2018

The Readiness Monitoring Tool: Investigating The Psychometric Evidence For Group-Level Aggregation In Two Samples

Samantha N. Hartley
University of South Carolina

THE READINESS MONITORING TOOL: INVESTIGATING THE PSYCHOMETRIC EVIDENCE FOR GROUP-LEVEL AGGREGATION IN TWO SAMPLES

by

Samantha N. Hartley

Bachelor of Arts
Stanford University, 2011

Submitted in Partial Fulfillment of the Requirements

For the Degree of Master of Arts in

Clinical-Community Psychology

College of Arts and Sciences

University of South Carolina

2018

Accepted by:

Abraham Wandersman, Director of Thesis

Mark Weist, Reader

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

DEDICATION

To my friends, family, and mentors, who expressed confidence in my readiness when I was uncertain: thank you. Any innovation would be lucky to have you as a Program Champion, but I am particularly grateful to have you in my life.

ACKNOWLEDGEMENTS

Many thanks to Jonathan Scaccia, Andrea Lamont, and other members of the Readiness Team for offering connections to existing readiness projects and providing an initial orientation to a brave new world of psychometrics and measure development. Without your help, this thesis would not have been possible.

ABSTRACT

Organizational readiness is an essential factor for successful implementation of a particular innovation. Although there is general consensus within the research literature on the importance of organizational readiness, there has been significantly less agreement on the nature of readiness as a construct and how it should be operationalized. Previous research has focused on organizational readiness for change at both the individual and organizational levels, but measures based on these theories of readiness have typically lacked evidence of reliability and validity.

The $R=MC^2$ heuristic and its associated measure, the Readiness Monitoring Tool (RMT), offer a compelling and comprehensive approach by which organizations can identify specific areas of readiness which could benefit from capacity-building efforts to strengthen implementation supports. Initial psychometric analysis of the RMT has included establishing internal reliability, content validity, criterion validity, and preliminary factor structure, but additional testing is required to establish the RMT as a theoretically-informed and psychometrically-sound measure. This study presents the results of a psychometric assessment of data from two project samples, including 1) a comparison of internal consistency measures to determine whether this property is maintained across project-specific adaptations and 2) calculation of interrater agreement and interrater reliability statistics to provide evidence for group-level aggregation of individual-level RMT data.

Individual-level data was collected from seven projects that had previously adapted the RMT; two projects were selected for analysis based on strong per-site participation. Internal consistency was assessed by calculating alpha coefficients for each RMT subscale. Interrater agreement and interrater reliability was assessed using r_{WG}, AD_M, *ICC(1),* and *ICC(2)* estimates. Internal reliability analysis showed very good internal consistency for the majority of subscales. Interrater reliability and interrater agreement statistics supported group-level aggregation of individual-level responses.

This study provides evidence in support of RMT as an adaptable measure capable of reliably and validly representing an organization's readiness by surveying its members. Occasional variability in results between subscales, sites, and projects informs recommendations for future study and implications for practical RMT use. Although further development is required, the RMT shows promise as an adaptable measure of organizational readiness capable of informing targeted capacity-building support.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Organizational Readiness Overview

In a world with a seemingly limitless need for efficacious innovations, but limited resources to provide for their implementation, it's important that innovations are given the greatest chance possible of realizing good outcomes when put into practice. Meta-analyses have shown that good outcomes are more likely when innovations are implemented with quality and that quality implementation is facilitated by a host of factors captured by the term "organizational readiness" (Durlak & DuPre, 2008). Accordingly, there is general consensus within the research literature that organizational readiness, conceptualized as the extent to which an organization is both willing and able to implement a given innovation, is an essential component of successful implementation (Drzensky, Egold, & Van Dick, 2012; Greenhalgh, Robert, MacFarlane, Bate, & Kyriakidou, 2004; Hall & Hord, 2015; Holt & Vardaman, 2013; Simpson, 2002; Weiner, 2009).

Researchers concede that although it is possible for organizational change to occur under conditions of low readiness, the probability of success is reduced (Backer, 1995). Up to half of failed organizational change efforts may be attributable to inadequate organizational readiness for change (Kotter, 1996, as cited in Weiner, 2009): when readiness is overlooked and organizational readiness is insufficient, an otherwise appropriate intervention may not produce desired outcomes (Pasmore & Fagans, 1992).

To produce desired outcomes, readiness must be 1) assessed before attempting to adopt an innovation, 2) improved where otherwise lacking, and 3) monitored and addressed as needed throughout the process of implementation (Armenakis, Harris, & Mossholder, 1993).

For this reason, readiness has been embedded in several program planning and implementation frameworks aimed at bridging the gap between evidence-based innovations and implementation in real-world settings (Aarons, Hurlburt, & Horwitz, 2011; Powell et al., 2011; Wandersman et al., 2008). However, though agreement exists on the essential role of organizational readiness in successful implementation, there has been significantly less unanimity on the nature of readiness as a construct and how it should be operationalized and measured for a given innovation.

Previous Conceptualization and Measurement of Readiness for Change

Readiness for organizational change has been examined extensively at both the individual and organizational levels. This introduction offers a brief and nonexhaustive review of those efforts.

Individual readiness for organizational change.

Researchers who focus on the role of individual readiness for organizational change reason that "organizations only change and act through their members" (Choi & Ruona, 2011). Individual readiness for organizational change is a conceptualized as a state of mind held by individuals (Backer, 1995) which is influenced by their beliefs, attitudes, and intentions regarding 1) the need for an innovation, 2) the individuals' and organization's capacity to successfully make the change, and 3) the anticipated benefits for themselves and the organization that are likely to result from doing so. Individual

readiness can be shaped by the readiness of others, as well as by attitudes and beliefs about the organizational context in which the innovation is to be implemented (Jansen, 2000). However, creating organizational readiness for change ultimately involves altering across a set of employees individual cognitions about the innovation (Armenakis et al., 1993).

Two models of assessing individual readiness for organizational change are A VICTORY (Davis, 1978) and the Concerns-Based Adoption Model (Hall & Hord, 2015). A VICTORY is an eight-factor model for assessing readiness for change comprised of a 75-item rating scale that can be completed via interview or self-administered survey. The survey results can be used to develop a profile of individuals' attitudes and beliefs about the innovation for that touches on factors regarding ability, values, information about the innovation, circumstances and timing, felt need to change, resistance, and anticipated rewards from implementation (Davis, 1978).

Hall and Hord's (2015) Concerns-Based Adoption Model (CBAM) holds that, regardless of the ultimate scale of a change, organizational change ultimately begins with and is accomplished at the level of the individual. CBAM uses three diagnostic components to assess staff members' attitudes and beliefs about an innovation (*Stages of Concern*), actual use of an innovation (*Levels of Use*), and quality of implementation (*Innovation Configuration Map*). In the pursuit of successfully implementing a particular innovation, these elements emphasize the importance of understanding and addressing the "personal side of change." Although individuals serve as the initial point of measurement, profiles based on these individual assessments can be aggregated to represent the distribution of staff concerns and degree/quality of innovation use within an

organization. This information can then be used to provide support that targets specific concerns impacting high-quality implementation of the innovation.

Organizational readiness for change.

Researchers focusing on organizational readiness at the supra-individual level have conceptualized readiness as a multi-faceted construct referring to an organization's collective ability and commitment to change (Weiner, Amick, & Lee, 2008; Weiner, 2009). Definitions and measurement strategies are divided between 1) those which focus on evaluating key elements of an organization's infrastructure (e.g., culture, leadership, communication, structures, systems) important for facilitating and sustaining change efforts and 2) those which assess organizational readiness as attitudes and beliefs shared by members of an organization that either hinder or promote implementation of an innovation.

Approaches to organizational readiness that focus on context assess the extent to which structural or procedural elements necessary for successful implementation are in place within an organization planning to adopt an innovation. Three examples of this form of readiness assessment include the School Context Analysis Form (D'Amico & Corbett, 1988), the Implementation Leadership Scale (Aarons, Ehrhart, & Farahnak, 2014), and the General Organizational Index (Lynne, Finnerty, & Boyle, 2005).

The School Context Analysis Form is a 19-item informal rating system, which rates schools as high, moderate, or low on eight contextual factors (e.g., resources, incentives and disincentives, linkages, priorities, factions, turnover, current practices, prior projects) known to influence innovation implementation and sustainability

(D'Amico & Corbett, 1988). The results of the rating system are intended to inform support efforts to create a school context more conducive to successful change.

The Implementation Leadership Scale recognizes that effective leadership is a critical contextual element in the successful implementation of a given innovation. The measure is a 12-item scale with four subscales (i.e., proactive leadership, knowledgeable leadership, supportive leadership, perseverant leadership) designed to measure organizational leadership for implementation of evidence-based practice (Aarons et al., 2014). The scale has been assessed for internal consistency reliability, convergent validity, and discriminant validity, and is designed for use as a predictor of implementation and as a guide for leadership development within an organization.

The General Organizational Index (GOI) was developed as a process assessment which program evaluators could use to measure organizational-level characteristics that have been found to affect an agency's overall capacity to implement and sustain evidence-based interventions (Lynne et al., 2005). Using the GOI rubric, organizations are scored in 12 categories, including program philosophy, training, outcome monitoring, quality assurance, and key elements of evidence-based interventions. Although psychometric properties have not yet been assessed, the scale has undergone multiple revisions based on feedback from pilot testing.

Measurements of organizational readiness that focus on measuring shared attitudes, perceptions, and beliefs that impact the climate in which an innovation is implemented are more numerous within the implementation science literature. Examples reviewed here include the Implementation Climate Scale (Ehrhart, Aarons, & Farahnak, 2014), the Dimensions of Organizational Readiness – Revised (DOOR-R) (Hoagwood,

Schoenwald, & Chapman, 2003), the Organizational Climate Measure (OCM) (Patterson et al., 2005), the Organizational Readiness for Change (ORC) (Lehmen, Greener, & Simpson, 2002), and the Organizational Readiness for Implementing Change (ORIC) (Shea, Jacobs, Esserman, Bruce, & Weiner, 2014).

The Implementation Climate Scale (Ehrhart et al., 2014) was developed to capture a broad range of issues important for effective implementation of evidence-based practice. The resulting 18-item scale evaluates six dimensions of organizational climate that influence employees' perceptions of the extent to which the organization values and prioritizes implementation of the innovation. Psychometric analyses support the factor structure, reliability, and construct validity of the measure, as well as aggregation of individual responses to the organizational level. Results from this assessment are intended to shape support strategies aimed at accelerating effective implementation.

The Dimensions of Organizational Readiness – Revised (DOOR-R) instrument was developed as a planning tool that could be used to assess stakeholder beliefs and attitudes about organizational processes in anticipation of taking on a new innovation (Hoagwood et al., 2003). The instrument covers six broad domains, including characteristics regarding the invention, practitioner, client, service delivery, service agency, and service system.

The Organizational Climate Measure (OCM) is a psychometrically validated multidimensional measure that assesses aspects of organizational climate that impact how effectively organizations adapt and respond to innovation (Patterson et al., 2005). The measure contains 17 distinct scales, including constructs such as acceptance of new ideas,

support in developing new ideas, and flexibility in responding to changes needed in procedures.

