



Regional Capacity Assessment (RCA)

Technical Manual

Developed in collaboration between NIRN and MIBLSI

March 2019

Suggested Citation: Ward, C., Harms, A., St. Martin, K., Cusumano, D., Levy, R., & Russell, C. (2019). Regional Capacity Assessment Technical Manual. National Implementation Research Network, University of North Carolina at Chapel Hill.

Table of Contents

PREFACE	3
Purpose of This Manual	3
Audience	3
OVERVIEW OF RATIONALE FOR THE REGIONAL CAPACITY ASSESSMENT	3
Role of Implementation Science within Education	3
Need for a Measure of Regional Capacity for Implementation	5
OVERVIEW OF THE REGIONAL CAPACITY ASSESSMENT (RCA)	5
Description of the RCA	5
VALIDATION OF ASSESSMENTS	7
Approaches to Validity.....	7
Focus and Process of Current Validity Work.....	9
Initial Development of the RCA	10
Construct Definitions	10
Items and Rubric	10
Content Validation Process: 4-Part Survey Protocol	11
General Survey Development.....	11
Test Content Participants	11
Content Validation Survey Elements	12
RCA Test Content Validation Results	13
Response Process: Think Aloud Protocols	17
Response Process Overview	17
General Protocol Development.....	17
Response Process Protocol Usage for the RCA.....	18
Response Process Participants	18
Response Process Results and Modifications to the Measure.....	19
Usability Testing Overview	19
Usability Testing Results and Modifications to the Measure	20
APPROPRIATE USE AND FUTURE DIRECTIONS OF THE REGIONAL CAPACITY ASSESSMENT	21
Appropriate Use of the RCA	21
Future Validation of the RCA	21
REFERENCES	22

Preface

Purpose of This Manual

The purpose of the Regional Capacity Assessment (RCA) Technical Manual is to provide background information on the technical adequacy of the RCA (St. Martin, Ward, Fixsen, Harms, & Russell, 2015). This current version draws upon a rich history and background of previous work assessing regional 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 RCA was released in 2016 following a thorough development process and early validation work resulting in a high quality assessment of regional capacity for implementation of effective innovations within local education agencies (LEAs). 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 regional education agencies.

Audience

This manual was written for state, regional, and local agencies that are considering or already using the RCA to assess regional 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, RCA Administrators, facilitators, and respondents may use this manual to deepen their background knowledge on the development and validation of the RCA.

Overview of Rationale for the Regional Capacity Assessment

Role of Implementation Science within Education

Increased attention is being paid to how innovations are implemented because students cannot benefit from evidence-based 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 (see Table 1) 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 (i.e., Effective 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 REAs and LEAs to use the AIFs 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 (Fuchs & Deshler, 2007; Skiba, Middelberg, & McClain, 2013).

Need for a Measure of Regional 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. Aarons, Cafri, Lugo, & Sawitzky, 2012; Klein, Conn, Smith, Speer, & Sorra, 2001). In order for REAs and 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 region'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 are 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.

Overview of the Regional Capacity Assessment (RCA)

Description of the RCA

The Regional Capacity Assessment (RCA) is a 27-item team-based self-assessment developed to assist Regional Education Agencies (REAs) in the implementation of effective innovations that benefit students (St. Martin et al., 2015). A Regional Implementation Team, inclusive of Regional Leadership, uses the RCA 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 region-wide consistent structure for supporting initiatives and practices across LEAs. Additionally, the RCA can be used as part of a feedback structure to a state education body to improve and focus the work of individuals who support districts.

In completing the RCA, the regional team works through items that ask them to consider how they are providing support to districts in their implementation of an Effective Innovation (EI). An EI 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)” (St. Martin et al., 2015, p. 36). Consequently, a team is able to utilize the RCA with any/all innovations that are occurring within the system or use the assessment with their most prominent initiative.

