Behavior Learning Support Initiative (MIBLSI) or partners of NIRN (e.g., staff from different partnering states and district) who provided technical assistance to the implementation of effective innovations at the state or regional levels; or
4. School district practitioners directly involved in the training and/or coaching structure for district implementation teams (DITs) within a SISEP active state or a partnering MIBLSI district.

Table 3. Test Content Validation Participants

Research/National Technical Assistance Providers	State/Regional Technical Assistance Providers	District Practitioners	Total Number
4	19	11	34

A large response pool was desired to serve multiple functions. Putting the DCA in front of individuals who had used or administered previous versions of the DCA gave those who had been involved with early capacity assessment work an opportunity to build an understanding of the proposed changes. The DCA developers also valued concurrent feedback from those who had not previously interacted with an assessment of this type. Additionally, the larger participant group allowed input from those who facilitate, train and support district

implementation teams to be examined along with practitioners on district teams who serve in a broad range of educational roles.

Participants who were not a part MIBLSI or NIRN staff were offered a stipend of $200 for their participation in the content validation process if this task fell outside of the scope of their work responsibilities or if they needed to allocate time outside of their normal work schedule. Compensation was provided to their employer if survey completion fell within the scope of their work responsibilities, occurred during work hours and participants had permission from their employer. Alternatively, a participant could offer his or her services in kind with no payment exchanged. Involvement as a reviewer provided participants with a unique opportunity to preview the DCA, shape the next version of the assessment and be recognized as DCA contributors.

Content Validation Survey Elements

An array of questions can be asked to elicit feedback from participants within a content validation survey. The most consistently addressed portion of a content validation survey is the rating of items in areas such as relevance and clarity. Haynes, Richard and Kubany (1995) suggest including all sections of the assessment within content validation. This includes: instructions, response formats and response scales, relevance and representativeness along with probing respondents to share what inferences they believe will be able to be drawn from the information gathered after the assessment has been completed. As an additional support, Haynes et al. outline a number of elements that may be relevant for a content validation survey. It is stated that not all questions may be relevant for all assessments, but that intentional consideration of the suggested elements should help inform the development of a content validation survey. Table 4 outlines test content elements suggested by Haynes et al. Table columns outline whether the element was considered appropriate for the content surveys related to the DCA and the survey in which included elements are addressed. Table 5 lists each survey and what components were included within that specific survey. A copy of each survey is included in Appendix A.

Table 4. Test Content Elements

Elements Suggested by Haynes and Included in the Survey	Test Content Validation Survey			
	Consent and Edits	Item Analysis	DCA Construct	Sequencing, Frequency, Format
Precision of wording or definition of individual items	X	X		
Item response form (e.g., scale)	X			
Temporal parameters of responses	X			
Instructions to participants	X			X
Method and standardization of administration	X			X
Components of an aggregate, factor, response class			X	
Definition of domain and construct			X	
Sequence of items or stimuli				X
Function-instrument match				X
Elements Suggested by Haynes but not Considered Necessary for the Surveys: • Array of items selected (questions, codes, measures) • Situations sampled • Behavior or events sampled • Scoring, data reduction, item weighting • Method-mode match				

Table 5. Test Content Validity Survey Components

Survey	Components Included in the Survey
Consent and Edits	• Consent form • Opt in/out of listing as a DCA contributor • Downloadable PDF of DCA • Upload DCA with edits, suggestions, questions provided through track changes
DCA Item Analysis	• Attainability of each DCA item rated on a 3-point scale • Importance of each DCA item rated on a 3-point scale • Opportunity to select the 5 most critical DCA items
DCA Constructs	• Comprehensiveness of each DCA construct definition rated on a 3-point scale • Clarity of each DCA construct definition rated on a 3-point scale • Open-ended comments on construct definition • Indication of the best fit for each DCA item within a subscale
Sequencing, Frequency, Format	• Suggestions for reordering DCA items • Frequency DCA should be administered • Comprehensiveness of each DCA section rated on a 3-point scale • Clarity of each DCA section rated on a 3-point scale • Open-ended comments on sections of the DCA • If participant had experience administering a previous version of the DCA or another capacity assessment asked: 1) whether current version of the DCA is an improvement from previous versions and 2) to give input on what benefits have been experienced from using the DCA or DMCA in the past

Each segment of the content validation survey began with a welcome statement and a short video outlining how to interact with that specific segment of the survey (length of video ranged from approximately 1.20 min to 3 min). At the conclusion of each survey segment, a question was posed asking participants to report how long (in minutes) it took to complete the survey, along with a thank you page containing the link to the next segment of the survey. Table 6 shares the average duration for each segment of the survey along with the average total completion time for all four surveys. Within the first survey, participants were asked first to read the current version of the DCA and make track changes within a word document denoting questions, suggestions for rewording, re-ordering, etc. Following the initial read and track changes, participants completed the remaining three sections of the 4-part survey. A 3-point Likert scale response with anchors of “Very”, “Somewhat” and “Not at All” coded as 3, 2, or 1 respectively for analysis purposes was provided as response options for appropriate items on the content validity survey.

Table 6. Minutes Spent Completing Test Content Validation Survey

Consent and Edits	Item Analysis	Construct	Sequencing Frequency and Format	Total
Average 89	26	23	20	157
Range (23 – 200)	(10 - 60)	(5 – 75)	(6 – 60)	(40 – 275)

Test Content Survey Description

Consent and Edits

A unique aspect of the DCA content validity survey included an opportunity to provide feedback, questions, and comments directly within the DCA assessment itself. After watching a video that provided training in how to utilize review/track changes options in Microsoft Word, participants were directed to download a word version of the DCA and then provide their feedback and suggestions directly to the downloaded DCA. Once the review was completed, they were asked to upload their comments and track changes of the DCA to the electronic content validation survey so that the DCA developers had access to their comments and track changes. This review activity enabled participants to focus and provide their feedback to certain sections, items, or even specific sentences within the DCA.

The goal of this atypical process was to have participants working within the DCA as they were providing feedback, rather than solely collecting feedback within an online survey. Participants were asked to pay particular attention to the scoring rubric, as having a rubric was a new addition to the DCA.. The emphasis on gathering feedback in this detailed way right within the document provided DCA developers with a significant quantity of comments, questions and edits through track changes to assist with improving clarity and improving the quality of the detailed scoring rubric. The developers also hypothesized that the process of providing comments and track changes directly into the DCA would result in participants having greater familiarity with the DCA; thus improving the quality of their feedback in subsequent segments of the content validation survey.

Item Analysis

Following the completion of the comments and track changes edits to the DCA, participants were next directed to a survey where they had an opportunity to rate the attainability and importance of each item included within the DCA. A three-point Likert scale was provided to quantify this information. Additionally, after rating the items, participants were asked to select what they believed to be the top five most critical items included in the DCA. The purpose of this step was to help to further discern which items participants viewed as critically important.

DCA Construct

To assess the clarity and comprehensive nature of specific definitions used for constructs included within the DCA, each construct definition was provided with an opportunity for participants to give feedback through both quantitative and qualitative means. Specifically, participants were asked to rate on a 3-point scale the comprehensiveness as well as the level of clarity of each construct definition. Finally, comments were collected about definitions provided.

Content validation also offers an opportunity to gather information from respondents on whether there is a strong match between the item and the construct under which it is clustered. Participants were asked to indicate for each item within the DCA which

Implementation Driver: Leadership, Organization, or Competency, the item best fits within based on the definition of the driver. This information was used to assist in final mapping of DCA items into Implementation Drivers, or domains, within the assessment.

Sequencing, Format and Frequency

Considering the flow of moving from item to item, participants were asked if the had any suggestions for changes to the order of the items. If yes, they were asked to provide their sequencing suggestions as well as any other comments in an open-ended format. Frequency of administration also was addressed through a series of questions related to how often the DCA should be used by a district team, the latency of time in the cycle from action to improvement, and the frequency with which this type of assessment would assist with problem solving. As with most other items included in this survey, open-ended response was also available for items related with frequency.

In an effort to dive deeply into the quality of the sections within the DCA including the Introduction and Purpose, Administration and Fidelity Checklist, DCA Scoring Form, Scoring Guide, Action Planning, and Glossary, questions about comprehensiveness and clarity were included. Respondents again used a 3-point Likert scale to respond and had space for open-ended comments.