The Organizational Readiness for Change (ORC) is a comprehensive assessment of organizational functioning and readiness for change that evaluates motivation for change, personality attributes of program leaders and staff, institutional resources, and organizational climate (Lehmen et al., 2002). Originally developed to study innovation and change efforts in substance abuse treatment agencies, the instrument includes 115 Likert-type items representing 18 content domains measuring the four major content areas previously mentioned. Two versions of the measure were developed, one for use with leadership and one for staff, to accommodate the different perspectives of these positions within an agency. Initial testing has shown ORC scales to possess adequate psychometric properties, including construct validity and reliability.

Finally, the Organizational Readiness for Implementing Change (ORIC) was developed as a brief, valid, reliable, theory-based measure to directly measure an organization's readiness for change (Shea et al., 2014). Weiner's (2009) influential theory of organizational readiness for change conceptualizes organizational readiness as a psychological trait shared by members of an organization. Organizational readiness is determined by two elements: change commitment (i.e., organizational members' shared resolve to implement a change) and change efficacy (i.e., organizational members' shared belief in their collective capability to do so).

Change commitment is influenced by how much individuals value the change, whereas change efficacy depends on their perceptions of 1) task knowledge (their understanding of what to do and how to do it), 2) resource availability (whether they have

the resources needed to implement the change), and 3) situational factors (whether factors like timing are favorable to achieving the change). Organizations with high organizational readiness, whose members are both committed to making a change and confident in their ability to do so, are theoretically more likely to demonstrate effective implementation. Weiner (2009) states that this particular conceptualization of change is well-suited for innovations that require collective behavior change to ensure effective implementation and to produce desired results.

A 12-item scale based on Weiner's theory of organizational readiness for change (Weiner, 2009), the ORIC differentiates two facets of readiness: change commitment and change efficacy. In application, multiple respondents from the same organization are encouraged to complete the measure, answering questions such as "People who work here are committed to implementing this change" (item on change commitment scale) and "People who work here feel confident that they can keep the momentum going in implementing this change" (item on change efficacy scale) on a 5-point Likert scale ranging from "disagree" to "agree."

Psychometric studies have produced evidence in support of content adequacy, factor structure and reliability, and interrater reliability and interrater agreement (Shea et al., 2014). Although additional work is necessary to test for convergent, discriminant, and predictive validity, whereas most publicly available measures of organizational readiness have lacked evidence of reliability and validity (Weiner et al., 2008), the ORIC is currently considered the gold standard measure for organizational readiness.

The Current Model of Readiness: $R=MC^2$

In 2015, Scaccia and colleagues offered a "practical implementation science heuristic for organizational readiness": Readiness = Motivation x Innovation-Specific Capacity x General Capacity (abbreviated by the heuristic $R=MC^2$). The theory behind $R=MC^2$ drew from multiple sources, including Weiner's theory of organizational readiness for change, described above (Weiner, 2009). However, influenced by their prior work in empowerment evaluation (Fetterman & Wandersman, 2005), the authors strove to create a readiness heuristic that could help organizations identify specific areas which could benefit from capacity building efforts to strengthen implementation supports.

In this model, organizational readiness is conceptualized as involving both the willingness (motivation) and the ability (capacity) to implement a particular innovation. Readiness components are considered to be dimensional, interrelated, and dynamic (Scaccia et al., 2015). Organizational readiness is considered a dimensional, rather than categorical, property: organizations do not fall into dichotomous categories of "ready" or "not ready," but rather demonstrate differences in readiness as "a matter of degree." Organizations can demonstrate variability in their relative level of readiness across the three major components, reporting high readiness in some components and low readiness in others. The multiplicative relationship between the readiness components of the $R=MC^2$ heuristic suggests that they interact. One implication of this relationship is that, if any of the components is absent or exceptionally low, the organization cannot be considered ready to implement an innovation. Attempts to implement the innovation will likely fail until all areas of inadequate readiness are addressed. Additionally, the components of readiness are believed to represent dynamic organizational properties: an

organization's level of readiness may fluctuate over the course of implementation. Regular monitoring and discussion of readiness components therefore becomes an essential part of providing adequate implementation support in order to respond to new challenges as they arise.

The model proposes that organizational readiness includes three distinct components: general capacity, innovation-specific capacity, and motivation. The two types of capacity were identified from Flaspohler et al.'s (2008) concepts of 1) general capacity, which includes attributes of a well-functioning organization associated with the ability to successfully implement *any* innovation, and 2) innovation-specific capacity, which constitutes the knowledge, skills, and conditions important for successfully implementing a *particular* innovation. These two types of capacity were previously identified as being necessary factors for the successful widespread implementation of evidence-based interventions within the Interactive Systems Framework (Wandersman et al., 2008). Motivation was added as a third essential component of organizational readiness in recognition of the observation that capacity is necessary, but not sufficient, for quality implementation (Weiner et al., 2008; Weiner, 2009). Influenced by Rogers's (2003) work, motivation captures the strength of the organization's intent to change, which is defined as the perceived incentives and disincentives that contribute to the desirability to use an innovation (Scaccia et al., 2015).

To create a more nuanced and comprehensive conceptualization of organizational readiness, each of the three main components are further divided into subcomponents. These subcomponents were identified through an extensive literature review of past research (Scaccia et al., 2015). Although not exhaustive (the enumeration of previously

identified factors affecting implementation leads one to believe that a completely comprehensive subcomponent list would have strained the limits of practical use; Durlak & DuPre, 2008), these subcomponents are designed to orient stakeholders to major, specific key variables within general capacity, innovation-specific capacity, and motivation. Subcomponents and their definitions, as described by Scaccia et al. (2015), are detailed below.

Seven subcomponents were identified within general capacity: culture, climate, organizational innovativeness, resource utilization, leadership, structure, and staff capacity. Culture entails an organization's operational norms and values. Climate refers to how employees collectively feel about their current work environment. Organizational innovativeness reflects the organization's general receptiveness toward change. Resource utilization describes how well the organization is able to acquire and use potential resources. Leadership describes the effectiveness of organizational leaders in communicating and supporting organizational efforts. Structure refers to processes that affect the quality of day-to-day organizational functioning. Staff capacity assesses the experience and skill of staff, as well as their ability to meet organizational demands.

Innovation-specific capacities include four subcomponents: innovation-specific knowledge, skills, and abilities; program champion; specific implementation climate supports; and interorganizational relationships. Innovation-specific knowledge, skills, and abilities refers to the degree to which the organization collectively possesses the knowledge and skills to implement the innovation. Program champion assesses whether the organization contains influential individual(s) who support the innovation. Specific implementation climate supports, or supportive climate, captures the extent to which the

organization is providing demonstrable, essential support to enable innovation implementation. Interorganizational relationships reflects whether organizations have established relationships necessary for facilitating successful implementation.

Six factors were identified as influencing an organization's motivation to implement an innovation: relative advantage, compatibility, complexity, trialability, observability, and priority. Relative advantage represents the degree to which a particular innovation is perceived as being better than current practice or an alternative innovation. Compatibility is the extent to which an innovation is perceived to be consistent with the organization's existing values, norms, experiences, and needs. Complexity demonstrates how difficult the innovation appears to be to understand and use. Trialability refers to how well the innovation can be tested and adapted in small steps. Observability reflects the ease with which "small wins" or positive outcomes that result from the innovation can be witnessed in the short term. Finally, priority represents the extent to which implementing the innovation is considered more important than other organizational obligations.

The Current Measure: Readiness Monitoring Tool (RMT)

Initial Development

The $R = MC^2$ heuristic was originally operationalized into a measure for use in the Institute for Healthcare Improvement's Spreading Community Accelerators through Learning and Evaluation initiative (SCALE; Robert Wood Johnson Foundation, 2015), although the tool was created with the intent of being customizable for use with other innovations. Development of the Readiness Monitoring Tool (RMT) was informed by DeVellis' scale development process (2016). The authors first inductively developed

items for each of the readiness subcomponents after reviewing existing measures (Aarons et al., 2014; Lehman et al., 2014; Patterson et al., 2005; Shea et al., 2014; Weiner, Belden, Bergmire, & Johnston, 2011). A team of content experts then reviewed the generated items and offered feedback to improve item comprehension and content adequacy, or "the degree to which a measure's items are a proper sample of the theoretical domain of a construct" (Schriesheim, Powers, Scandura, Gardiner, & Lankau, 1993).

After incorporating suggested revisions, the RMT was piloted with two SCALE community representatives to ensure item clarity and measure usability. Their comments led to further refinement of the measure by removing items perceived to be redundant and clarifying item language that was confusing or unclear to community-based respondents. The resulting 79-item scale was administered to a community sample within the SCALE project using a 7-point Likert scale (*Strongly Disagree to Strongly Agree)* to encourage wide range of responses. Consistent with their original intent to adapt the RMT for a variety of different settings and innovations, the SCALE measure was soon modified to form the Readiness for Integrated Care Questionnaire (RIC-Q), a tool for assessing readiness to integrate behavioral health and primary care (Scott et al., 2017).

Initial Use of the RMT

As of this writing, the RMT has been used in a total of 17 projects in a variety of settings -- including by healthy community coalitions, federally qualified health centers, and schools – with different project-specific modifications. Similar to the ORIC (Weiner et al., 2008), administration of the RMT emphasizes the necessity of multiple respondents from the same organization for an accurate assessment of the organization's collective

readiness. The tool has been used for a variety of purposes: as a pre-implementation readiness assessment, with the results shared with respondents to plan for implementation and training/technical assistance needs; as a monitor of changes in readiness during implementation as a result of unplanned events; as a predictor of innovation-related outcomes; and as an outcome measure in interventions focused on organizational capacity-building and coaching (Hartley, 2016). Depending on the degree of adaptation, the RMT has contained from 63 to 100 items focused on various subcomponents of general capacity, innovation-specific capacity, and motivation.

Psychometric Analysis of the RMT

Initial psychometric analysis of the RMT has focused on establishing internal reliability and criterion validity and engaging in preliminary factor analysis (Wandersman & Scaccia, 2017). Content validity was not analyzed quantitatively, but was addressed through consultation with content experts and community stakeholders during the initial item development process.

Internal reliability.

Based on data collected from the SCALE initiative, internal reliability of the subcomponents within each of the three constructs of readiness were assessed by calculating Cronbach's alpha. Good internal reliability was found ($\alpha \geq 0.70$ for 89% of the subcomponent index scores). The remaining two subcomponents, staff capacity and ability to pilot, had acceptable ($\alpha = 0.6294$) and very poor ($\alpha = 0.1215$) internal reliability, respectively.

Criterion validity.

Initial evidence of criterion validity was found by comparing readiness scores of SCALE communities to Pathway to Pacesetter (P2P) communities -- a subset of communities in the IHI initiative not accepted to the SCALE initiative because of an initial assessment of low readiness -- who were invited to participate in a lower dose version of SCALE. P2P communities were expected to have lower readiness than SCALE communities, resulting from either their initial lower levels of readiness or the comparatively smaller amount of support they were receiving from the initiative. A comparison of RMT scores at two timepoints quantitatively demonstrated that SCALE communities had statistically higher readiness scores in nearly all subcomponents compared to P2P communities. This provides initial evidence that the RMT is able to distinguish between lower and higher readiness organizations.

Preliminary factor analyses of the RMT.