The RCA is grounded in the understanding that regions must develop capacity in the Active Implementation Frameworks (AIFs; Fixsen et al., 2005) to reach desired outcomes from an innovation. According to St. Martin and colleagues (2015) regional capacity “refers to the systems, activities, and resources that are necessary for a REA to be able to facilitate district-level implementation of [EIs]” (p. 4). Key organizational activities required for strong implementation and sustainability of efforts are organized into four scales that include Leadership, Competency, Organization, and Stage-based Functioning.

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 Drivers includes: Leadership and Action 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, Staff 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.*

Stage-based Functioning— Key activities to be engaged in systematically with districts to build their knowledge and skills in using implementation science practices. Activities are organized by the AIF of Implementation Stages: Exploration, Installation, Initial Implementation, and Full Implementation.

The suggested schedule for conducting a RCA is twice a year, about every six months. Throughout the administration of the 27-item self-assessment, a rubric is utilized to anchor current functioning with a score of 0, 1 or 2.

The RCA requires specific roles, including: (a) RCA Administrator, (b) Facilitator, (c) Note Taker, and (d) Respondents. Preparation for the administration of assessment includes commitment to the time for the RCA administration, identification of roles, and securing leadership support for the administration and use of the results for action planning. 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 RCA scoring guide (i.e., rubric). The facilitator contributes to the process by providing necessary context 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 RCA, the team enters

their results into an excel file, which can house repeated administrations of the RCA for the purposes of tracking data over time. It is important to note that the RCA 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 region’s capacity to support the identified EI.

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: (a) content validity, (b) criterion validity, and (c) 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: (a) test content, (b) response process, (c) internal structure, (d) relationship to other variables, and (e) 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	<ul style="list-style-type: none"> • Basis for items/literature review • Qualification of authors and reviews • Item writing process • Review by panel of experts

		<ul style="list-style-type: none"> • 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 <ul style="list-style-type: none"> • 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.

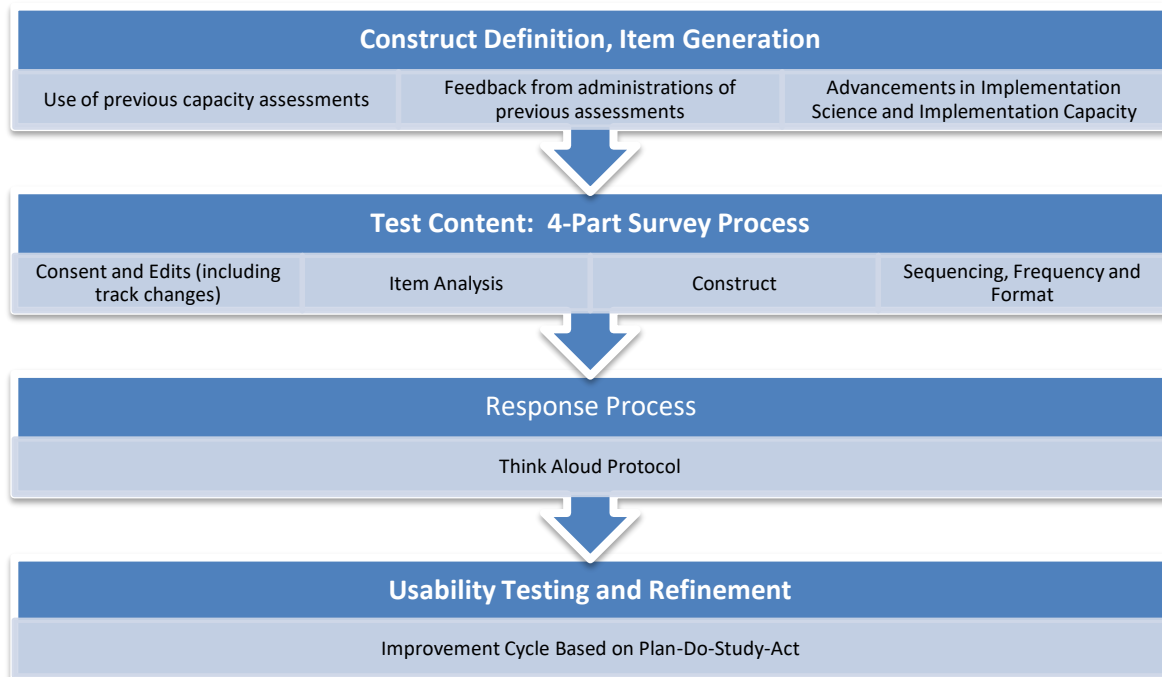
A critical first step in instrument development is gathering evidence of how soundly the test content measures the construct. This should be gathered as a part of the test development process. Gathering evidence of test content establishes the appropriateness of the conceptual framework and how well 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.