The content validation survey concluded with a brief question specifically for those who had experience with a previous version of the DCA or the DMCA. If a participant had experience with these tools, he or she was asked two open-ended questions. These items asked about perceived benefits of the revised DCA in comparison to previous administrations. Participants then rated, on a sliding scale from 0-10 with 0 representing “Not an Improvement” and 10 indicating “Significant Improvement”, whether the current version of the DCA was an improvement from the assessment with which they had previously worked.

Analysis of Content Validity Survey Results/Decision Rules

A variety of quantitative and qualitative responses were elicited throughout the test content survey. Quantitative and qualitative responses were organized together across areas of the DCA in an effort to triangulate data and enhance decision making. Qualitative responses from the open-ended questions within the survey were combined with comments, edits, questions and suggestions from the track changes documents provided by each participant.

Prior to analysis of the DCA results, decision rules were developed and agreed upon by the DCA developers. Decision rules support an unbiased use of results. All comments, edits, questions and suggestions from survey results and the track changes documents were read and considered by developers. However, the level of editing and changes that were employed was mediated by quantitative results. Items using a Likert rating, such as a 3- or 10-point scale, were analyzed using a content validity index (CVI) score for each item. A CVI of at least 0.80 is considered to be a good criterion for accepting an item as valid (Davis, 1992). Other qualitative data were reported by number of participants responding a particular way with predetermined cut scores set for analysis.

DCA Test Content Validation Results

Improvement Compared to Other Measures

Seventy-six percent (n=16) of respondents to the DCA test content validation survey had previously completed a similar assessment, such as the DCA or DMCA. On a response scale of 0-10, the new version of the DCA was given an average improvement rating of an eight. Comments provided indicated that the new version was simplified, had shorter and more concise items and was improved by the addition of a scoring rubric. This information was used to confirm the continued work on the new version of the DCA. In fact, a high number of responses not only endorsed this version as an improvement over previous measures, they also endorsed this tool as positively impacting the assessment of implementation capacity.

Construct Definitions

A major task within a content validation study is to determine whether those working within the field accept and support the definitions provided within the assessment. How comprehensive and clear the foundational constructs are affects whether respondents will begin an assessment with a common language and understanding of the foundations of the assessment. Due to the complexity of the DCA constructs and the infancy of the study of implementation science, it was essential to gather feedback on how well the construct definitions were developed.

Within the survey, construct definitions for capacity, leadership, organizational environment and competency were provided. Respondents reviewed definitions within their downloaded copy of the DCA by providing comments and/or track changes in the Consent and Edits portion of the survey. Additionally, within the DCA Constructs segment, respondents were asked to rate the comprehensiveness and clarity of each definition. Definitions that received an average rating below 2.5 would be revised.

Quantitative Survey results indicated no significant revisions were needed for the capacity and competency definitions. Comments provided within the DCA Construct survey and track changes were considered; however few to no changes were made to the definitions. Leadership and Organizational Environment did not fair as well on the quantitative feedback. Due to this finding, comments provided were used to rewrite both definitions.

Table 7. Construct Definition Decision Rules and Results

Construct	Met or Exceeded Comprehensiveness Threshold	Met or Exceeded Clarity Threshold	Revisions
Capacity	Yes	Yes	No Revisions
Competency	Yes	Yes	No Revisions
Leadership	Yes	No	Rewrote definition
Organization	No	No	Rewrote definition

Decision cut point: average rating of less than 2.5 for comprehensiveness or clarity
Decision rule: revise definition based on comments

Frequency of Assessment

An important question in determining the suggested schedule and use of the assessment was addressed in items asking about the frequency with which district teams should use the assessment. Respondents were asked not only about how often to administer the assessment but also about the duration between assessments. In other words, respondents were asked about the time needed for a team to see growth/change in their DCA results along with the timing of teams using the DCA results to inform district-level action planning. The decision cut point was set at more than 70% of respondents suggesting one option for frequency.

Results for this section of the content validation survey were inconclusive. The decision rule cut point was not met. For all questions related to frequency, most results were split between annual assessment and bi-annual assessment of the DCA. While the results were similar between the two options, more respondents shared that a twice a year assessment scheduled would be most beneficial to teams. Comments also indicated that a more rigorous schedule was warranted for such a difficult area of work for teams. Reviewers felt that directing teams to return to the assessment on a regular basis would help keep the teams focused and moving forward on tasks related to capacity for support of effective innovations. The majority recommendation to assess twice annually was adopted.

Reviewers did point out that less frequent assessment may be appropriate at later stages of implementation and/or once a threshold had been met. The DCA development team determined that this feedback could be further considered after additional use of the assessment, usability testing, and when a significant number of teams interacting with the assessment have reached further stages of implementation coupled with above threshold results on the DCA.

Table 8. Frequency of DCA Administration Decision Rules and Results

Frequency	Percent of Respondents Suggesting the Frequency	Decision to Suggest this Frequency
Monthly	0%	No
Quarterly	9%	No
Bi-Annually	50%	Yes: Majority response
Annually	38%	No
Every 2 Years	0%	No
Other	3%	No

Decision cut point: More than 70% of respondents suggest one option for frequency
Decision rule: Use the recommendation as the suggested frequency

Comprehensive and Clear Sections

The DCA provides a wealth of information and support through the inclusion of sections such as the Introduction and Purpose, Scoring Guide, Fidelity Assessment, Glossary, etc. These sections help to ensure a common understanding of why the assessment is completed and provide clear directions on how to complete the administration and scoring process in a standardized way. To ensure that these sections were well developed, survey participants were asked to read through the sections and provide edits, comments, and questions within the track changes document. Respondents also shared feedback regarding the comprehensiveness and clarity of these sections.

All sections included in the DCA met the threshold for comprehensive and clear format and language. DCA developers continued to analyze and attend to the track changes edits to ensure that the highest quality discussion was provided in each section. Significant rewriting was not needed, but based on suggestions, small edits to the content and format did occur.

Table 9. Section Comprehensiveness and Clarity Decision Rules and Results

Section	Average Comprehensiveness Rating	Average Clarity Rating	Decision
Introduction and Purpose	2.9	2.8	No changes
Administration and Fidelity Checklist	3.0	2.8	No changes
DCA Scoring Form	3.0	2.9	No changes
Scoring Guide	3.0	2.8	No changes
Action Planning	2.7	2.6	No changes
Glossary	2.8	2.8	No changes

Decision cut point: Average rating of less then 2.5 for comprehensiveness or clarity
Decision Rule: Revise sections based on comments

Item Analysis

The hallmark of content validation is to ensure comprehensive and clear items. The item analysis portion of the content validation process is the most time consuming and the important aspect. Within the validation process of the DCA, items and item detail included in the scoring rubric were analyzed together to each met high quality standards.

When analyzing the data compiled for each item, developers first considered item ratings on importance and how many reviewers rated the item as one of the top five most important items within the DCA. This information was used initially to determine whether the item would need significant rewriting or only small edits based on suggestions. If an item met the Content Validity Index (CVI) criteria as an important item, DCA developers kept the item and only used comments and edits from the track changes document as a guide for identifying small edits like spelling, grammar, or word order which ultimately led to enhancing the item. If an item was rated low on importance, reviewers considered whether the item was necessary and if so, used feedback from the track changes document to rewrite the item. Information from the attainability rating gave insight into which items reviewers considered difficult for districts to attain. This information was used by the developers to prioritize resources to assist districts in their efforts to develop capacity.

At the conclusion of the item analysis, the DCA developers combined two of the items and deleted one item. Edits to each item were made; including edits to how the item was defined in the scoring guide. The edits made were based on reviewer feedback that was provided through the track changes provided within the assessment tool.

The results of the survey and track changes feedback related to item analysis are organized in Table 10. It is noteworthy that the number of comments, questions, and edits received for an item did not depend on the importance rating. Feedback offered related to improving the quality of the item and therefore the number of suggested edits fluctuated from other ratings. Another key discovery was that the majority of lower ratings for attainability were for items related to the Competency Driver. This clustering of items reviewers considered difficult for districts to attain facilitated action planning for developing future resources.