To determine whether items in the RMT statistically correspond to conceptual readiness subcomponents, several exploratory models have been run to understand the underlying factor structure of the measure using both the data from the RMT used in the SCALE project and responses to the Readiness for Integrated Care Questionnaire (RIC-Q). Methods used included: exploratory factor analysis, to determine the number of distinct factors measured by the RMT and which items were related to each of those factors; hierarchical cluster analysis, to visualize how items clustered at various levels of the overall scale; multidimensional modeling, to detect underlying dimensions that underlie similarities between individual items in a data set; and Mokken scaling, to

determine the scalability and reliability of the RMT's subscales (Wandersman & Scaccia, 2017).

Although the data used for these analyses drew from a relatively small set of respondents, results indicated 1) RMT items grouped together as expected for the three major readiness components (general capacity, innovation-specific capacity, motivation), 2) there was minor overlap for a few of the subcomponents within the three major components, and 3) there was substantial overlap for some subcomponents, especially for subcomponents within general capacity. These results suggested a need for additional item revisions to create a measure that could better discriminate between subcomponents that, while theoretically expected to be related, represent distinct constructs.

Revision of the RMT

Results from the factor analysis work and qualitative feedback from interviews with early adopters of the RMT (Hartley, 2016) informed a revision process that included developing new items and revising, dropping, and reassigning existing items. These revisions were intended to drop statistically redundant items, eliminate items that were too project-specific to improve adaptability, and revise subcomponent indices to better reflect factor loadings from the exploratory analyses. The resulting revised version of the RMT demonstrated relatively strong factor loadings and better fit indices compared to the original RMT.

Plans for Continued Development

RMT developers plan on continuing their work to develop a theoretically-informed and psychometrically-sound measure of organizational readiness suitable for use in a variety of settings and innovations. Developmental aims center around continued

1) item development, 2) psychometric testing, and 3) evaluation to optimize scale length. Psychometric testing of the validity and reliability of the revised version of the RMT will include confirmatory factor analysis and investigation of internal consistency, interrater agreement, temporal stability, and sources of within-organization variation in RMT response patterns.

The Current Study

Although preliminary analyses have been completed to establish psychometric properties of the RMT and additional work is planned, many areas remain where typical use of the tool is justified by convention, reasonable assumptions, or historical precedent rather than conclusions rooted in statistical evidence.

Similar to the ORIC, the RMT conceptualizes organizational readiness at a supra-individual organizational level, even though data for that measurement is derived from individual-level sources. The RMT is interpreted using subcomponent scores. Subcomponent scores are calculated by averaging the individual item scores across participants from an organization (creating the organization's average score for that item) and then further averaging the combined averaged item scores across all items within a subscale. These scores are compared relative to each other to identify strengths and areas for improvement within the three major constructs of readiness so that technical assistance and implementation supports can be provided in a targeted manner to build readiness in areas of relative weakness (Scaccia et al., 2015).

Though it may intuitively make sense to compute organizational-level means about organizational readiness from the collection of individual RMT responses provided by staff at a particular site – and, indeed, this has been standard practice from the

beginning of RMT use – it is important to check the reliability and validity of that site's mean to determine whether it's an adequate representation of the organization's readiness (Klein, Dansereau, & Hall, 1994; Klein & Kozlowski, 2000).

The RMT represents a composition approach to bottom-up processing in its desired combination of individual-level data to reflect an organizational-level variable (Kozlowski & Klein, 2000). Composition models rest on the assumption that individual-level data are effectively equivalent to the higher-level construct, so it's necessary to show that individual-level data are in agreement with one another (e.g., staff at School A have very similar perceptions of their school's readiness, which differ from how ready staff at School B feel their school is for the same project).

Determining whether individuals at a site share a common perception of their site's readiness requires examining the degree of interrater reliability and interrater agreement amongst their RMT scores (Bliese, 2000; LeBreton & Senter, 2008). Site-level means that reliably and validly reflect the perceptions of the organization's members as a group can be calculated when there is sufficient interrater agreement and interrater reliability. However, if individuals at a site disagree about their organization's readiness to such an extent that sufficient interrater agreement and interrater reliability do not exist, then it would be inappropriate to create site-level means based on the average scores of individual RMT responses. In this case, readiness as a shared team property does not exist, and capacity-building strategies may need to be targeted to individuals based on their response pattern (akin to the individual-level approach to promoting innovation implementation used in the CBAM; Hall & Hord, 2015).

This study looks at the performance of adapted versions of the RMT in two of the seventeen projects in which the RMT has been adapted for use. Research questions are aimed at 1) assessing basic psychometric properties related to scale construction in a measure that encourages project-specific adaptation (e.g., inter-item consistency), and 2) gaining insight into whether averaging individual-level RMT responses into an overall mean score validly and reliably represents an organization's level of readiness.

Research questions within the current study include:

1. Do project-specific adaptations of the RMT maintain sufficient internal consistency within subscales?

   The RMT was designed to be adaptable to different projects, settings, and innovations. However, this has resulted in widely differing versions of the scale, with item counts ranging between 63 and 100 items (Wandersman & Scaccia, 2017). Inter-item consistency on included subscales has not consistently been calculated within these adaptations. Alpha coefficients will be calculated for each subscale and compared both across projects and to established benchmarks for sufficient inter-item consistency (DeVellis, 2003; Nunnally, 1978).

2. To what extent is a participant's ratings on the RMT affected by their site?

   Ratings of organizational readiness are expected to be driven by perceptions of the organization that are specific to that individual's site. Calculating *ICC(1)* values, a form of intraclass correlation coefficient that compares variance between-groups versus within-groups, helps us determine statistically whether this is so.

3. Do mean subscale scores calculated for each site reliably distinguish the sites in each project? Is there sufficient interrater reliability and interrater agreement to justify aggregating the data into site-level means?

    Sites within a project are expected to be distinguishable based on their group-level mean subscale scores. Calculating *ICC(2)* values, a form of intraclass correlation coefficient that accounts for the number of respondents, provides evidence for the reliability of mean ratings in distinguishing different sites within a sample and is used in justifying creating group-level means (Weiner et al., 2011). At the recommendation of LeBreton & Senter (2008), who suggest examining multiple indicators of interrater agreement, two additional indices of interrater agreement, r_{WG} and AD_M, will also be calculated based on mean subscale scores.

Addressing these questions with RMT data from two project samples will provide insight into how consistently the RMT performs across settings and determine whether there is evidence to support the aggregation of individual responses into organizational means. Results from this study will inform practical recommendations for future RMT administration and analysis.

CHAPTER 2

METHOD

This study performed secondary data analyses to assess the reliability and validity of using organizational-level means with data from project-specific adaptations of the Readiness Monitoring Tool.

Participants

With the help of the developers of the Readiness Monitoring Tool (Scaccia and Wandersman), seven projects were identified as having used the RMT in their work to assess contextual issues critical to implementation success. Consistent with the belief that the RMT is highly adaptable to both context and innovation, these projects represented work in a variety of settings interested in implementing a diverse set of innovations, including a nationwide initiative of community coalitions undertaking health improvement projects, middle schools adopting a school safety initiative, federally qualified health centers (FQHCs) transitioning to integrated health/mental health care, and community-based fatherhood programs assessing their readiness to use data for continuous quality improvement.

Deidentified data sets of RMT results with individual-level responses were requested from the project lead in each of the seven identified projects. Data sets were cleaned and organized to determine the number of respondents and sites represented within each (see Table 2.1). Following the example set by Shea et al. (2014) in their assessment of the reliability and validity of an organization-level measure of readiness

based on aggregated individual-level data from a field sample of international non-governmental organizational staff, only sites represented by five or more survey respondents were included in the analysis. From the original pool of seven projects, 97 sites, and 352 total respondents, two projects were selected for analysis based on having the largest number of sites that met inclusion criteria: Community Health Coalition – Project 2 and Fatherhood Program (hereafter referred to as Project 1: CHC and Project 2: FP, respectively).

Project 1: CHC represents an ambitious nationwide initiative to build the capability of community coalitions to successfully promote health, well-being, and equity in their communities through the use of spread and quality improvement techniques in community health improvement projects. After respondents from sites with fewer than five respondents were removed, the sample for Project 1: CHC included 10 sites with a total of 69 respondents (6.9 respondents per site on average). This represents 56% of the sites included in the original project sample.

Project 2: FP represents a statewide initiative to assess organizational readiness to use data for continuous quality improvement at local sites of a community-based program for fathers. After respondents from sites with fewer than five respondents were removed, the sample for Project 2: FP included 8 sites with a total of 62 respondents (7.75 respondents per site on average). This represents 66.7% of the sites included in the original project sample.

Procedure

In Project 1: CHC and Project 2: FP, the adapted version of the RMT was delivered in the form of an online survey. Respondents were provided with instructions

explaining the purpose of the survey, the expected duration of time to completion, and clarification on the nature of the innovation for which they were being asked to assess readiness. Additionally, they were asked to respond to items based on the current status of their site/organization (instead of past or anticipated future functioning) and based on what they believed best represented their site/organization as a whole, not just their own individual views. Although the manner in which individuals completed the survey was not confirmed, project leaders intended respondents to complete the RMT individually without conferring with their colleagues or coming to consensus.

Measures

Organizational readiness as conceptualized by Scaccia et al. (2014) and as measured by the RMT is composed of three primary components (general capacity, innovation-specific capacity, motivation) and multiple subcomponents. Although the RMT was initially developed for use in a project similar to Project 1: CHC, it was intended to be easily adapted for a variety of different innovations and settings. Interviews conducted with early users of the RMT revealed that project-specific adaptation could vary considerably in the degree of substantive modifications made to the original measure. While the majority of projects administered all subscales and made only minor modifications to items within the RMT (e.g., including project- and setting-specific terminology to items to ensure fit with the project's innovation but otherwise conserving item content and count), some made more significant modifications (e.g., replacing jargon with plain language to improve accessibility to community respondents, removing items and subscales depending on the assessment point, and making greater adaptation to Innovation-Specific Capacity scale items). As a result of these intended

adaptations, the RMT measure administered in Project 1: CHC and the RMT measure administered in Project 2: FP are not interchangeable. Included subscales and item counts administered within each project are described below.

The RMT in Project 1: CHC.

The Readiness Monitoring Tool administered in Project 1: CHC represents a 63-item version adaptation of the original RMT. Items are grouped within three major scales (General Capacity, Innovation-Specific Capacity, and Motivation) and 17 subscales. The General Capacity scale consists of seven subscales, including Climate (4 items), Culture (6 items), Leadership (7 items), Organizational Innovativeness (6 items), Resource Utilization (4 items), Staff Capacity (4 items), and Structure (4 items). A subscale for Process Capacity was not included in this version of the RMT. Within the Innovation-Specific Capacity scale, four subscales were administered: Implementation Climate Supports (3 items), Innovation-Specific Knowledge & Skills (2 items), Interorganizational Relationships (2 items), and Program Champion (2 items). The final Motivation scale included six subscales: Compatibility/Alignment (4 items), Complexity (3 items), Observability (4 items), Priority (2 items), Relative Advantage (3 items), and Trialability (3 items). The text of the items included in this version of the RMT can be found in Table 2.2.

For each item, participants were asked to respond using a 7-point Likert scale (1=*Strongly Disagree*, 4=*Neither Agree or Disagree*, 7=*Strongly Agree*). The three items within the Complexity subscale were reverse-scored before computing subscale averages. For each respondent, an average response value was calculated for each of the 17 administered subscales.