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 RCA items and scoring rubric in 2015, a 4-part survey was developed to collect feedback from experts and practitioners regarding the assessment during Spring of 2015. In Spring of 2015 think aloud protocols were completed and usability testing was completed. Following each phase of work with the RCA, the assessment was refined based on the feedback and information gathered (see Figure 1). The results of each stage and the modifications made are discussed in further sections of this technical report.

Figure 1. Phases of Work



Initial Development of the RCA

Construct Definitions

The RCA is designed to measure practices as operationalized within Active Implementation Frameworks, meaning that the regional agency 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 RCA 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 the District Capacity Assessment (DCA) were used to set the stage for initial item generation of the Regional Capacity Assessment (RCA). Factors influencing item generation included: careful consideration of items relevant for regional education agencies; use of feedback collected from administrators and facilitators of previous DCA administrations, and recent advances in the field of implementation science. Items included in others measures were deleted when deemed inappropriate or ineffective or revised as appropriate for the regional level of the education system. In addition, 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, RCA Administration Fidelity Checklist, RCA 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 RCA 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 RCA content validity survey results included feedback from 23 participants. Initially, 35 individuals received the request for survey participation resulting (66% 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

Participant Category	<i>n</i>
Expert Participant Category	3
Research/National Technical Assistance Providers	2
State/Regional Technical Assistance Providers	13
District Practitioners	5
Total (N)	23

Table 3 shares descriptors of the participants ($N = 23$) who engaged in the RCA content validation survey. A majority of participants were state/regional technical assistance providers ($n = 13$).

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 RCA and the survey in which included elements are addressed. Table 5 lists each survey and what components were included within that specific survey.

Table 4. Test Content Survey Components

Survey	Included Survey Components
RCA Consent and Edits	<ul style="list-style-type: none"> • Consent form • Opt in/out of listing as an RCA contributor • Downloadable Microsoft® Word® document of the RCA • Upload RCA with edits, suggestions, questions provided through track changes
RCA Item Analysis	<ul style="list-style-type: none"> • Attainability of each RCA item rated on a 3-point scale • Importance of each RCA item rated on a 3-point scale • Opportunity to select the 5 most critical RCA items • For items including at timeframe (e.g. within two weeks) appropriateness of timeframe on a 3-point scale and space provided for comments related to timeframe
RCA Construct	<ul style="list-style-type: none"> • Comprehensiveness of each RCA construct definition rated on a 3-point scale • Clarity of each RCA construct definition rated on a 3-point scale • Open-ended comments on construct definitions • Appropriateness of current RCA item mapping to constructs rated on a 3-point scale, ability to suggest a better matching construct and open ended comments
Sequencing, Frequency, Format	<ul style="list-style-type: none"> • Suggestions for reordering RCA items • Frequency RCA should be administered

- Comprehensiveness of each RCA section rated on a 3-point scale
 - Clarity of each RCA section rated on a 3-point scale
 - Open-ended comments on sections of the RCA
-

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 minute to 3 minutes). 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.

Within the first survey, participants were asked first to read the current version of the RCA and make track changes within a word document denoting questions, suggestions for rewording, re-ordering, etc. Following the initial read and track changes, the participants completed the remaining sections of the four-part survey.