Table 10. Item Analysis Decision Rules and Results

DCA Item	Average Importance Rating	Average Attainability Rating	Number of Times Rated as a Top 5 Most Important Item
1	3.00	2.97	19
2	2.92	2.86	18
3	2.92	2.88	4
4	2.94	2.85	12
5	2.60	2.52	4
6	2.75	2.67	2
7	2.83	2.31*	7
8	2.94	2.92	14
9	2.86	2.83	9
10	2.69	2.50	7
11	2.50	2.50	0
12	2.67	2.39*	8
13	3.00	2.83	5
14	3.00	2.81	25
15	2.75	2.78	2
16	2.92	2.75	2
17	2.97	2.78	15
18	2.89	2.75	3
19	3.00	2.83	8
20	2.64	2.25*	0
21	2.42* Deleted	2.11*	0
22	2.81	2.39*	2
23	2.56* Combined with item #20	2.17*	0
24	2.81	2.47*	1
25	2.69	2.39*	4
26	2.83	2.33*	4
27	2.75	2.22*	1
28	2.86	2.44*	4
	Decision cut point: Content Validity Index below 2.5 (CVI) Decision rule: *Eliminate or substantially change the item	Decision cut point: CVI below 2.5 Decision Rule: *Develop an action plan to create resources to assist teams with action planning and attaining item	Used to further validate CVI rating

Item Match with Constructs

Matching items to constructs of the assessment helps to ensure that all constructs are a match to the overall assessment. Construct matching can assist in setting up item mapping to subscales of the assessment. In the case of the DCA, reviewers were not consistently matching items to the constructs identified by the authors. This may also have been mediated by the variety of definitions used within the field of implementation science and the novelty of the concepts to some practitioners. For these reasons, the DCA development team determined that developers who were most knowledgeable about the implementation science constructs would map the items using the comments provided by reviewers as an additional guide/resource.

Table 11. Item Match with Constructs Decision Rules and Results

Greater than 70% of Respondents Aligned Item to the same construct	50-70% of Respondents Aligned Item to Same Construct	Less than 50% of Respondents Aligned Item to Same Construct
3 Items: 2, 13, 22	20 Items: 1, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 17, 18, 19, 20, 21, 24, 25, 26	3 Items: 11, 23, 27

Decision cut point: Less than 70% of respondents align an item to the same construct
Decision Rule: Authors will use results, comments and personal knowledge of the constructs to map an item to a construct

Sequencing of Items

The large majority of reviewers reported that the current order of the items was sufficient. Due to the feedback only minor edits were made and these were based on the edits made during the item analysis and/or a comment made by a reviewer that did suggest an item change. Overall, the order of the items seemed appropriate for the tool and was adequately sequenced for district teams to complete the assessment. The decision rule for reordering items was if more than 50% of respondents suggested moving an item. 77% of reviewers suggested no reordering of items. A few items were reordered based on reviewer comments and due to edits to the assessment items.

Response Process: Think Aloud Protocols

Response Process Overview

Response process is used as a part of the validation process to collect further evidence of the alignment between assessment purpose and directions and resulting thinking and acting by those using the assessment. Alignment between participant responses and intended interpretation of the assessment are evaluated (Smith & Smith, 2007). While response process is not historically a widely used source of validity, it is highlighted as a critical element of

validation within the Standard for Education and Psychological Testing (American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education, 2014).

To gather these data, participants are expected to verbally report their thoughts during a section-by-section and item-by-item walk through of the assessment. The objective is to capture participants' cognitions, performance strategies, thoughts, feelings, beliefs, and experiences as they respond to assessment items. Johnstone, Bottsford-Miller, and Thompson (2006) point to this process as a way to gather valuable information around potential assessment design problems that may lead to inconsistency in how items or assessment directions are interpreted. This step enables assessment developers to rethink or reformat items that have the potential to be misinterpreted increases internal validity of the assessment.

General Protocol Development

To more deeply understand a participant's thoughts while working through response process cognitive interviews, otherwise known as think aloud protocols, can be used. While working through a think aloud protocol, participants share aloud what they are thinking, doing and feeling as they engage in an assessment. In an effort to standardize the think aloud process and add ease to the collection of feedback, a Think-Aloud Protocol Guide (TAP Guide) was developed to collect this information while participants progressed through the DCA. The TAP Guide is intended to be an efficient strategy for gathering evidence of validity. The TAP Guide is a script used during a think aloud containing several best practices promoted by researchers who have developed methods to standardize the observation and recording of verbal reporting data (Conrad, Blair, and Tracy, 1999; Conrad & Blair 1996, Willis, 1999). The TAP Guide is included in Appendix B.

The TAP Guide includes scripted instructions, a practice phase, and clear instructions on how someone administering the think aloud protocol should respond to participant input. The TAP Guide begins by briefly explaining the purpose of the response process. Then it describing the work he or she will do while reading the assessment aloud. Reviewers will voice everything that comes to mind as he or she verbally answer each of items. To acclimate reviewers to voicing aloud what comes to mind as they complete the assessment, a practice phase is conducted; which provides the participants an opportunity to practice.

During the think aloud protocol, DCA content developers collected qualitative data in real time along with occasionally probing reviewers to encourage further dialogue about what came to their mind as they read the assessment aloud. At the conclusion of the think aloud, protocol follow-up questions are used as an opportunity to address reviewer questions that arose during the think-aloud protocol, ask clarifying questions regarding specific items and directions, and summarize reviewers' general impressions of the assessment.

Response Process Participants

Willis (1999) suggests that recruitment of participants should emphasize diversification based on characteristics of interest that will support a variety of viewpoints providing feedback on the assessment. Large sample sizes are not required because the purpose is not statistical estimation; rather, qualitative analysis. Within individual interviewing procedures, Virzi (1992) recommends the use of four or five participants, which has been shown to adequately uncover 80% of the construct-irrelevant variance.

For this aspect of the validation process, four participants were identified. Efforts were made to select individuals that have either differing roles in supporting district implementation or have various levels of experience in using previous iterations of district capacity assessments. Roles included an administrator, school psychologist, and an MTSS Coordinator representing a variety of experiences with implementation work. Within the reviewer group two individuals had worked closely with MIBLSI to implement an effective innovation and two others had more ancillary participation in the MIBLSI work.

Each participant provided feedback through a one-on-one meeting with one of the DCA developers using the described TAP Guide. One participant reviewed the entire assessment from start to finish. In the interest of receiving high-quality feedback without fatiguing the reviewers, one additional respondent reviewed the Introduction and Purpose and DCA Administration Fidelity Checklist; while, two additional participants reviewed the Scoring Guide, which also included a review of the DCA items. All participants were asked to refer to the glossary as needed and when this occurred reviewers were asked to give feedback on the portion of the glossary that they accessed.

Response Process Results and Modifications to the Measure

The time needed to complete the response process varied from one reviewer to another due to the variety of sections upon which each reviewer provided feedback. On average, this process took two to three hours per reviewer. Results were documented in the notes section of the TAP Guide by capturing, whenever possible, what the reviewer said verbatim. Those administering the Think Aloud Protocol did so using prescribed directions within the protocol.

The response process results were analyzed and acted upon by the developers following the completion of the think aloud procedures. Qualitative results were summarized and actionable feedback was shared with the group for consideration. For this portion of the content validation process, no significant changes to the DCA were necessary; however, minor improvements were made (e.g., item and scoring guide re-wording) in an effort to improve the clarity of the assessment. The response process was considered valuable as it highlighted difficult to read sentences and inconsistencies in language, and wording that could be interpreted multiple ways. Comments and suggested edits were used within a final editing process to ensure consistency and clarity in wording across the DCA.