The RMT in Project 2: FP.

The Readiness Monitoring Tool administered in Project 2: FP represents an 81-item version adaptation of the original RMT. Similar to other versions of the RMT, items are grouped within three major scales (i.e., General Capacity, Innovation-Specific Capacity, and Motivation) and further divided into 15 subscales within these three major areas. The General Capacity scale consists of seven subscales, including Climate (8 items), Culture (6 items), Leadership (13 items), Organizational Innovativeness (8 items), Process Capacity (11 items), Resource Utilization (3 items), Staff Capacity (1 item), and Structure (6 items). Innovation-Specific Capacity was assessed using four subscales, including Implementation Climate Supports (5 items), Innovation-Specific Knowledge & Skills (2 items), Interorganizational Relationships (2 items), and Program Champion (3 items). Finally, four subscales were included within Motivation: Compatibility/Alignment (4 items), Complexity (3 items), Priority (3 items), and Relative Advantage (3 items). Two subscales, Observability and Trialability, were not administered as part of the Motivation scale in this project. Table 2.3 presents the text of items included in this adaptation of the RMT.

For each item, respondents were asked to respond using a 7-point Likert scale (1=*Strongly Disagree*, 4=*Neither Agree or Disagree*, 7=*Strongly Agree*). Two of the three items within the Complexity subscale were reverse-scored before computing subscale averages. For each respondent, an average response value was calculated for each of the 15 administered subscales.

Analysis

Assessing Internal Consistency

Alpha coefficients were calculated to assess inter-item consistency within each subscale of the RMT. Subscales within the RMT are designed to each measure a single subcomponent that contributes to readiness. Providing evidence that items within subscales are highly intercorrelated is important for suggesting that all items within that scale are measuring the same construct. Cronbach's coefficient alpha is a widely used measure of internal consistency, accounting for the proportion of total variance amongst a scale's items that is due to the latent variable (DeVellis, 20012). Nunnally (1978) determined .70 as an acceptable lower bound for alpha.

Assessing Reliability and Validity of Organizational Means

Interrater reliability and interrater agreement address questions concerning whether ratings provided by one judge are "similar" to ratings provided by one or more other judges. Whereas interrater agreement is concerned with whether ratings provided by judges are interchangeable or equivalent based on their absolute value, interrater reliability focuses on the relative consistency in ratings provided by multiple judges on multiple targets (LeBreton, Burgess, Kaiser, Atchley, & James, 2003). Estimates of interrater agreement and interrater reliability are often used to justify the aggregation of individual-level data to form organizational-level variables (LeBreton & Senter, 2008).

To assess the reliability and validity of organizational-level means for each subscale in the adapted RMT, several indices of interrater reliability and interrater agreement were computed, including r_{WG}, AD_M, *ICC(1),* and *ICC(2)* estimates.

r_{WG} indices.

r_{WG} indices (James, Demaree, & Wolf, 1993) are one of the most popular ways to assess within-group interrater agreement. Estimates are interpreted as the proportional reduction in error variance, with higher scores indicating higher levels of agreement. A value of .80 suggests that there has been an 80% reduction in error variance because of group membership; the remaining 20% of the observed variance between judges' ratings in that group is due to random responding.

Historically, an r_{WG} value of .70 has been used as a cut point for determining high versus low interrater agreement (LeBreton et al., 2003). Wary of artificial dichotomization, LeBreton & Senter (2008) proposed more inclusive guidelines for interpreting r_{WG} estimates, with ranges indicating very strong agreement (.91 to 1.00), strong agreement (.71 to .90), moderate agreement (.51 to .70), weak agreement (.31 to .50), and a lack of agreement (.00 to .30). Additionally, they suggest that the minimal level of agreement necessary to justify aggregation may also depend on the importance of the decisions being made based on those ratings and the psychometric qualities of the measure used. Newly developed measures used to make decisions that do not pose serious consequences for individuals, such as the RMT when used for directing organizational capacity-building supports, may be able to justify aggregation with lower levels of agreement.

Average deviation (AD_M) indices.

AD_M indices represent another measure of interrater agreement (Burke, Finkelstein, & Dusig, 1999) for use with multiple judges rating a single target on a

variable using an interval scale of measurement. Unlike r_{WG} indices, AD_M estimates agreement in the metric of the original interval scale.

Definitive critical values for determining adequate interrater agreement from AD_M indices have yet to be established, but, as a preliminary guideline, Burke and Dunlap (2002) suggest that for a 7-point scale high agreement is obtained when AD_M estimates based on a uniform response distribution are less than 1.2.

Intraclass correlation coefficients (ICC).

ICCs represent the "proportion of observed variance in ratings that is due to systematic between-target differences compared to the total variance in ratings" (LeBreton & Senter, 2008). ICCs based on the one-way random effects ANOVA are used to provide information about interrater reliability and interrater agreement for a construct measured among multiple organizations rated by different sets of raters (e.g., members at each site rate their own organization rather than all organizations) on an interval measurement scale. Within this study, values for two intraclass correlation coefficients – *ICC(1)* and *ICC(2)* – were calculated from a one-way random-effects ANOVA. ICCs were calculated for each site (organization) on each subscale administered in the two projects.

ICC(1) represents the extent to which raters within a group are alike by estimating the extent to which an individual's ratings on a given measure are explained by higher level units, such as group membership (Bliese, 2000; Klein & Kozlowski, 2000; LeBreton & Senter, 2008). The larger the value of *ICC(1)*, the more alike and the more interchangeable the raters within that group are. Low values can result from low consensus, low consistency, or both, within the group (LeBreton et al., 2003). As *ICC(1)*

values reflect the size of the "effect" of group membership on individual ratings (Bliese, 2000), traditional conventions for interpreting effect sizes may be applied, such that *ICC(1)* = .01 would represent a "small" effect, .10 is interpreted as a "medium" effect, and .25 is considered a "large" effect (Murphy & Myors, 2004).

ICC(2) is a mathematical function of *ICC(1)*, adjusted for group size. *ICC(2)* indicates the extent to which the mean rating assigned by a group of raters is reliable (LeBreton & Senter, 2008). Results can be interpreted as representing the interrater reliability and interrater agreement of the group's average rating. As *ICC(2)* values indicate how reliably the mean rating distinguishes between groups, values between .70 and .85 may be used to justify aggregation.

Table 2.1 *Respondents per Site per Project from Initiatives Using the RMT*

Project	Total No. of Sites (*n*)	No. of Sites with 3 or More Respondents *(n)*	No. of Sites with 5 or More Respondents *(n)*
Community College Academic Programming	9 (22)	4 (17)	2 (10)
Secondary School Safety Initiative	5 (28)	5 (28)	4 (25)
FQHC Integrated Care	10 (36)	8 (34)	4 (21)
Low-Dose Community Health Coalition	21 (27)	2 (8)	0 (0)
Community Health Coalition – Project 1	22 (64)	8 (43)	5 (33)
Fatherhood Program	12 (78)	12 (78)	8 (62)
Community Health Coalition – Project 2	18 (97)	17 (95)	10 (69)

Note. Data from Community Health Coalition – Project 2 and Fatherhood Program were ultimately selected for analysis. Throughout the paper, they are referred to as Project 1: Community Health Coalitions (CHC) and Project 2: Fatherhood Programs (FP), respectively.

Table 2.2 *Item List for RMT Administered in Project 1: CHC*

Scale	Subscale	Item Text
General Capacity	Climate (4 items)	Morale is positive in our community coalition.
		Turnover is not a problem in our community coalition.
		The members of our community coalition who work on our projects generally feel valued.
		We feel positively about our community coalition's work.
	Culture (6 items)	Our community coalition's mission statement is understood by all of us.
		We all know our community coalition's vision.
		We have a strong sense of belonging and identification with our community coalition.
		Our community coalition has a common purpose.
		We know the goals of our community coalition.
		We put in extra effort to make sure our community coalition succeeds.
	Leadership (7 items)	We have clear leadership in our community coalition.
		Our community coalition's leadership has a plan(s) to implement our projects.
		Our community coalition's leadership knows what they are talking about when it comes to our projects.
		Our community coalition's leadership recognizes and appreciates team efforts that help us to successfully implement projects.
		Our community coalition's leadership supports our efforts by learning more about our projects.
		Our community coalition's leadership carries on through the challenges of implementing our projects.
		Our community coalition's leadership reacts to critical issues regarding the implementation of our projects by openly and effectively addressing the problem(s).
	Organizational Innovativeness (6 items)	Our community coalition regularly takes time to consider ways to improve how we do things.
		People in our community coalition actively try to improve how we do things.
		When we experience a problem in our community coalition, we make a serious effort to find a new way of doing things.
		Our community coalition is strategic in how we approach change.
		Overall, our community coalition adapts well to change.
		Our community coalition can quickly change procedures to meet new conditions and solve problems as they arise.
	Process Capacity (0 items)	--
	Resource Utilization (4 items)	Our community coalition has the ability to access diverse sources of revenue.
		There is a clear financial plan for our community coalition to create sustainability of our projects.
		There is a clear process by which our community coalition prioritizes and distributes resources.
		Our community coalition knows how to sustain progress when something is going well.

Table 2.2, cont. *Item List for RMT Administered in Project 1: CHC*

Scale	Subscale	Item Text
General Capacity	Staff Capacity (4 items)	The people within our community coalition have sufficient skills to carry out our day-to-day tasks.
		The people within our community coalition have sufficient knowledge to carry out our day-to-day tasks.
		People who work with our community coalition have adequate experience.
		We have enough people in our community coalition to work toward our major goals.
	Structure (4 items)	The way our community coalition is organized makes it possible to do things well.
		Our community coalition's structure is effective.
		Our community coalition functions well.
		We communicate well with each other within our community coalition.
Innovation-Specific Capacity	Implementation Climate Supports (3 items)	Our community coalition actively supports becoming a Community of Solutions.
		Our community coalition has a system in place to monitor how well Community of Solutions activities are implemented.
		Our community coalition has ways to promote ongoing activities to become a Community of Solutions.
	Innovation-Specific Knowledge & Skills (2 items)	Our community coalition has the knowledge needed to become a Community of Solutions
		Our community coalition has the concrete skills needed to become a Community of Solutions.
	Interorganizational Relationships (2 items)	Our community coalition communicates well with other coalitions who are also working on Community of Solution's transformation.
		Our community coalition obtains support from IHI/coaches to help us become a Community of Solutions.
	Program Champion (2 items)	An influential person in our community coalition strongly promotes becoming a Community of Solutions.
		At least one person we work with clearly communicates the needs and benefits of becoming a Community of Solutions.
Motivation	Compatibility/ Alignment (4 items)	Becoming a Community of Solutions fits well with other initiatives in our community coalition.
		Becoming a Community of Solutions is timely given the current needs of our community.
		Becoming a Community of Solutions fits well with the culture and values of our community coalition
		Becoming a Community of Solutions is feasible for our community coalition.
	Complexity (3 items)	*It is difficult for us to become a Community of Solutions.
		*The Community of Solutions model is hard for us to understand.
		*The many different parts in a Community of Solutions makes it complicated for us to implement.
	Observability (4 items)	Our community coalition can already see some results of becoming Community of Solutions.
		Our community coalition has seen the Community of Solutions model work in other places.
		We see other coalitions becoming Communities of Solutions.
		We are likely to see benefits soon from the Community of Solutions model.