In the second part of the survey, participants were asked to rate the RCA items on attainability and importance with a three-point Likert scale response with anchors of *Very*, *Somewhat*, and *Not at All* coded as 3, 2, or 1, respectively. Additionally, participants were asked to pick their top five critical items; and, when applicable, participants were asked to share appropriate timeframes for items. The purpose of this step was to help to further discern which items participants viewed as critically important.

Next, participants were asked to share information on the RCA construct definitions. Specifically, comprehensiveness and clarity of constructs were rated on the aforementioned three-point scale. Participants were also asked to provide open-ended comments on construct definitions and the appropriateness of the mapping of constructs to Implementation Drivers.

Finally, during the last part of the four-part survey, participants were asked to provide suggestions regarding sequencing, frequency, and formatting of the RCA. Refer to Table 4 for more specifics regarding the survey components.

RCA Test Content Validation Results

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 RCA in an effort to triangulate data and enhance decision-making. Each participant provided qualitative responses from the open-ended questions within the survey which were combined with comments, edits, questions and suggestions from the track changes documents. Decision rules were developed and agreed upon by the RCA developers prior to analysis of the results. 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. Other qualitative data were reported by number of participants responding a particular way with predetermined cut scores set for analysis.

Table 4. Item Analysis Results

RCA Item	Importance CVI	Attainability CVI	Number of Times Rated as Most Important Item
1	3.00	3.00	19
2	2.70	2.70	3
3	2.90	2.60	10
4	3.00	3.00	8
5	2.80	2.90	5
6	2.80	2.60	6
7	2.70	2.30**	1
8	2.80	2.50	0
9	2.90	2.70	3
10	3.00	2.70	3
11	2.90	2.70	8
12	2.90	2.60	3
13	2.90	2.30**	7
14	2.90	2.70	8
15	2.90	2.60	7
16	2.70	2.50	2
17	2.80	2.40**	2
18	2.50	2.20**	0
19	3.00	2.70	3
20	3.00	2.50	5
21	3.00	2.70	10
22	2.70	2.60	0
23	2.50	2.30**	1
24	2.60	2.30**	0
25	2.80	2.40**	0
26	2.90	2.50	1
27	2.80	2.50	1
28	2.80	2.70	0
29	2.70	2.20**	0
30	2.70	2.40**	1
31	2.80	2.30**	0

Note. * = if Importance CVI is below 2.5, eliminate or substantially change the item;
 ** = if Attainability CVI is below 2.5, develop an action plan to create resources to assist teams with action planning and attaining item.

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 most important aspect. Within the validation process of the RCA, 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 RCA. 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 (i.e., > 2.5), RCA 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, developers 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, and the developers used this information to prioritize resources to assist regions in their efforts to develop capacity.

At the conclusion of the item analysis, the RCA developers made edits resulting in the reduction of items from 31 items total to 27 items. The edits made were based on reviewer feedback that was provided through the track changes provided within the assessment tool. In addition, edits were made to each item and to how the item was defined in the scoring guide. Specifically, these four items were deleted due to low ratings of importance or overlap with other items.

- RIT conducts an external review of District Implementation Plans
- Members of the Regional Education Agency Executive Management Team are designated as executive leaders for the RIT work
- RIT utilizes a process of aligning EIs
- RIT provides and on-site review of overall DIT functioning

Table 5. Construct Definition Results

Construct	Comprehensive CVI	Clarity CVI
Leadership	2.60	2.70
Organizational Environment	2.30*	2.22*
Competency	2.65	2.70
Planning	2.74	2.83
Performance Assessment	2.91	2.78
Selection	2.65	2.65
Training	2.65	2.78
Coaching	2.30*	2.48*
Decision Support Data System	2.78	2.57
Facilitative Administration	2.65	2.26*
Systems Intervention	2.83	2.65
Stage Based Functioning	2.65	2.48*

Note. * = if CVI is below 2.5 for comprehensiveness or clarity, revise definition based on comments and track changes.