Usability Testing: Continuous Improvement Process

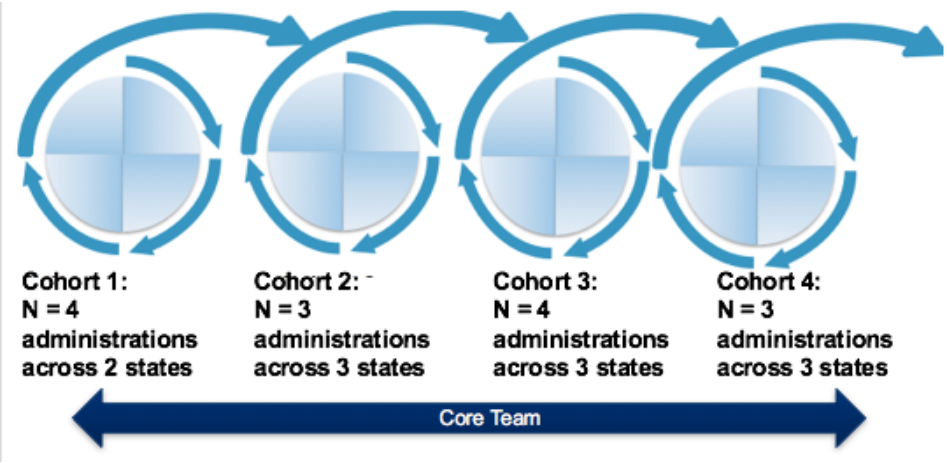
Usability Testing Overview

Usability Testing was completed to test the feasibility of the assessment and administration processes. Usability Testing is a planned series of improvement cycles (Plan-Do-Study-Act Cycles). Specifically, small cohorts of DCA administrations (N= 4-5) were completed in four intentional improvement cycles (see Figure 2). The goal of usability testing is to progressively improve the administration and scoring process by identifying and addressing challenges encountered before broadly using the assessment. The key to Usability Testing is having a team that:

- Plans - Leads the improvement planning process and develops the scope of the test for use of the assessment
- Does - Engages in using the assessment as outlined in the planning phase
- Studies – After each data collection cycle, the team studies what is working (or not) using data
- Acts - Identifies actions the team will take and implements those actions in another data collection cycle with a different cohort

By engaging in four to five improvement cycles, approximately 80% of the problems with the assessment itself can be eliminated (Nielsen, 2000). This improves the administration and scoring experience of those using the assessment (for more information on usability testing: http://implementation.fpg.unc.edu/module-5/topic-2-usability-testing).

Figure 2. Usability Testing



Usability Testing Plan

The usability testing plan was comprised of several components including 1) selection criteria for determining who would participate in each cohort of a cycle; 2) scope of the test including elements to study; 3) criteria for success or reconciling information, and 4) processes for data collection. Details of the usability testing plan undertaken for the DCA are outlined in Table 12.

In each improvement cycle, 3-4 District Capacity Assessments were completed within two to three State Education Agencies including Michigan, Washington, and North Carolina. Districts varied in size (e.g. urban, rural), need, and demographics.

Table 12. Usability Testing Plan

1. Selection Criteria	**Administrators:** • Trained on the administration of the DCA • NIRN staff or MIBLSI staff **Sites for Administration:** • Districts actively engaged in installing an effective innovation • Districts actively engaged in developing their capacity to support effective use of innovations in partnership with MIBLSI or NIRN • Districts with executive leadership support for use of the DCA

2. Scope of Test	**Areas of Study:**
	• Communication and preparation for the DCA administration process
	  ○ Obtaining commitment to the time for the administration
	  ○ Obtaining leadership support for administration
	  ○ Identifying appropriate respondents
	• Administration protocol with fidelity
	• Items and scoring rubric
	  ○ Clarity of language
	  ○ Sequencing of items
	• Participant Responsiveness
	• Training Implications

3. Criteria for Success	**Communication and Preparation** • Commitment obtained to the time and leadership support as evidenced by 90% or more of respondents staying or the entire length of administration • Appropriate respondents were identified for administration **Administration Protocol with Fidelity** • 90% or greater of items on administration protocol reported to be completed **Items and Scoring Rubric** • Fewer than 10 items needing revision for clarity purposes • Less than 2 changes in sequence **Participant Responsiveness** • Majority of respondents (80%) reported to be engaged and positive
4. Process for Data Collection	At the end of improvement cycle, the development team met to review the results of the administrations including DCA results including length of administration, respondents, adherence to fidelity checklist results, and qualitative feedback from administrators.

Usability Testing Results and Modifications to the Measure

The number of improvements identified for the different areas (e.g., communication, administration protocol, items and scoring rubric, participant responsiveness, and training implications) decreased by the end of the fourth improvement cycle. In addition, the criteria for success for each of study was met by the end of the fourth improvement cycle. Examples of improvement for different areas studied and acted upon are listed below in Table 13.

Table 13. Areas Identified for Improvement Based on Usability Testing

Area	Examples of Improvement
Communication & Preparation	• More guidance developed around team composition and respondents • Length of time requested for administration increased from 1.5 hours to 2 hours to include time for review of results • One page handout introducing the DCA was developed for communication purposes
Administration Protocol	• 100% adherence to the administration protocol on administrations
Items & Scoring Rubric	• Minor wording changes to a total of 4 items (3 in cycle 1, 2 in cycle 2); • Sequencing of items was reviewed but not changed
Training Implications	• Facilitation skills identified and training materials refined • Process developed for the prioritization of areas for action planning using DCA results
Participant Response	• Engaged and positive throughout all administrations within 4 improvement cycles

Preliminary Reliability Results

Descriptive Statistics

One hundred and ninety-five DCAs within 18 states have been conducted since the release of the instrument. A series of descriptive statistics, initial reliability analyses, and exploratory factor analyses were conducted. Results of these analyses are presented below. Given the low scores and little variability in several of the items, it was determined a flooring effect is currently occurring for those items such as the items comprising the coaching subscale. Before making any significant changes to the items themselves and rescaling the instrument, the developers determined additional data was needed.

Table 14. DCA Scale and Subscale Descriptive Statistics (*N*=195)

Scale/Subscale	*M*	*SD*	Range
Total DCA Score	22.34	9.85	0-48
Leadership Composite	9.67	3.76	0-16
Competency Composite	5.10	3.32	0-16
Organization Composite	7.57	4.09	0-19
Leadership Subscale	7.17	2.32	0-10
Planning Subscale	2.50	2.02	0-6
Performance Assessment Subscale	1.75	1.05	0-4
Selection Subscale	1.47	1.11	0-4
Training Subscale	1.37	1.17	0-4
Coaching Subscale	0.51	0.99	0-4
Decision Support Subscale	2.32	1.87	0-6
Facilitative Administration Subscale	4.23	2.34	0-11
Systems Intervention Subscale	1.03	0.80	0-2

Table 15. DCA Item Descriptive Statistics (*N*= 195)

Item	*M*	*SD*
Leadership Subscale		
1. There is a DIT to support implementation of EI.	1.63	0.66
2. DIT includes someone with executive leadership authority	1.58	0.69
3. DIT includes an identified coordinator(s)	1.20	0.77
7. Funds are available to support the implementation of the EI	1.47	0.65
17. BITs are developed and functioning to support implementation of EIs	1.29	0.68
Planning Subscale		
8. District has an implementation plan for the EI	0.76	0.74
9. DIT actively monitors the implementation of the plan	0.60	0.75
18. BIT implementation plans are linked to district improvement plan	1.14	0.90
Performance Assessment Subscale		
13. DIT supports the use of a fidelity measure for implementation of the EI	1.11	0.79
26. Staff performance feedback is on-going	0.64	0.61
Selection Subscale		
20. District uses a process for selecting staff (internal and/or external) who will implement and support the EI	0.67	0.67
21. Staff members selected to implement or support the EI have a plan to continuously strengthen skills	0.80	0.67
Training Subscale		
22. DIT secures training on the EI for all district/school personnel and stakeholders	0.95	0.72
23. DIT uses training effectiveness data	0.42	0.66
Coaching Subscale		
24. DIT uses a coaching service delivery plan	0.34	0.61
25. DIT uses coaching effectiveness data	0.17	0.45
Decision Support Data Systems Subscale		
14. DIT has access to data for the EI	1.06	0.73
15. DIT has a process for using data for decision making	0.61	0.74
19. BITs have a process for using data for decision making	0.66	0.71
Facilitative Administration Subscale		
4. DIT uses an effective team meeting process	0.97	0.72
5. District outlines a formal procedure for selecting EIs through the use of guidance documents	0.42	0.57
6. District documents how current EIs link together	0.65	0.66
10. District utilizes a communication plan	0.60	0.60
11. District uses a process for addressing internal barriers	0.83	0.48
16. District provides a status report on the EI to the school board	0.76	0.72
Systems Intervention Subscale		
12. District uses a process to report policy relevant information to outside entities	1.03	0.80

Note. All item scores range from 0-2. DCA= District Capacity Assessment; DIT= District Implementation Team; EI= Effective Innovations; BIT= Building Implementation Team.

Bivariate Correlations

Pearson product-moment correlation coefficients were conducted to determine relationships between the District Capacity Assessment (DCA) subscales. This analysis determined that the majority of subscales (Leadership, Planning, Performance Assessment, Selection, Training, Coaching, Decision Support, Facilitative Administration, Systems Intervention) are significantly correlated with each other (see Table 3). The only exceptions are: Leadership and Coaching (r= -.002, n= 195, p > .05), Systems Intervention and Training (r= .089, n= 195, p > .05), and Systems Intervention and Coaching (r= .113, n= 195, p > .05). It is important to note that the mean score for Coaching (M= .51) is lower than other subscale mean scores, which may be contributing to the aforementioned exceptions.