Table 2.2, cont. *Item List for RMT Administered in Project 1: CHC*

Scale	Subscale	Item Text
Motivation	Priority (2 items)	Becoming a Community of Solutions is one of our community coalition's top three priorities.
		Our community coalition emphasizes that becoming a Community of Solution is very important.
	Relative Advantage (3 items)	Becoming a Community of Solutions is a better strategy than others we have tried before in our community coalition.
		Becoming a Community of Solutions has advantages for our community coalition.
		Becoming a Community of Solutions represents an advance over other models of community change that are already available to our community coalition.
	Trialability (3 items)	Our community coalition is able to try out becoming a Community of Solutions in a limited way.
		Our community coalition can test small parts of Community of Solutions to see if it is working.
		If we try becoming Community of Solutions and things don't go well, our community coalition can go back to the way we used to do things.

Note. Three additional items measuring "joy" as a subcomponent of motivation were included in this survey at the request of the project funder. They have been excluded from consideration here because they are not a recognized subcomponent within the $R = MC^2$ heuristic.
-- indicates subscales that were not included in the project's adapted version of the RMT; * indicates reverse-scored item

Table 2.3 *Item List for RMT Administered in Project 2: FP*

Scale	Subscale	Item Text
General Capacity	Climate (8 items)	Most of the time, people in this local Site want to perform to the best of their abilities.
		People are enthusiastic about their work.
		We put in extra effort to make sure our local Site succeeds.
		Our workload is reasonable.
		Morale is positive in our local Site.
		We have a positive attitude toward the work of the local Site.
		Turnover is not a problem in our local Site.
		People who work within our local Site feel valued.
	Culture (6 items)	Our local Site's mission statement is understood by all of us.
		We all know our local Site's vision.
		We have a strong sense of belonging and identification within the local Site.
		Our local Site has a common purpose.
		We have good working relationships within our local Site.
		We know the goals of our local Site.
	Leadership (13 items)	We have clear leadership in our local Site.
		Our leadership supports ongoing projects.
		Our leadership approaches collaboration by relying heavily on building trust among stakeholders.
		Our leadership expresses confidence in the capabilities of others.
		Our leadership praises/recognizes when someone has done something well.
		Our leadership has a plan to implement our projects.
		Our leadership removes obstacles that prevent our programs from being implemented.
		Our leadership lays out the standards we need to aspire to when putting our programs into practice.
		Our leadership knows what they are talking about when it comes to our projects.
		Our leadership recognizes and appreciates team efforts to help us successfully implement.
		Our leadership supports our efforts to learn more about our projects.
		Our leadership carries on through the challenges of implementing our projects.
		Our leadership reacts to critical issues regarding the implementation of our projects by openly and effectively addressing the problem(s).
	Organizational Innovativeness (8 items)	We regularly take time to consider ways to improve how we do things.
		People in our local Site actively try to improve how we do things.
		Our local Site encourages everyone to share their ideas.
		Our local Site listens to people who have new ideas.
		Our local Site learns from its mistakes.
		When we experience a problem in the local Site, we make a serious effort to figure out what's really going on.
		We are deliberate in how we approach change.
		Overall, our local Site adapts to change well.

Table 2.3, cont. *Item List for RMT Administered in Project 2: FP*

Scale	Subscale	Item Text
General Capacity	Process Capacity (11 items)	We are able to use strategic planning frameworks to accomplish our goals.
		We know how to conduct a comprehensive needs assessment.
		We are able to develop appropriate goals for our local Site.
		We know how to select an evidence-based strategy that best fits with our local Site and community's needs.
		We know how to identify the capacities needed to put our strategies into place.
		We are able to develop strategic plans.
		We know what it takes to put our innovations into place.
		We know how to evaluate what we do.
		We know how to evaluate if our innovations are reaching our desired outcomes and goals.
		We are able to implement ongoing improvement activities (e.g., PDSA, Six Sigma, Model for Improvement, etc.).
		We know how to plan for sustainability at our local Site.
	Resource Utilization (3 items)	We have the ability to access sources of revenue and resources (e.g., multiple grants, public funds, third party private payers, etc.).
		There is a clear financial plan for us to create sustainability.
		There is a clear process by which the local Site prioritizes and distributes resources.
	Staff Capacity (1 item)	People who work with our local Site have experience working towards program improvement.
	Structure (6 items)	Our local Site can quickly change procedures to meet new conditions and solve problems as they arise.
		Our leadership committees (e.g. board, advisory, or steering) actively contribute to the goals of our local Site.
		We are able to communicate openly within our local Site.
		We understand each other when communicating within our local Site.
		We have a well-defined method to resolve internal problems.
		There is a clear method for sharing information within the local Site.
Innovation-Specific Capacity	Implementation Climate Supports (5 items)	Our local Site actively supports the regular review of data for program improvement.
		We have enough resources at our local Site to regularly review data for program improvement.
		Our local Site dedicates specific resources to regularly review data for program improvement.
		There is a system in place to monitor how well we regularly review data for program improvement.
		We have ways to promote ongoing participation in the regular review of data.
	Innovation-Specific Knowledge & Skills (2 items)	We have the knowledge we need to regularly review data for program improvement.
		We have the concrete skills to regularly review data for program improvement.

Table 2.3, cont. *Item List for RMT Administered in Project 2: FP*

Scale	Subscale	Item Text
Innovation-Specific Capacity	Interorganizational Relationships (2 items)	We communicate well with similar fatherhood organizations that focus on regularly reviewing data for program improvement.
		We seek consultation from other organizations to help us regularly review data for program improvement.
	Program Champion (3 items)	An influential person in our local Site strongly promotes the regular review of data for program improvement.
		At least one person we work with clearly communicates the needs and benefits of regularly reviewing data for program improvement.
		We have designated a person to share our progress in how we regularly review data for program improvement.
Motivation	Compatibility/ Alignment (4 items)	Regularly reviewing data for program improvement fits well with other initiatives at our local Site.
		Regularly reviewing data for program improvement will help us track the desired outcomes at our local Site.
		Regularly reviewing data for program improvement is timely given the current needs at our local Site.
		Regularly reviewing data for program improvement fits well with the culture and values of our local Site.
	Complexity (3 items)	At our local Site, regularly reviewing data for program improvement is simple and easy to implement.
		*There are so many components to regularly reviewing data for program improvement that it is hard to understand all of the pieces.
		*The complexity of regularly reviewing data for program improvement will make it difficult to put it into place.
	Observability (0 items)	--
	Priority (3 items)	Regularly reviewing data for program improvement is a top priority at our local Site.
		Our local Site emphasizes that regularly reviewing data for program improvement is very important to improve the quality of our services.
		We are aware of how important regularly reviewing data for program improvement is at our local Site right now.
	Relative Advantage (3 items)	Regularly reviewing data for program improvement represents an advance over other methods that are already available for our local Site.
		Regularly reviewing data for program improvement is better than other processes we have considered using at our local Site.
		Regularly reviewing data for program improvement is better than other processes we are currently using at our local Site to monitor and improve our outcomes.
	Trialability (0 items)	--

Note. -- indicates subscales that were not included in the project's adapted version of the RMT; * indicates reverse-scored item

CHAPTER 3

RESULTS

Data analyses for the current study were performed using IBM SPSS Statistics software version 24. Syntax for computing r_{WG} and AD_M indices was modeled on the work of LeBreton and Senter (2008).

Internal Consistency

A Cronbach's alpha coefficient was calculated in order to establish the reliability of each subscale within the Readiness Monitoring Tool (RMT) administered in Project 1: CHC and Project 2: FP. An alpha coefficient was calculated for each subscale administered in the project's adapted version of the RMT. The alpha coefficient was based on the entire population of respondents for Project 1 (n=67-69, some respondents had incomplete data for subscales and were excluded via listwise deletion for that particular subscale) and Project 2 (n=62). See Table 3.1 for results.

The majority of subscales demonstrated very good internal consistency ($\alpha \geq .80$) (DeVellis, 2012). Those which did not meet the acceptable lower bound for alpha suggested by Nunnally (1978) included Trialability ($\alpha = .645$) within Project 1 and Program Champion ($\alpha = .695$), Complexity ($\alpha = .633$), and Relative Advantage ($\alpha = .653$) within Project 2. It should be noted that none of these alpha coefficients fall below the unacceptable standard ($\alpha < .60$) described by DeVellis (2012) in his book on scale development.

r_{WG} Indices

r_{WG} index estimates for each project sample and site can be found in Tables 3.2-3.4 for Project 1: CHC and Tables 3.6-3.8 for Project 2: FP. Interpretation of these estimates was based on guidelines proposed by LeBreton and Senter (2008).

In Project 1: CHC, r_{WG} sample values from the collection of 10 community health coalitions indicated strong agreement for 82% of the RMT subscales and moderate agreement for 18% of the subscales (M=.76; SD=.07; range: .59-.85). The mean, range, and standard deviation for the r_{WG} estimate for each subscale, along with the percentage of sites with r_{WG} values indicating strong or very strong agreement, moderate agreement, and weak or no agreement can be found in Table 3.5.

r_{WG} sometimes varied extensively between sites within a subscale and between subscales within a site. One site, CHC 13, demonstrated much lower levels of interrater agreement (M=.47; SD=.20; range: .05-.76) than other sites within the sample, with 58.8% of subscales demonstrating weak or no agreement. Subscales demonstrating weak or no agreement at this site included six of the seven General Capacity subscales (Climate, Culture, and Resource Utilization demonstrated no agreement, whereas r_{WG} estimates for Organizational Innovativeness, Staff Capacity, and Structure suggested weak agreement), two of the four Innovation-Specific Capacity subscales (Interorganizational Relationships and Program Champion both demonstrated weak agreement), and two of the six Motivation subscales (r_{WG} estimates suggest weak agreement for Observability and no agreement for Complexity).

Results from the Project 2: FP sample were consistent with findings from Project 1: CHC. r_{WG} sample values from the collection of eight fatherhood programs indicated

strong agreement for 81% of the RMT subscales and moderate agreement for 19% of the subscales (*M*=.76; *SD*=.07; range: .62-.85). The mean, range, and standard deviation for the r_{WG} estimate for each subscale, along with the percentage of sites with r_{WG} values indicating strong or very strong agreement, moderate agreement, and weak or no agreement can be found in Table 3.9.

As with Project 1: CHC, r_{WG} in Project 2: FP sometimes varied extensively between sites within a subscale and between subscales within a site. One site, FP 10, demonstrated much lower levels of interrater agreement (*M*=.35; *SD*=.31; range: -.06-.79) than other sites within the sample, with 63% of subscales demonstrating weak or no agreement. These subscales included seven of the eight General Capacity subscales (the r_{WG} estimates for Climate, Culture, Leadership, Organizational Innovativeness, and Structure all indicated no agreement), two of the four Innovation-Specific subscales (r_{WG} estimates for Implementation Climate Supports and Innovation-Specific Knowledge indicated weak and no agreement, respectively), and one of the four Motivation subscales (Complexity demonstrated only weak agreement).

AD_M Indices

AD_M index estimates for each project sample and site can be found in Tables 3.2-3.4 for Project 1: CHC and Tables 3.6-3.8 for Project 2: FP. As suggested by Burke and Dunlap (2002), AD_M estimates were interpreted as indicating high interrater agreement when valued at less than 1.2.