The RCA developers used the same CVI cut-point (i.e., < 2.5) for both comprehensiveness and clarity of the RCA construct definitions. In this instance, two constructs needed to be revised for comprehensiveness, and four constructs needed to be revised for clarity (see Table 5).

Table 6. Frequency of Assessment Results

Frequency	Percent of Respondents Suggesting the Frequency
Monthly	0%
Quarterly	8%
Bi-Annually	52%
Annually	30%
Every 2 Years	0%
Other	8%

Note. Decision rule used: if more than 70% of respondents suggest one option for frequency, use the recommendation as the suggested frequency.

The RCA developers asked participants to suggest the frequency of which the RCA should be administered. Table 6 summarizes these results. The developers decided upon the following decision rule prior to collecting the results: if more than 70% of respondents suggest one option for frequency, use the recommendation as the suggested frequency. Per the results, there is no one recommended frequency that achieves the decision rule cut-point. However, over half of the participants recommended that the RCA be administered bi-annually.

Table 7. Item Mapping Results

Item match	Amount of Items
Greater than 70% of Respondents Aligned Item to the same construct	30
50-70% of Respondents Aligned Item to Same Construct	1
Less than 50% of Respondents Aligned Item to Same Construct	0

Note. Decision rule used: if less than 70% of respondents align an item to the same construct, use results, comments and personal knowledge of the constructs to map an item to a construct.

The developers asked participants to indicate if they agreed with the construct the item was originally aligned with in the RCA they reviewed. Table 7 summarizes these results. The participants agreed that almost all of the items ($n = 30$) aligned with the construct originally placed with using the following decision rule: if less than 70% of respondents align an item to the same construct, use results, comments and personal knowledge of the constructs to map an item to a construct. Only one item fell below the decision point cut point, so that item was

mapped onto a different construct.

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 (a) directions, (b) resulting thinking and, (c) acting by those using the assessment. Alignment between participant responses and intended interpretation of the assessment are evaluated by way of this process (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 RCA. 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 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, RCA 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 Protocol Usage for the RCA

The response process participants were provided the following directions, along with modeling and practice, prior to beginning the response process. For this response process, Think Aloud Protocols and consistent documentation were utilized across all participants.

“We are going to ask you to read portions of the document aloud. **The purpose of reading aloud is to ensure clarity and ease of reading the measure.** This process will allow us to capture any areas where wording needs to be adjusted. As you read please verbalize any thoughts, reactions, or questions that are running through your mind. 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.”

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 regional implementation or have various levels of experience in using previous iterations of district capacity assessments. Roles included a regional special education director, school psychologist/MTSS Coordinator, state education agency leader, and a regional school improvement facilitator representing a variety of experiences with implementation work. Within the reviewer group two individuals had worked closely with MIBLSI to implement an effective innovation, one individual had just started work with MIBLSI and the fourth was from another state.

Each participant provided feedback through a one-on-one meeting with one of the RCA developers using the described TAP Guide. In the interest of receiving high-quality feedback without fatiguing the reviewers, one additional respondent reviewed the Introduction and Purpose and RCA Administration Fidelity Checklist; while, two additional participants reviewed the Scoring Guide, which also included a review of the RCA items. Review of the items were split, with some overlap between participants (Items 1-12, 10-27, 1-21, 22-27) 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 one to two 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 RCA 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. In addition, additional words were identified for inclusion in the glossary. 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 RCA.