Table 16. Bivariate Correlations Between DCA Subscales

Subscale	1	2	3	4	5	6	7	8
1. Leadership	--							
2. Planning	.504**	--						
3. Performance Assessment	.500**	.633**	--					
4. Selection	.321**	.470**	.501**	--				
5. Training	.280**	.529**	.465**	.591**	--			
6. Coaching	-.002	.269**	.287**	.384**	.449**	--		
7. Decision Support Data System	.393**	.775**	.617**	.518**	.522**	.314**	--	
8. Facilitative Admin.	.526**	.682**	.529**	.515**	.536**	.274**	.632**	--
9. Systems Intervention	.181*	.196**	.148*	.154*	.089	.113	.146*	.318**

Note. DCA= District Capacity Assessment.
** p < .001 (2-tailed)
* p < .05 (2-tailed)

Cronbach's Alpha Coefficients

Cronbach's alpha coefficients were conducted to determine internal consistency of the District Capacity Assessment (DCA), DCA composite scales, and DCA subscales. The total DCA has strong internal consistency with a Cronbach's alpha coefficient of .908. The three composites also have adequate internal consistency: Leadership (α= .794), Competency (α= .791), and Organization (α= .805). The eight subscales vary in internal consistency. The Planning subscale (α= .797), Coaching subscale (α= .832), and Decision Support Data Systems subscale (α= .818) all have adequate internal consistency. The Leadership subscale (α= .689), Performance Assessment subscale (α= .224), Selection subscale (α= .563), Training subscale (α= .606), and Facilitative

Administration subscale (α= .678) are all below the adequate level of internal consistency (i.e., .700). See Table 4 for all results from these analyses.

Table 17. Cronbach's Alpha Coefficients

Scale/Subscale	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	Number of Items
DCA	.908	.908	26
Leadership Composite	.794	.794	8
Competency Composite	.791	.802	8
Organization Composite	.805	.811	10
Leadership Subscale	.689	.687	5
Planning Subscale	.797	.808	3
Performance Assessment Subscale	.224	.230	2
Selection Subscale	.563	.563	2
Training Subscale	.606	.607	2
Coaching Subscale	.832	.854	2
Decision Support Data Systems Subscale	.818	.818	3
Facilitative Administration Subscale	.678	.690	6

Note. DCA= District Capacity Assessment

Exploratory Factor Analysis

The 26 items of the District Capacity Assessment (DCA) were subjected to principle components analysis (PCA). Prior to interpreting the PCA results, the suitability of data for factor analysis was assessed by inspecting the Kaiser-Meyer-Oklin (KMO) value and Bartlett's Test of Sphericity. The KMO value (.865) met the necessary level to run a factor analysis, and Bartlett's Test of Sphericity reached statistical significance. The PCA revealed five components with eigenvalues exceeding 1 within the DCA. These five factors explain 32.16%, 9.66%, 6.36%, 5.51%, and 5.05% of the variance respectively. See Tables 17 and 18 for the results. It is interesting to note that the DCA was developed with nine different factors in mind, however only five different factors were identified in this analysis. Factor one consists of those items related to data and use of data for action planning at both district and building levels. Using data for evaluation competency supports (e.g. training and coaching) was prevalent theme of Factor 2. The third factor was comprised of items that address leadership coordination and team functioning. Items addressing competency and facilitative administration activities and

supports comprised the fourth factor. Finally, the fifth factor consisted of those times related to policy, communication, reporting, and other systems intervention supports.

The DCA developers determined that additional data are needed to further analyze the internal structure of the instrument before making significant changes to the items themselves and the composition of the subscales. The basis of this decision was made on factors such as theory the DCA is built upon, the flooring effect observed in low scores across several items due to the current state of the field, and sample size.

Table 18. Five Factor Solution Simplified

Factor	Items loaded into factor (DCA Subscale)	Factor Loadings
1	19 (DSDS)	.832
	18 (Planning)	.818
	17 (Leadership)	.714
	14 (DSDS)	.698
	15 (DSDS)	.688
	8 (Planning)	.651
	9 (Planning)	.583
	13 (Performance Assessment)	.560
	16 (Facilitative Administration)	.443
2	25 (Coaching)	-.856
	24 (Coaching)	-8.49
	23 (Training)	-.662
	21 (Selection)	-.494
	26 (Performance Assessment)	-.486
3	2 (Leadership)	.815
	1 (Leadership)	.770
	3 (Leadership)	.745
	4 (Facilitative Administration)	.693
4	7 (Leadership)	.711
	20 (Selection)	.533
	22 (Training)	.473
	11 (Facilitative Administration)	.434
5	12 (Systems Intervention)	.739
	5 (Facilitative Administration)	.666
	10 (Facilitative Administration)	.481
	6 (Facilitative Administration)	.363

Note. Major loadings for each item are bolded. DCA= District Capacity Assessment; DSDS= Decision Support Data Systems; DIT= District Implementation Team; EI= Effective Innovations; BIT= Building Implementation Team. DCA= District Capacity Assessment; DSDS= Decision Support Data Systems.

Current and Future Uses of the District Capacity Assessment

Appropriate Use of the DCA

As with all assessment instruments, there are appropriate uses of the District Capacity Assessment. These include the following:

- District self-assessment and progress monitoring used to guide and improve implementation capacity building
- Coaching for district and building implementation teams on developing of systems, structures, functions, and roles necessary to adopt and sustain implementation of EIs
- Coaching for implementation specialists at the regional and state level on the development of district and building implementation teams to engage in capacity building
- Feedback on materials, resources, and learning tools to support implementation specialists and implementation teams on capacity building
- Research on structures, roles, and functions necessary for effective and sustained implementation of EIs and the associations between these and fidelity measures of the EIs and student outcomes

The DCA should **not** be used as high stakes evaluation tool of a DIT. The DCA's validity and reliability is still being assessed. Its principal purpose is for use as an action assessment to assist districts and their schools to implement evidence based practices that benefit students.

Future Validation of the DCA

Next steps in the development and validation process of the DCA include designing and conducting research to further examine the DCA's internal structure (e.g., Factor Analysis), its relationship to other variables (e.g., Predictive, Concurrent, Convergent, and Divergent Validity analyses), and its consequential validity, that is the intended and unintended consequences of using the DCA. The DCA development team is currently in process of designing the research to address these areas of validation and securing the funds to accomplish this task.

References

Aarons G.A., Cafri G., Lugo L., & Sawitzky A. (2012). Expanding the domains of attitudes towards evidence-based practice: The Evidence Based Attitudes Scale-50. Administration and Policy in Mental Health and Mental Health Services Research, 5, 331-340.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing. (2014). Standards for educational and psychological testing. Washington, DC: AERA.

American Psychological Association. (2010). American Psychological Association ethical principles of psychologists and code of conduct. Retrieved from http://www.apa.org/ethics/code/index.aspx

Beck, C.T., & Gable, R.K. (2001). Ensuring content validity: An illustration of the process. *Journal of Nursing Measurement, 9,* 201–215.

Blase, K.A., Fixsen, D. L., Naoom, S. F., (2005). Operationalizing implementation: Strategies and methods. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute.

Carretero-Dios, H., & Pérez, C. (2007). Standards for the development and review of instrumental studies: Considerations about test selection in psychological research. *International Journal of Clinical and Health Psychology, 7,* 863-882.

Conrad, F. G., & Blair, J. (1996). From Impressions to Data: Increasing the Objectivity of Cognitive Interviews. *Proceedings of the Section on Survey Research Methods, Annual Meetings of the American Statistical Association.* 1–10. Alexandria, VA: American Statistical Association.

Conrad, F. G., Blair, J., & Tracy, E. (1999). Verbal Reports are Data! A Theoretical Approach to Cognitive Interviews. *Proceedings of the Federal Committee on Statistical Methodology Research Conference, Tuesday B Sessions,* Arlington, VA, 11–20.

Davis, L. (1992). Instrument review: Getting the most from your panel of experts. *Applied Nursing Research, 5,* 194–197.

DeVellis, R. F. (2012). *Scale development: Theory and applications.* Thousand Oaks, CA: SAGE Publications.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). Implementation research: A synthesis of the literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231). Retrieved May, 2016 http://nirn.fpg.unc.edu/resources/implementation-research-synthesis-literature

Fuchs, D., & Deshler, D. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research and Practice, 22,* 129-136.