In Project 1: CHC, 100% of the AD_M sample values from the collection of 10 community health coalitions indicated high agreement for each of the RMT subscales. Across subscales, the vast majority of CHC sites demonstrated high agreement based on

AD_M values. However, one site (CHC 13) demonstrated markedly lower interrater agreement than the others, with AD_M values on seven of the 17 subscales failing to indicate adequate interrater agreement. These subscales included five of the seven General Capacity subscales (Climate, Culture, Organizational Innovativeness, Resource Utilization, Staff Capacity), one of the four Innovation-Specific Capacity subscales (Program Champion), and one of the six Motivation subscales (Complexity).

Similar results were found in Project 2: FP: 100% of the AD_M sample values from the group of eight fatherhood programs demonstrated high agreement for each of the RMT subscales. Across subscales, the vast majority of fatherhood program sites demonstrated high agreement based on AD_M values. However, one site (FP 10) indicated markedly lower interrater agreement than the others, with AD_M values on seven of the 16 subscales failing to indicate adequate interrater agreement. These subscales included five of the eight General Capacity subscales (Climate, Leadership, Organizational Innovativeness, Process Capacity, Resource Utilization), one of the four Innovation-Specific Capacity subscales (Innovation-Specific Knowledge and Skills), and one of the four Motivation subscales (Complexity).

Intraclass Correlation Coefficients

Intraclass correlation coefficients with 95% confidence interval estimates for each subscale can be found in Table 3.10 and Table 3.11 for Project 1: CHC and Project 2: FP, respectively.

Project 1: CHC

General Capacity subscales.

The one-way ANOVA for the Climate subscale yielded an *ICC(1)* of .402 and an *ICC(2)* of .729 (p<.001). The one-way ANOVA for the Culture subscale yielded an *ICC(1)* of .610 and an *ICC(2)* of .904 (p<.002). The one-way ANOVA for the Leadership subscale yielded an *ICC(1)* of .546 and an *ICC(2)* of .894 (p<.05). The one-way ANOVA for the Organizational Innovativeness subscale yielded an *ICC(1)* of .653 and an *ICC(2)* of .919 (p=.115). The one-way ANOVA for the Resource Utilization subscale yielded an *ICC(1)* of .678 and an *ICC(2)* of .894 (p<.001). The one-way ANOVA for the Staff Capacity subscale yielded an *ICC(1)* of .539 and an *ICC(2)* .824 of (p<.001). The one-way ANOVA for the Structure subscale yielded an *ICC(1)* of .737 and an *ICC(2)* of .918 (p=.742).

Innovation-Specific Capacity subscales.

The one-way ANOVA for the Implementation Climate Supports subscale yielded an *ICC(1)* of .527 and an *ICC(2)* of .770 (p<.001). The one-way ANOVA for the Innovation-Specific Knowledge & Skills subscale yielded an *ICC(1)* of .746 and an *ICC(2)* of .854 (p=.172). The one-way ANOVA for the Interorganizational Relationships subscale yielded an *ICC(1)* of .641 and an *ICC(2)* of .781 (p<.001). The one-way ANOVA for the Program Champion subscale yielded an *ICC(1)* of .782 and an *ICC(2)* of .878 (p=.775).

Motivation subscales.

The one-way ANOVA for the Compatibility/Alignment subscale yielded an *ICC(1)* of .549 and an *ICC(2)* of .830 (p=.185). The one-way ANOVA for the

Complexity subscale yielded an *ICC(1)* of .737 and an *ICC(2)* of .893 (p=.08). The one-way ANOVA for the Observability subscale yielded an *ICC(1)* of .630 and an *ICC(2)* of .872 (p<.001). The one-way ANOVA for the Priority subscale yielded an *ICC(1)* of .584 and an *ICC(2)* of .738 (p<.001). The one-way ANOVA for the Relative Advantage subscale yielded an *ICC(1)* of .528 and an *ICC(2)* of .770 (p<.001). The one-way ANOVA for the Trialability subscale yielded an *ICC(1)* of .310 and an *ICC(2)* of .574 (p<.001).

Summary of ICC results for Project 1: CHC.

Per interpretive conventions, the *ICC(1)* values for the general capacity, innovation-specific capacity, and motivation subscales reflect a "large" effect size of group membership on individual ratings and the *ICC(2)* values suggest an adequate level of reliability to justify aggregation. Only the *ICC(2)* estimate for the Trialability subscale failed to support aggregation of individual responses into a site-based average score.

Project 2: FP

General Capacity subscales.

The one-way ANOVA for the Climate subscale yielded an *ICC(1)* of .505 and an *ICC(2)* of .891 (p<.001). The one-way ANOVA for the Culture subscale yielded an *ICC(1)* of .684 and an *ICC(2)* of .929 (p=.208). The one-way ANOVA for the Leadership subscale yielded an *ICC(1)* of .675 and an *ICC(2)* of .964 (p<.001). The one-way ANOVA for the Organizational Innovativeness subscale yielded an *ICC(1)* of .691 and an *ICC(2)* of .947 (p<.001). The one-way ANOVA for the Process Capacity subscale yielded an *ICC(1)* of .653 and an *ICC(2)* of .954 (p<.001). The one-way ANOVA for the Resource Utilization subscale yielded an *ICC(1)* of .631 and an *ICC(2)* of .837 (p<.001).

The one-way ANOVA for the Structure subscale yielded an *ICC(1)* of .577 and an *ICC(2)* of .891 (p<.001).

Innovation-Specific Capacity subscales.

The one-way ANOVA for the Implementation Climate Supports subscale yielded an *ICC(1)* of .484 and an *ICC(2)* of .824 (p<.001). The one-way ANOVA for the Innovation-Specific Knowledge & Skills subscale yielded an *ICC(1)* of.767 and an *ICC(2)* of .868 (p=.063). The one-way ANOVA for the Interorganizational Relationships subscale yielded an *ICC(1)* of .543 and an *ICC(2)* of .704 (p<.001). The one-way ANOVA for the Program Champion subscale yielded an *ICC(1)* of .407 and an *ICC(2)* of .673 (p<.01).

Motivation subscales.

The one-way ANOVA for the Compatibility/Alignment subscale yielded an *ICC(1)* of .360 and an *ICC(2)* of .692 (p<.001). The one-way ANOVA for the Complexity subscale yielded an *ICC(1)* of .326 and an *ICC(2)* of .592 (p<.001). The one-way ANOVA for the Priority subscale yielded an *ICC(1)* of .430 and an *ICC(2)* of .693 (p<.001). The one-way ANOVA for the Relative Advantage subscale yielded an *ICC(1)* of .350 and an *ICC(2)* of .618 (p<.001).

Summary of ICC results for Project 2: FP.

Per interpretive conventions, the *ICC(1)* values for the general capacity, innovation-specific capacity, and motivation subscales all indicate a “large” effect of group membership on individual ratings. *ICC(2)* values support group-level aggregation for the majority of subscales in general and innovation-specific capacity. The *ICC(2)* values for the Program Champion, Compatibility/Alignment, Complexity, Priority, and

Relative Advantage subscales approach but do not meet the minimum recommended standard for justifying aggregation (LeBreton & Senter, 2008).

Table 3.1 *Internal Consistency Estimates of RMT Subscales in Selected Projects*

		Cronbach's Alpha Coefficient (α)	
Scale	Subscale	Project 1: CHC (n=67-69)	Project 2: FP (n=62)
General Capacity	Climate	.809	.900
	Culture	.908	.929
	Leadership	.896	.966
	Organizational Innovativeness	.920	.950
	Process Capacity	--	.959
	Resource Utilization	.920	.852
	Staff Capacity	.844	--
	Structure	.917	.903
Innovation-Specific Capacity	Implementation Climate Supports	.813	.853
	Innovation-Specific Knowledge & Skills	.856	.874
	Interorganizational Relationships	.812	.754
	Program Champion	.876	.695
Motivation	Compatibility/Alignment	.831	.727
	Complexity	.896	.633
	Observability	.900	--
	Priority	.778	.725
	Relative Advantage	.812	.653
	Trialability	.645	--

Note. -- indicates subscales that were not included in the project's adapted version of the RMT

Table 3.2 *r_{WG} and AD_M Values for Project 1: CHC: General Capacity Subscales*

	Climate			Culture			Leadership			Organizational Innovativeness			Resource Utilization			Staff Capacity			Structure		
Site	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M
CHC 1	5.82	.94	.42	5.71	.95	.34	6.00	.96	.29	5.31	.85	.55	4.93	.75	.85	6.14	.94	.41	5.71	.99	.12
CHC 2	5.18	.73	.72	5.62	.87	.52	5.53	.81	.66	4.86	.55	.94	3.64	.06	1.38	4.99	.90	.51	4.79	.51	.98
CHC 3	6.83	.97	.22	6.94	.99	.10	7.00	1.00	.00	6.94	1.00	.09	6.86	.96	.25	6.94	.99	.10	7.00	1.00	.00
CHC 4	5.29	.83	.63	5.67	.74	.78	5.76	.72	.86	5.25	.77	.69	3.33	.77	.75	5.63	.56	.96	5.54	.91	.46
CHC 6	5.80	.97	.26	5.42	.79	.63	5.89	.99	.19	5.57	.96	.32	4.80	.85	.54	5.75	.95	.30	5.55	.75	.72
CHC 7	5.58	.82	.69	5.67	.78	.74	5.63	.80	.71	5.30	.83	.63	5.08	.84	.63	5.61	.76	.77	5.17	.58	1.00
CHC 10	5.68	.97	.28	5.81	.92	.37	6.06	.95	.29	5.76	.91	.44	5.18	.78	.74	5.82	.93	.40	5.86	.93	.35
CHC 13	4.80	.17	1.33	4.88	.27	1.28	5.27	.70	.87	5.04	.42	1.27	3.82	.05	1.58	5.25	.36	1.29	4.96	.41	1.11
CHC 17	5.25	.82	.60	5.00	.77	.67	5.86	.70	.74	4.90	.61	1.05	4.55	.90	.54	4.70	.57	.96	4.45	.55	1.04
CHC 18	5.46	.94	.40	5.93	.93	.41	5.45	.77	.79	5.48	.74	.79	4.57	.99	.15	5.71	.96	.33	5.64	.83	.62
CHC Mean	*5.57*	*.82*	*.56*	*5.67*	*.80*	*.58*	*5.85*	*.84*	*.54*	*5.44*	*.76*	*.68*	*4.68*	*.70*	*.74*	*5.65*	*.79*	*.60*	*5.47*	*.75*	*.64*

Table 3.3 *r_{WG} and AD_M Values for Project 1: CHC: Innovation-Specific Capacity Subscales*