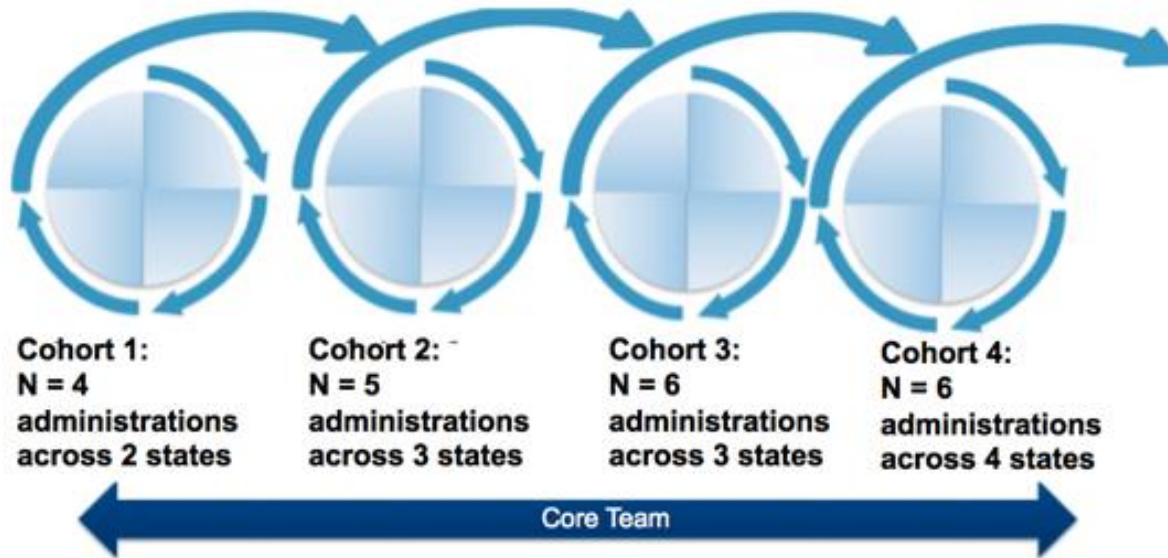
Usability Testing Overview

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 RCA administrations (N= 4-6) 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:

- a. Plans - Leads the improvement planning process and develops the scope of the test for use of the assessment
- b. Does - Engages in using the assessment as outlined in the planning phase
- c. Studies – After each data collection cycle, the team studies what is working (or not) using data
- d. 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 Process and Participants



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 were met by the end of the fourth improvement cycle. Examples of improvement for different areas studied and acted upon are listed below in Table 8.

Table 8. Areas Identified for Improvement Based on Usability Testing

Area	Example Improvement
Communication and Preparation	More guidance developed around team composition and respondents
Administration Protocol	100% on fidelity protocol and rating of importance (4 or higher)
Items and Scoring Rubric	Minor wording changes to items; sequencing of items was reviewed but not changed
Training Implications	Facilitation skill identified; prioritization of areas for action planning
Participant Response	Engaged and positive

Note. Number of improvements in each of the five areas decreased over the cycles and all goals were met

Appropriate Use and Future Directions of the Regional Capacity Assessment

Appropriate Use of the RCA

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

- Region self-assessment and progress monitoring used to guide and improve implementation capacity building
- Coaching for region, 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 RCA should **not** be used as high stakes evaluation tool of a Regional Implementation Team. The RCA's validity and reliability is still being assessed. Its principal purpose is for use as an action assessment to assist regions and their districts to implement evidence-based practices that benefit students.

Future Validation of the RCA

Next steps in the development and validation process of the RCA include designing and conducting research to further examine the RCA'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 RCA. The RCA 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.
- 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., & Blair, J. (1996, August). From impressions to data: Increasing the objectivity of cognitive interviews. In *Proceedings of the section on survey research methods, annual meetings of the American Statistical Association* (Vol. 1, No. 10).
- Conrad, F., Blair, J., & Tracy, E. (1999, November). Verbal reports are data! A theoretical approach to cognitive interviews. In *Proceedings of the Federal Committee on Statistical Methodology Research Conference* (pp. 11-20). Tuesday B Sessions.
- 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).
- DeVellis, R. F. (2012). *Scale development: Theory and applications*. Thousand Oaks, CA: SAGE Publications.
- 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. Technical Report 44. *National Center on Educational Outcomes, University of Minnesota*.
- 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, 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.
- 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.
- 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.

St Martin, K., Ward, C., Fixsen, D., Harms, A., & Russell, C. (2015). *Regional Capacity Assessment*. National Implementation Research Network, University of North Carolina at Chapel Hill.

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

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

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