Gable, R., & Wolf, M. (1994). *Instrument Development in the Affective Domain: Measuring Attitudes and Values in Corporate and School Settings.* New York, NY: Evaluation in Education and Human Services.

Grant, J. S., & Davis, L. L. (1997). Selection and use of content experts for instrument development. *Research in Nursing & Health, 20*, 269–274.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods. *Psychological Assessment, 3,* 238-247.

Johnstone, C. J., Bottsford-Miller, N. A., & Thompson, S. J. (2006). *Using the think aloud method (cognitive labs) to evaluate test design for students with disabilities and English language learners.* Minneapolis, MN: National Center on Educational Outcomes.

Klein, K. J., Conn, A. B., Smith, D. B., & Sorra, J. S. 2001. Is everyone in agreement? An exploration of within-group agreement in employee perceptions of the work environment. *Journal of Applied Psychology, 86*, 3–16.

Landenberger, N. A., & Lipsey, M. W. (2005). The positive effects of cognitive-behavioral programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology, 1,* 451–476.

Lynn, M. (1986). Determination and quantification of content validity. *Nursing Research, 35,* 382–385.

Martinez, R. G., Lewis, C. C., & Weiner B. J. (2014). Instrumentation issues in implementation science. *Implementation Science, 9,* 118.

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

Mihalic, S., & Irwin, K. (2003). From research to real world settings: Factors influencing the successful replication of model programs. *Youth Violence and Juvenile Justice, 1,* 307-329.

Nielsen, J. (2000, March 19), *Why you only need to test with 5 users.* Retrieved June, 2016 from http://www.useit.com/alertbox/20000319.html

Olds, D.L., Hill, P.L., O'Brien, R., Racine, D., & Moritz, P. (2003). Taking preventive intervention to scale: The nurse–family partnership. *Cognitive and Behavioral Practice, 10,* 278–290.

Popham, W. J. (2008). All About Assessment / A Misunderstood Grail. *Educational Leadership*, *66,* 82-83.

Schoenwald, S. K., Sheidow, A. J., & Letourneau, E. J. (2004). Toward effective quality assurance in evidence-based practice: Links between expert consultation, therapist fidelity, and child outcomes. *Journal of Clinical Child and Adolescent Psychology, 33,* 94-104.

Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema, 26,* 100-107.

Skiba, R. J., Middelberg, L., & McClain, M. (2013). Multicultural issues for schools and EBD students: Disproportionality in discipline and special education. In H. Walker F. Gresham (Eds.), *Handbook of Evidence-Based Practices for Students Having Emotional and Behavioral Disorders.* New York: Guilford.

Smith, E. V., & Smith, R. M. (2007). *Rasch measurement: Advanced and specialized applications*. Maple Grove, MN: JAM Press.

Tilden, V. P., Nelson, C. A., & May, B. A. (1990). Use of qualitative methods to enhance content validity. *Nursing Research, 39,* 172–175.

Virzi, R. A. (1992). Refining the test phase of usability evaluation: How many subjects are enough? *Human Factors 34*, 457-468.

Waltz, C. F., Strickland O. L., & Lenz E. R. (1991) Reliability and validity of norm-referenced measures. *Measurement in Nursing Research,* 161-194.

Ward, C., St. Martin, K., Horner, R., Duda, M., Ingram-West, K., Tedesco, M., Putnam, D., Buenrostro, M., & Chaparro, E. (2015). District Capacity Assessment. University of North Carolina at Chapel Hill.

Willis, G. (1999). Cognitive Interviewing: A "How To" Guide (Reducing Survey Error through Research on the cognitive and Decision Processes in Surveys) *Meeting of the American Statistical Association* (pp. 1-37). Durham, NC: RTI International.

Appendix A: Content Validation Surveys

DCA Consent and Edits

Welcome

Play the video below to learn about how to interact with Survey #1.

Consent

First and Last Name *

By clicking on the consent checkbox below, I acknowledge that I have read and agree to the terms of participation outlined the ***District Capacity Assessment (DCA) Content Validation Research Consent Form.*** *

- [ ] I consent to participate as an expert reviewer for the District Capacity Assessment.

May we include your name as a DCA reviewer? Your name will appear after the cover page. *

- ○ Yes
- ○ No

Response Time

Please upload your edited version of the District Capacity Assessment. *

Browse... Upload

DCA Item Analysis

Welcome

1. First and Last Name

View the video below to learn about how to interact with Survey #2.

Item Feedback

Each item within the DCA is provided in the left hand column. Read through each individual item and respond regarding:

- How attainable the specific DCA item is for the District Implementation Team
- How important the specific DCA item is for successful implementation of Effective Innovations (EI)
- Which items are the top 5 most critical items to include in the DCA

*

	How attainable is this item for a district team? *	How important is this item for successful implementation of Effective Innovations? *	Select the 5 most critical items to include in the DCA
1. There is a District Implementation Team (DIT) to support implementation efforts of Effective Innovations (EI)	Very Somewhat Not at All	Very Somewhat Not at All	☐
2. The DIT includes someone	Very	Very	

2. The DIT includes someone with executive leadership authority	Very Somewhat Not at All	Very Somewhat Not at All	☐
3. DIT uses an effective team meeting process	Very Somewhat Not at All	Very Somewhat Not at All	☐
4. DIT includes an identified coordinator (or coordinators)	Very Somewhat Not at All	Very Somewhat Not at All	☐
5. District guidance documents outline a formal procedure for selecting EIs	Very Somewhat Not at All	Very Somewhat Not at All	☐
6. District documents how current initiatives/ practice link together	Very Somewhat Not at All	Very Somewhat Not at All	☐
7. Funds are available to support the implementation of EIs	Very Somewhat Not at All	Very Somewhat Not at All	☐
8. District has an implementation plan	Very Somewhat Not at All	Very Somewhat Not at All	☐
9. DIT actively monitors the implementation of the plan	Very Somewhat Not at All	Very Somewhat Not at All	☐
10. The district uses a process for addressing	Very Somewhat Not at All	Very Somewhat Not at All	☐

internal barriers	Not at All	Not at All	
11. District uses a process to report policy relevant information to outside entities	Very Somewhat Not at All	Very Somewhat Not at All	☐
12. DIT uses a measure of fidelity for the use of the EI	Very Somewhat Not at All	Very Somewhat Not at All	☐
13. DIT has access to data	Very Somewhat Not at All	Very Somewhat Not at All	☐
14. DIT has a process for using data for decision making	Very Somewhat Not at All	Very Somewhat Not at All	☐
15. District provides a status report to the school board	Very Somewhat Not at All	Very Somewhat Not at All	☐
16. District utilizes a communication plan	Very Somewhat Not at All	Very Somewhat Not at All	☐
17. Building Implementation Teams (BITs) are developed and functioning to support implementation of EIs	Very Somewhat Not at All	Very Somewhat Not at All	☐
18. BIT implementation plans are linked to district improvement plan	Very Somewhat Not at All	Very Somewhat Not at All	☐

19. BITs have a process for using data for decision making	Very Somewhat Not at All	Very Somewhat Not at All	☐
20. A process is followed for recruiting staff (internal and/or external) to implement the EI	Very Somewhat Not at All	Very Somewhat Not at All	☐
21. A process is in place to evaluate selection outcomes	Very Somewhat Not at All	Very Somewhat Not at All	☐
22. Staff members selected have a plan to strengthen skills necessary for success	Very Somewhat Not at All	Very Somewhat Not at All	☐
23. A process is in place to evaluate selection outcomes	Very Somewhat Not at All	Very Somewhat Not at All	☐
24. DIT secures training for all district/school personnel and stakeholders	Very Somewhat Not at All	Very Somewhat Not at All	☐
25. DIT uses training effectiveness data	Very Somewhat Not at All	Very Somewhat Not at All	☐
26. DIT uses a Coaching service delivery plan	Very Somewhat Not at All	Very Somewhat Not at All	☐
	Very	Very	

27. DIT uses coaching effectiveness data	Somewhat Not at All	Somewhat Not at All	☐
28. Staff performance feedback is perpetual	Very Somewhat Not at All	Very Somewhat Not at All	☐

Time to Respond

Approximately how many minutes did it take you to complete this "DCA Item Analysis" survey?

DCA Construct

Welcome

1. First and Last Name *

Play the video below to learn about how to interact with Survey #3.