	Implementation Climate Supports			Innovation-Specific Knowledge & Skills			Interorganizational Relationships			Program Champion		
Site	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M
CHC 1	5.10	.82	.75	6.29	.92	.47	4.64	.71	.88	6.14	.86	.59
CHC 2	5.14	.72	.73	5.29	.77	.69	4.64	.19	1.08	6.00	.85	.57
CHC 3	7.00	1.00	.00	7.00	1.00	.00	7.00	1.00	.00	7.00	1.00	.00
CHC 4	4.67	.58	1.00	5.58	.94	.42	4.83	.73	.83	5.50	.50	1.00
CHC 6	5.47	.83	.69	5.50	.75	.80	5.20	.92	.44	4.70	.83	.64
CHC 7	4.37	.54	1.00	5.50	.52	1.00	4.33	.30	1.37	5.61	.78	.68
CHC 10	6.05	.91	.35	6.36	.94	.41	5.86	.86	.49	6.07	.82	.53
CHC 13	4.57	.59	.94	5.36	.59	.98	4.50	.48	1.14	5.07	.49	1.20
CHC 17	5.163	.81	.64	5.00	.84	.60	5.00	.84	.60	5.90	.61	.92
CHC 18	4.67	.85	.67	5.43	.78	.65	4.79	.84	.67	4.50	.31	1.29
CHC Mean	*5.22*	*.77*	*.68*	*5.73*	*.81*	*.60*	*5.08*	*.69*	*.75*	*5.65*	*.71*	*.74*

Table 3.4 *r_{WG} and AD_M Values for Project 1: CHC: Motivation Subscales*

	Compatibility/Alignment			Complexity			Observability			Priority			Relative Advantage			Trialability		
Site	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M
CHC 1	6.13	.95	.32	4.62	.57	1.10	4.64	.87	.55	4.50	.83	.57	5.24	.78	.72	4.50	.64	.86
CHC 2	5.89	.83	.63	4.52	.64	.84	5.24	.62	.94	5.64	.69	.88	5.86	.74	.78	5.24	.67	.78
CHC 3	7.00	1.00	.00	7.00	1.00	.00	7.00	1.00	.00	7.00	1.00	.00	7.00	1.00	.00	2.30	.26	1.44
CHC 4	6.04	.99	.14	5.00	.74	.78	4.42	.65	.92	4.42	.44	1.06	5.33	.91	.44	4.83	.81	.67
CHC 6	6.00	1.00	.00	4.27	.52	1.01	4.90	.96	.32	5.40	.93	.48	5.47	.92	.43	4.67	.83	.53
CHC 7	5.78	.75	.73	5.37	.80	.63	4.42	.43	1.22	4.56	.56	1.06	5.41	.74	.84	4.52	.92	.50
CHC 10	6.21	.95	.38	4.95	.38	1.12	5.18	.80	.78	5.29	.77	.82	5.76	.95	.37	4.57	.78	.68
CHC 13	5.50	.56	1.00	3.19	.30	1.36	4.93	.49	1.15	5.07	.66	.94	5.24	.76	.75	5.33	.76	.57
CHC 17	5.75	.69	.90	4.67	.68	.93	5.25	.80	.70	4.80	.89	.56	5.53	.63	.96	4.40	.76	.72
CHC 18	5.75	.74	.79	4.62	.30	1.37	4.71	.87	.54	4.64	.90	.55	5.29	.76	.71	4.24	.90	.49
CHC Mean	*6.01*	*.85*	*.49*	*4.82*	*.59*	*.91*	*5.07*	*.75*	*.71*	*5.13*	*.77*	*.69*	*5.61*	*.82*	*.60*	*4.46*	*.73*	*.72*

Table 3.5 *Summary of r_{WG} Estimate Statistics by Subscale for Project 1: CHC*

Scale	Subscale	*M*	Range	*SD*	% of sites with Very Strong or Strong Agreement	% of sites with Moderate Agreement	% of sites with Weak or No Agreement
General Capacity	Climate	.82	.17-.97	.24	90%	0%	10%
	Culture	.80	.27-.99	.21	90%	0%	0%
	Leadership	.84	.70-1	.12	80%	20%	0%
	Organizational Innovativeness	.76	.42-1	.19	70%	20%	10%
	Process Capacity	--	--	--	--	--	--
	Resource Utilization	.70	.05-.99	.35	80%	0%	20%
	Staff Capacity	.79	.36-.99	.22	70%	20%	10%
	Structure	.75	.41-.99	.22	60%	30%	10%
Innovation-Specific Capacity	Implementation Climate Supports	.77	.54-1	.15	70%	30%	0%
	Innovation-Specific Knowledge & Skills	.81	.52-1	.16	80%	20%	0%
	Interorganizational Relationships	.69	.19-1	.27	70%	0%	30%
	Program Champion	.71	.31-1	.22	60%	10%	30%
Motivation	Compatibility/Alignment	.85	.56-1	.16	80%	20%	0%
	Complexity	.59	.30-1	.23	30%	40%	30%
	Observability	.75	.43-1	.19	60%	20%	20%
	Priority	.77	.44-1	.18	60%	30%	10%
	Relative Advantage	.82	.63-1	.12	90%	10%	0%
	Trialability	.73	.26-.92	.19	70%	20%	10%

Table 3.6 *r_{WG} and AD_M Values for Project 2: FP: General Capacity Subscales*

	Climate			Culture			Leadership			Organizational Innovativeness			Process Capacity			Resource Utilization			Staff Capacity			Structure		
Site	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M
FP 2	4.98	.78	.82	5.27	.56	1.12	5.15	.64	.92	4.75	.56	1.15	4.75	.94	.32	4.67	.76	.80	5.00	.75	.80	4.40	.70	.81
FP 3	6.58	.98	.22	6.78	.97	.26	6.53	.97	.30	6.61	.96	.35	6.28	.85	.62	6.30	.90	.49	6.22	.76	.69	6.56	.96	.32
FP 5	5.88	.88	.50	6.14	.91	.47	6.19	.93	.42	6.04	.88	.54	5.42	.91	.43	5.11	.87	.56	5.83	.86	.56	5.56	.87	.59
FP 6	5.80	.91	.46	6.14	.78	.67	5.69	.85	.61	5.68	.85	.55	5.35	.78	.78	5.03	.74	.88	5.91	.83	.51	5.68	.81	.68
FP 7	6.31	.91	.52	6.39	.95	.39	6.51	.95	.41	6.21	.80	.67	6.03	.91	.42	6.17	.66	.83	5.83	.76	.61	6.14	.83	.58
FP 10	5.44	-.01	1.44	6.02	.24	1.14	5.69	.11	1.21	5.36	-.03	1.49	5.23	.05	1.40	4.79	.12	1.34	5.38	.58	1.03	5.35	.23	1.18
FP 11	5.96	.96	.26	6.02	.90	.41	5.37	.74	.78	5.80	.82	.65	5.00	.66	.99	4.24	.43	.97	5.71	.86	.61	5.50	.86	.52
FP 12	5.84	.92	.46	6.62	.95	.39	6.28	.93	.42	5.95	.91	.47	5.45	.90	.49	5.47	.87	.57	6.10	.81	.54	5.32	.95	.32
FP Mean	*5.85*	*.79*	*.59*	*6.17*	*.78*	*.61*	*5.93*	*.77*	*.63*	*5.80*	*.72*	*.73*	*5.44*	*.75*	*.68*	*5.22*	*.67*	*.81*	*5.75*	*.78*	*.67*	*5.56*	*.78*	*.63*

Table 3.7 *r_{WG} and AD_M Values for Project 2: FP: Innovation-Specific Capacity Subscales*

	Implementation Climate Supports			Innovation-Specific Knowledge & Skills			Interorganizational Relationships			Program Champion		
Site	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M
FP 2	5.24	.86	.61	4.6	.89	.52	4.00	.66	.80	4.93	.77	.75
FP 3	6.49	.92	.46	6.50	.89	.44	5.67	.81	.63	6.48	.87	.62
FP 5	5.53	.91	.49	5.58	.89	.58	4.83	.56	1.11	6.06	.85	.63
FP 6	5.69	.90	.54	6.18	.90	.50	4.73	.73	.88	6.36	.91	.52
FP 7	5.97	.73	.66	5.83	.73	.72	5.67	.83	.56	5.89	.68	.81
FP 10	5.55	.45	1.16	5.94	-.06	1.34	5.06	.56	1.06	6.25	.70	.94
FP 11	5.29	.80	.76	5.86	.90	.45	5.07	-.09	1.61	5.48	.81	.60
FP 12	4.80	.90	.44	4.50	.64	1.00	6.50	.92	.50	5.50	.79	.73
FP Mean	*5.57*	*.81*	*.64*	*5.62*	*.72*	*.69*	*5.19*	*.62*	*.89*	*5.87*	*.80*	*.70*

Table 3.8 *r_{WG} and AD_M Values for Project 2: FP: Motivation Subscales*

	Compatibility/Alignment			Complexity			Priority			Relative Advantage		
Site	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M	*M*	r_{WG}	AD_M
FP 2	5.5	.74	.70	4.47	.76	.69	5.53	.63	1.04	5.27	.93	.45
FP 3	6.19	.92	.45	5.96	.91	.48	6.44	.96	.35	5.74	.81	.67
FP 5	5.96	.92	.46	4.94	.72	.83	6.50	.96	.33	5.39	.92	.39
FP 6	5.98	.94	.31	4.79	.65	1.08	6.24	.91	.51	5.15	.78	.87
FP 7	6.04	.71	.71	5.17	.55	1.06	6.00	.70	.67	5.50	.81	.67
FP 10	6.22	.73	.73	4.71	.39	1.21	6.46	.79	.73	6.00	.76	.67
FP 11	6.21	.94	.40	5.38	.80	.72	6.29	.97	.24	5.10	.85	.65
FP 12	5.60	.87	.58	4.30	.77	.71	6.03	.89	.45	4.63	.86	.64
FP Mean	*5.96*	*.85*	*.54*	*4.97*	*.69*	*.85*	*6.19*	*.85*	*.54*	*5.35*	*.847*	*.63*

Table 3.9 *Summary of r_{WG} Estimate Statistics by Subscale for Project 2: FP*

					% of sites with		
Scale	Subscale	*M*	Range	*SD*	Very Strong or Strong Agreement	Moderate Agreement	Weak or No Agreement
General Capacity	Climate	.79	-.01-.98	.33	87.5%	0%	12.5%
	Culture	.78	.24-.97	.26	75%	12.5%	12.5%
	Leadership	.77	.11-.97	.29	75%	12.5%	12.5%
	Organizational Innovativeness	.72	-.03-.96	.33	75%	12.5%	12.5%
	Process Capacity	.75	.05-.94	.30	75%	12.5%	12.5%
	Resource Utilization	.67	.12-.90	.27	62.5%	12.5%	25%
	Staff Capacity	.78	.58-.86	.09	87.5%	12.5%	0%
	Structure	.78	.45-.92	.16	87.5%	0%	12.5%
Innovation-Specific Capacity	Implementation Climate Supports	.81	.45-.92	.16	87.5%	0%	12.5%
	Innovation-Specific Knowledge & Skills	.72	-.06-.90	.33	75%	12.5%	12.5%
	Interorganizational Relationships	.62	-.09-.92	.32	50%	37.5%	12.5%
	Program Champion	.80	.68-.91	.08	75%	25%	0%
Motivation	Compatibility/Alignment	.85	.39-.91	.17	62.5%	25%	12.5%
	Complexity	.69	.39-.91	.17	62.5%	25%	12.5%
	Observability	--	--	--	--	--	--
	Priority	.85	.63-.97	.13	75%	25%	0%
	Relative Advantage	.85	.76-.93	.06	100%	0%	0%
	Trialability	--	--	--	--	--	--

Table 3.10 *Intraclass Correlation Coefficients and 95% Confidence Interval Estimates for Project 1: CHC*

				95% CI for ICC(1)			*95% CI for ICC(2)*	
Scale	Subscale	N	ICC(1)	*Lower Bound*	*Upper Bound*	*ICC(2)*	*Lower Bound*	*Upper Bound*
General Capacity	Climate	67	.402	.277	.534	.729	.605	.821
	Culture	68	.610	.513	.707	.904	.864	.935
	Leadership	69	.546	.448	.648	.894	.850	.928
	Organizational Innovativeness	67	.653	.560	.743	.919	.884	.945
	Process Capacity	--	--	--	--	--	--	--
	Resource Utilization	69	.678	.579	.767	.894	.846	.930
	Staff Capacity	67	.539	.420	.655	.824	.743	.884
	Structure	67	.737	.648	.814	.918	.880	.946
Innovation-Specific Capacity	Implementation Climate Supports	69	.527	.390	.654	.770	.657	.850
	Innovation-Specific Knowledge & Skills	69	.746	.620	.834	.854	.765	.910
	Interorganizational Relationships	68	.641	.477	.762	.781	.646	.865
	Program Champion	69	.782	.671	.859	.878	.803	.924
Motivation	Compatibility/Alignment	68	.549	.432	.663	.830	.753	.887
	Complexity	69	.737	.639	.817	.893	.841	.931
	Observability	68	.630	.523	.730	.872	.814	.915
	Priority	69	.584	.406	.720	.738	.577	.837
	Relative Advantage	69	.528	.390	.654	.770	.658	.850
	Trialability	68	.310	.160	.466	.574	.364	.724

Note. Intraclass correlation coefficients calculated using a one-way random effects model.