Construct Definitions

Provide feedback on the definitions and descriptions that are used within the DCA *

	Is the definition ***comprehensive*** *	Is the definition ***clear*** *	Provide any commen definition / desc
Capacity: organization, activities, and systems that exist at the district level and have a direct effect on the success of building leadership teams to adopt and sustain evidence-based practices.	Very Somewhat Not at All	Very Somewhat Not at All	
Competency: mechanisms			

mechanisms to develop, improve, and sustain one's ability to implement an intervention as intended in order to benefit children, families, and communities.	Very Somewhat Not at All	Very Somewhat Not at All	
Organization: mechanisms to create and sustain hospitable organizational and system environments for effective services.	Very Somewhat Not at All	Very Somewhat Not at All	
Leadership: emphasis on providing the right leadership strategies for types of leadership challenges. These leadership challenges often emerge as part of the change management process needed to make decisions, provide guidance, and support organization.	Very Somewhat Not at All	Very Somewhat Not at All	

Item-Construct Alignment

For each item within the DCA, indicate which Implementation Driver: Leadership, Organization or Competency, that the item best fits within based on the definition of each driver listed below.

Competency: mechanisms to develop, improve, and sustain one's ability to implement an intervention as intended in order to benefit children, families, and communities.

Organization: mechanisms to create and sustain hospitable organizational and system environments for effective services.

Leadership: emphasis on providing the right leadership strategies for types of leadership challenges. These leadership challenges often emerge as part of the change management process needed to make decisions, provide guidance, and support organization.

*

	Implementation Driver *
1. There is a District Implementation Team (DIT) to support implementation of Effective Innovations (EI)	Leadership Organization Competency Unsure
2. The DIT includes someone with executive leadership authority	Leadership Organization Competency Unsure
3. DIT uses an effective team meeting process	Leadership Organization Competency Unsure

4. DIT includes an identified coordinator (or coordinators)	Leadership Organization Competency Unsure
5. District guidance documents outline a formal procedure for selecting EIs	Leadership Organization Competency Unsure
6. District documents how current initiatives/ practice link together	Leadership Organization Competency Unsure
7. Funds are available to support the implementation of EIs	Leadership Organization Competency Unsure
8. District has an implementation plan	Leadership Organization Competency Unsure
9. DIT actively monitors the implementation of the plan	Leadership Organization Competency Unsure
10. The district uses a process for addressing internal barriers	Leadership Organization Competency Unsure
11. District uses a process to report policy relevant information to outside entities	Leadership Organization Competency Unsure
	Leadership

12. DIT uses a measure of fidelity for the use of the EI	Organization Competency Unsure
13. DIT has access to data	Leadership Organization Competency Unsure
14. DIT has a process for using data for decision making	Leadership Organization Competency Unsure
15. District provides a status report to the school board	Leadership Organization Competency Unsure
16. District utilizes a communication plan	Leadership Organization Competency Unsure
17. Building Implementation Teams (BITs) are developed and functioning to support implementation of EIs	Leadership Organization Competency Unsure
18. BIT implementation plans are linked to district improvement plan	Leadership Organization Competency Unsure
19. BITs have a process for using data for decision making	Leadership Organization Competency Unsure
20. A process is followed for recruiting staff (internal and/or external) to implement the EI	Leadership Organization Competency

	Unsure
21. A process is followed for selecting staff who will implement the EI	Leadership Organization Competency Unsure
22. Staff members selected have a plan to strengthen skills necessary for success	Leadership Organization Competency Unsure
23. A process is in place to evaluate selection outcomes	Leadership Organization Competency Unsure
24. DIT secures training for all district/school personnel and stakeholders	Leadership Organization Competency Unsure
25. DIT uses training effectiveness data	Leadership Organization Competency Unsure
26. DIT uses a Coaching service delivery plan	Leadership Organization Competency Unsure
27. DIT uses coaching effectiveness data	Leadership Organization Competency Unsure
28. Staff performance feedback is perpetual	Leadership Organization Competency Unsure

DCA Sequencing, Frequency and Format

Welcome

1. First and Last Name

Play the video below to learn about how to interact with Survey #4.

Order of Items

Read through the items on the DCA considering the order in which the items are currently organized. Consider the flow and ease of moving from item to item for a district team.

Do you have suggestions for changes to the order of questions? *

- Yes, significant reordering (6 or more suggestions)
- Yes, some reordering (5 or less suggestions)
- No reordering suggestions

Provide any other comments about the order of items within the assessment

Sequencing of Questions

Drag items from the left-hand list into the right-hand list to order them.

1. There is a District Implementation Team (DIT) to support implementation of Effective Innovations (EI)
2. DIT includes someone with executive leadership authority
3. DIT uses an effective team meeting process
4. DIT includes an identified coordinator (or coordinators)
5. District guidance documents outline a formal procedure for selecting EIs
6. District documents how current initiatives/ practice link together
7. Funds are available to support the implementation of EIs
8. District has an implementation plan
9. DIT actively monitors the implementation of the plan
10. The district uses a process for addressing internal barriers
11. District uses a process to report policy relevant information to outside entities
12. DIT uses a measure of fidelity for the use of the EI
13. DIT has access to data
14. DIT has a process for using data for decision making
15. District provides a status report to the school board
16. District utilizes a communication plan

17. Building Implementation Teams (BITs) are developed and functioning to support implementation of EIs

18. BIT implementation plans are linked to district improvement plan

19. BITs have a process for using data for decision making

20. A process is followed for recruiting staff (internal and/or external) to implement the EI

21. A process is followed for selecting staff who will implement the EI

22. Staff members selected have a plan to strengthen skills necessary for success

23. A process is in place to evaluate selection outcomes

24. DIT secures training for all district/school personnel and stakeholders

25. DIT uses training effectiveness data

26. DIT uses a Coaching service delivery plan

27. DIT uses coaching effectiveness data

28. Staff performance feedback is perpetual

Provide any other comments about the order of items within the assessment

Clearly state the ordering changes you would suggest. If it is helpful, also include the rationale for this order change.

	Item Number	Description of what needs to change
Suggestion 1		
Suggestion 2		
Suggestion 3		
Suggestion 4		
Suggestion 5		

Provide any other comments about the order of items within the assessment

Frequency of Assessment

Please provide feedback on the 3 items below related to frequency of using the DCA. *

	Monthly	Quarterly	Bi-Annually	Annually	Every Two Years	Other
How often would it would be helpful for a district implementation team to administer this assessment?	○	○	○	○	○	○
How long between assessments would a team likely see growth/change in their DCA results?	○	○	○	○	○	○
How often would the DCA results inform district-level action planning?	○	○	○	○	○	○

Please give more information about what other frequency would be helpful for this assessment

Comments about the frequency with which district teams should complete this assessment

Format

2. Provide feedback on the following sections of the DCA:

	Is the section ***comprehensive***?	Is the section ***clear***?	Comments
Introduction and Purpose	Very Somewhat Not at All	Very Somewhat Not at All	
Administration and Fidelity Checklist	Very Somewhat Not at All	Very Somewhat Not at All	
DCA Scoring Form	Very Somewhat Not at All	Very Somewhat Not at All	
Scoring Guide	Very Somewhat Not at All	Very Somewhat Not at All	
Action Planning	Very Somewhat Not at All	Very Somewhat Not at All	
Glossary	Very Somewhat Not at All	Very Somewhat Not at All	

Previous Experience

Have you previously completed an assessment of district capacity such as the District Capacity Assessment (DCA; SISEP) or the District MTSS Capacity Assessment (DMCA; MiBLSi) *

- Yes, the District Capacity Assessment (DCA; SISEP)
- Yes, the District MTSS Capacity Assessment (DMCA; MiBLSi)
- No
- Unsure
- Other

*

Consider your experience with the DCA *

What benefits have you and / or your team experienced from your previous experience comp

What would have improved your previous DCA experience?

Is this current version of an assessment of district capacity an improvement compared to what you have worked with before?

Not an Improvement — Significant Improvement

Comments

Consider your experience with the DMCA *

What benefits have you and / or your team experienced from your previous experience comp

What would have improved your previous DMCA experience?

Is this current version of an assessment of district capacity an improvement compared to what you have worked with before?

Not an Improvement — Significant Improvement

Comments

Response Time

Appendix B: Think Aloud Protocol Guide

Think-Aloud Protocol Guide (TAP-Guide): Instructions

The TAP-Guide is a data collection blueprint used by researchers who are in the process of developing an instrument. It contains several best practices advocated by researchers who have developed rigorous verbal reporting methods called think-aloud protocols. The TAP-Guide is intended to be an efficient strategy for gathering evidence of validity based on response process and may provide valuable information around various design problems that introduce construct-irrelevant variance (e.g., unclearly defined instructions, items, and response categories).

How to Complete the TAP-Guide

Step #1 & #2: Complete Demographic Information & Individual/ Team Profile:
Indicate the name of the observer, date, number of participants, and the name of the participating individual/ team. Begin each cognitive interview with a review of individual/ team characteristics regarding professional role, level of professional experience, and any other identifying information pertinent to the study.

Step #3: Review Conditional Probes and Taxonomy of Possible Response Problems
In order to improve validity and objectivity during the think-aloud session, the TAP-Guide provides a standardized format for probing and selecting a taxonomy of possible participant problems. Investigators can use condition-specific probes when participant's verbal reports signal they are having a potential problem and thus, warrants their use. Investigators should be well versed on the major response stages that a participant is likely to pass through and major problem types for which participants provide evidence when answering an item.

Step #4: Begin TAP Protocol Part I: Establish Rapport, TAP Directions, Modeling and Practice of Examples
Investigators should present standardized instructions, model an example question, and provide a practice phase. The practice phase provides the participants an opportunity to practice thinking aloud and to ensure that it meets the expectations of the investigator. Following, the investigator should ask participants if they have any questions and then proceed.

Step #5, #6, & #7: TAP Protocol Part I, II, & III: Collect Data
Investigators should be prepared to record data in the first two sections, which support the collection of introspective data on items by making available a section for qualitative observation data and a checklist of the major response stages and problem types.

Step #7: TAP Protocol Part IV: Collect Retrospective Data on Instrument Directions, Items, Response Categories, Scoring Rubric, and General Observations/ Questions
There is no script for follow-up questions. Instead, investigators should address issues and questions that arise as a result of the think-aloud protocol. However, the TAP-Guide does provide a framework for facilitating the follow-up interview. TAP Protocol Part IV provides qualitative data recording sections on instrument directions, items, response categories, scoring rubric, and general observation or questions.

Think-Aloud Protocol (TAP) Guide: Introspective & Retrospective Data Collection

Step 1

Observer: _____ **Date**: _____
# of Participants: _____ **Participating Individual/ Team**: _____

Step 2

Team Profile: Please identify at least two characteristics regarding participants' role and teams level of experience (e.g., job title, years of experience, trainings attended).

_____
_____
_____

Step 3

Review Conditional Probes: C stands for "conditional;" P for "probe"

C	Participant cannot answer; does not provide a protocol	P	"What was going through your mind as you tried to answer the question?"
C	Answer after a period of silence	P	"You took a little while to answer that question. What were you thinking about?"
C	Answer with uncertainty (e.g., "um" and "ah," changing an answer, etc.)	P	"It sounds like the question may be difficult. If so, can you tell me why?" "What occurred to you that caused you to change your answer?"
C	Answer contingent on certain conditions being met ("…if you don't need a super precise number.")	P	"You seem a little unsure. Was there something unclear about the question?"
C	Erroneous answer; verbal report indicates misconceptions	P	Clarify participant's understanding of particular word, concept, etc. Probe this misconception.
C	Participant requests information initially instead of answering.	P	"If I weren't available or able to answer, what would you decide it means?" "Please elaborate."

Step 4

Review Taxonomy Of Possible Respondent Problems

	RESPONSE STAGE		
PROBLEM TYPE	Understanding (Problems with comprehension of item)	Task Performance (Understands item but difficulty executing - retrieval, deduction, etc.)	Response Formatting (Differences in response and response options)
Lexical (Meaning of words)	Trouble comprehending the meaning of words/ concepts, phrases, etc. (e.g., defining "capacity building")	Understands meaning but trouble differentiating (e.g., "capacity building": does CHAMPS training count?)	Differences in meaning of response and provided category labels (e.g., 8 participants vs. "many")
Omission/ Inclusion (Understanding the scope of a word/ item)	Trouble understanding the scope and limits of a word (e.g., "capacity building": individual and/or organizational")	No explicit decision rule for including/ excluding instances from a category (e.g., "organizational capacity": include "community level"?)	Involves using a response option that was not explicitly provided (e.g., answering 1.5 vs. 0, 1, or 2).
Temporal (Involve time period)	Trouble grasping the meaning of temporal terms (e.g., "last year": calendar year vs. past 12 months	Understands meaning but assigns an incorrect reference (e.g., "current month" mistaken for previous month b/c month just recently changed)	Differences in response time period and provided category labels (e.g., 6 months vs. "often")
Logical (Involves semantics)	• Trouble comprehending the inclusion of semantic devices (e.g., and, or, non-, other than, un-, etc.) • The inclusion of false presuppositions in a question (e.g., how many times a month do you provide teacher consultation?) • Item involves contradictions/ tautologies (e.g., necessary requirement, forward planning, the truth is false, great fidelity but bad implementation)		
Computational	• Residual category: assign after all others have been considered		

Step 4

TAP Protocol Part I: Establish Rapport, TAP Instructions, Modeling and Practice of Examples

Checklist

□ Establish Rapport
□ Provide Explicit Instructions □ Provide Clear Expectations

"We will begin the survey soon, starting from the beginning by reading the directions and then proceeding to the items. After reading each item, think aloud as you reflect and problem solve. After you have read each item, respond to them in your own words. Do not feel pressured to answer each item correctly, as no evaluation of you or your rating performance will occur. Please act and talk as if you are talking to yourself and be completely natural and honest about your rating process and reactions. Also, feel free to take as long as needed to adequately verbalize."
"Do you have any questions?" "Can you say in your own words, what the expectations are?"
"To sum up, we are less interested in the answer participants provide as we are with how they are thinking about them. Remember, we are interested in how participants solve, think about, feel, and the beliefs they have as they respond to survey items."

□ Model TAP
Model Example: "Lets start with an example. I will go first and then you can do the next one":

- "ISD implementation Plan operationally defines steps for addressing equity issues related to educational programming." "Indicate the level of implementation from 0 to 2, 0 indicating *not in place*, 1 indicating *partially in place*, and 2 indicating *fully in place*."
    - When I read this, I'm not exactly sure how "equity issues" are defined. It could mean trying to reduce gaps between subgroups, or it could mean something more specific. It's been a while since I've actually seen our ISD implementation plan, so I'm not really sure if our plan does address equity issues. Because I don't know and I don't remember us ever talking about it, I would score this a 0, but will probably want to check with my other team members at some point.

Step 5

□ Provide Practice
Participant Example: "Now you try: ISD implementation team consists of a diverse group of professionals."

- "Indicate the level of implementation from 0 to 2, 0 indicating *not in place*, 1 indicating *partially in place*, and 2 indicating *fully in place*."
- Provide suggestions to correct and praise to encourage.
- Ask participants whether they have any questions and then proceed.

TAP Protocol Part II: Collect Concurrent Data on Instrument Directions

Qualitative Data
Respondent Problems
☐Understanding Task (Comprehension of item) ☐Performance (Difficulty executing) ☐Response Formatting (Response options) ☐Lexical (Meaning of words) ☐Omission/ Inclusion (Scope of a word/ item) ☐Temporal (Time period) ☐Logical (Semantic: Semantic devices [and, or, other than, etc.], false presuppositions, contradictions, etc.) ☐Computational (Residual Category: Assign after all others have been considered)

Step 6

TAP Protocol Part III: Collect Concurrent Data on Items

Item	Qualitative and Respondent Problem Data
1	Qualitative: *Respondent Problems*: □Understanding Task (Comprehension of item) □Performance (Difficulty executing) □Response Formatting (Response options) □Lexical (Meaning of words) □Omission/ Inclusion (Scope of a word/ item) □Temporal (Time period) □Logical (Semantic: Semantic devices [and, or, other than, etc.], false presuppositions, contradictions, etc.) □Computational (Residual Category: Assign after all others have been considered)
2	Qualitative: *Respondent Problems*: □Understanding Task □Performance □Response Formatting □Lexical □Omission/ Inclusion □Temporal □Logical □Computational

***Continue for all 28 items**

Step 7

TAP Protocol Part IV: Collect Retrospective Data on Instrument Directions, Items, Response Categories, Scoring Rubric, and General Observations/ Questions

Directions
Items

Response Categories
Scoring Rubric
General Observations/ Questions