Table 3.11 *Intraclass Correlation Coefficients and 95% Confidence Interval Estimates for Project 2: FP*

				95% CI for ICC(1)			*95% CI for ICC(2)*	
Scale	Subscale	N	ICC(1)	*Lower Bound*	*Upper Bound*	*ICC(2)*	*Lower Bound*	*Upper Bound*
General Capacity	Climate	62	.505	.405	.616	.891	.845	.928
	Culture	62	.684	.592	.771	.929	.897	.953
	Leadership	62	.675	.592	.759	.964	.950	.976
	Organizational Innovativeness	62	.691	.606	.775	.947	.925	.965
	Process Capacity	62	.653	.567	.741	.954	.935	.969
	Resource Utilization	62	.631	.502	.742	.837	.751	.896
	Staff Capacity	--	--	--	--	--	--	--
	Structure	62	.577	.473	.683	.891	.843	.928
Innovation-Specific Capacity	Implementation Climate Supports	62	.484	.368	.606	.824	.744	.885
	Innovation-Specific Knowledge & Skills	62	.767	.642	.853	.868	.782	.920
	Interorganizational Relationships	62	.543	.342	.697	.704	.510	.821
	Program Champion	62	.407	.252	.560	.673	.502	.792
Motivation	Compatibility/Alignment	62	.360	.230	.501	.692	.545	.800
	Complexity	62	.326	.169	.487	.592	.378	.740
	Observability	--	--	--	--	--	--	--
	Priority	62	.430	.275	.579	.693	.533	.805
	Relative Advantage	62	.350	.193	.510	.618	.418	.757
	Trialability	--	--	--	--	--	--	--

Note. Intraclass correlation coefficients (ICC) calculated using a one-way random effects model.

CHAPTER 4

DISCUSSION

Organizational readiness is an essential element for successful implementation of a particular innovation. The $R=MC^2$ heuristic and its associated measure, the Readiness Monitoring Tool, offer a compelling and comprehensive approach for organizations to identify specific areas of readiness which could benefit from capacity-building efforts to strengthen implementation supports. The RMT was designed to be adaptable to any innovation or setting, and is intended for use as a pre-implementation assessment, a periodic monitor of readiness during implementation, a predictor of innovation-related outcomes, and/or as an outcome measure for capacity-building interventions. Initial psychometric analysis has included establishing internal reliability, content validity, criterion validity, and preliminary factor structure. Further development of the RMT as a theoretically-informed and psychometrically-sound measure of organizational readiness requires additional psychometric testing, including the calculation and comparison of measures of interrater agreement and internal consistency across different applications of the tool.

Primary Findings

This study looked at the performance of adapted versions of the RMT in two of the 17 projects in which the RMT has been adapted for use. Research questions centered around two essential aims: 1) determining the stability of psychometric properties related to scale construction in a measure with adaptable items by computing and comparing

subscale alpha coefficients, and 2) providing preliminary evidence for the appropriateness of group-level aggregation of individual-level RMT responses to represent an organization's level of readiness.

The RMT was developed from a conceptualization of organizational readiness as a multidimensional construct, composed of three principal components: general capacity, innovation-specific capacity, and motivation. Each component contains multiple subcomponents sufficiently specific to permit targeted capacity-building efforts. The subscales of the RMT were designed to be unidimensional measures of each of these specified readiness subcomponents; as such, it is important that the subscales demonstrate sufficient internal consistency reliability to suggest each is measuring a single construct. Preliminary psychometric analysis of the RMT has found that the majority of subscales show good internal reliability (Wandersman & Scaccia, 2017). However, as a measure that encourages adaptation to fit each project's particular setting and innovation, this property isn't always guaranteed.

In support of the study's first aim, analyses showed that the majority of subscales in both projects demonstrated very good internal consistency. One subscale within Project 1: CHC (Trialability) and three subscales within Project 2: FP (Program Champion, Complexity, and Relative Advantage) fell below the acceptable lower bound for alpha set by Nunnally (1978), but not within the unacceptable range described by DeVellis (2012). Although these results are encouraging for the overall internal consistency of RMT subscales, variability in subscale alpha coefficients between projects suggests that RMT users who make significant adaptations to the measure (e.g., adding or

deleting multiple items) would do well to compute alpha coefficients to check that internal consistency has been maintained.

The second aim of this study focused on informing scoring procedures for the RMT, a survey measure administered to individuals within an organization to provide an overall assessment of the organization's readiness. Prior work has provided evidence that organizational readiness can be assessed via survey research methodology (Fox, Ellison, & Keith, 1988; Pond, Armenakis, & Green, 1984). However, Shea et al. (2014) note that three issues must be considered when attempting to measure readiness at supra-individual levels from individual respondents.

First, items within the measure should be written in such a way that they are group-referenced rather than self-referenced (e.g., "We are ready to…" rather than "I am ready to…") to orient individuals to the importance of providing scores that are based on perceived collective readiness of their organization rather than their personal readiness (Weiner, 2009). The RMT accomplishes this task by maintaining consistent use of "we" and "our organization/[site name]" in the measure's items, and by instructing participants to provide ratings that reflect their perception of how their organization overall would answer, rather than their own personal reaction to the item. Reiterating the importance of maintaining this item format for future RMT adapters should be included in any RMT adaptation protocol or administration guide.

Second, it is important that multiple respondents provide input on their organization's readiness. Proxy reporting from a single respondent, such as the organization's leader or a designated project liaison, is unlikely to generate accurate, representative data (Weiner, 2008). Strongly encouraging robust participation from each

site where the RMT is administered is necessary to provide an accurate picture of the organization's readiness as a whole, to determine within-organization variations in perception of readiness, and to generate sufficient power for statistical analysis. In many administrations of the RMT, researchers request that at last three individuals from each site complete the survey (Hartley, 2016). However, from the seven projects on which deidentified RMT data was initially collected, only 57% of sites had at least three participants, and just over one-third (34%) of surveyed sites had information from at least five participants, the minimum used by Shea et al. (2014) to determine appropriateness of individual-level response aggregation. Recommendations for researchers interested in calculating interrater reliability and interrater agreement indices when testing multilevel models – a natural future direction for the $R=MC^2$ body of research – call for ten or more participants per site to consistently see emergent group-level effects (LeBreton & Senter, 2008). Depending on the purpose of RMT administration and the statistical capacity of project staff, a higher per site participation ask may be warranted in future projects.

And, third, Weiner (2009) recommends that before averaging individuals' readiness perceptions into an organizational score, inter-rater agreement should be checked to ensure that the group-level aggregation is valid. When interrater agreement is severely lacking, an averaged group-level mean may fail to represent anyone's perception of that organization's readiness (e.g., if 50% of a site feel that they are woefully "not ready" and 50% feel that they are extremely ready to implement an innovation, the resulting group average of "moderately ready" accurately represents no one's view).

In this study, several measures of interrater agreement and interrater reliability were calculated, including r_{WG}, AD_M, *ICC(1),* and *ICC(2)* estimates. Although estimates

of interrater agreement calculated using r_{WG} and AD_M indices tend to be highly correlated (Burke et al., 1999), interpreting multiple indices can help researchers better understand the degree to which their data demonstrates interrater agreement (LeBreton & Senter, 2008) and can thus be validly aggregated into group-level means.

r_{WG} and AD_M indices provide estimates of interrater agreement. Estimates based on project samples indicated high interrater agreement in both Project 1: CHC and Project 2: FP. r_{WG} values for approximately 80% of the subscales in each project demonstrated strong agreement; the remaining 20% of subscales showed moderate agreement, which may be sufficient for justifying aggregation in a scale like the RMT (LeBreton & Senter, 2008). All subscales in each of the project samples demonstrated high agreement based on AD_M estimates. There was greater variability when estimates were examined between sites within a subscale and between subscales within a particular site. In each project, there was one site that demonstrated markedly lower levels of interrater agreement, with r_{WG} index values suggesting either weak or no agreement in over half of administered subscales (the majority of these subscales fell under the General Capacity scale).

ICC(1) estimates determine the extent to which a participant's ratings on the RMT is affected by their site. In both Project 1: CHC and Project 2: FP, *ICC(1)* values for RMT subscales reflected a "large" effect size of group membership on individual ratings. This suggests that the ratings given by individual participants are significantly influenced by the characteristics of their particular site.

ICC(2) estimates address the reliability of a site's mean rating. *ICC(2)* values for Project 1: CHC suggest an adequate level of reliability to justify aggregation for all

subscales but Trialability. In Project 2: FP, *ICC(2)* values supported aggregation for the majority of subscales in general and innovation-specific capacity. *ICC(2)* values for the Program Champion, Compatibility/Alignment, Complexity, Priority, and Relative Advantage subscales approached but did not meet the minimum recommended standard for justifying aggregation. As *ICC(2)* estimates are sensitive to the number of raters per group, it's possible that reliable mean differences between sites could have emerged for these subscales if the sample had included more respondents per site.

Collectively, these measures of interrater agreement and interrater reliability demonstrate that there is evidence to justify the aggregation of individual-level RMT responses into site-level subscale means.

Strengths of the Current Study

This study represents a necessary and important step in the development of the RMT as a psychometrically sound and useful measure of organizational readiness.

Naturally, much of the work completed thus far on the RMT has focused on looking at its performance within a single project. However, to establish that an adaptable measure performs consistently in different settings, it's necessary to compare performance across samples. By assessing internal consistency and interrater agreement in two samples, this study was able to draw conclusions about RMT performance that extended beyond a particular project or setting. Additionally, the initial effort to collect, clean, and compile individual-level responses from multiple projects that have used adapted versions of the RMT has resulted in the creation of a large database accessible for future RMT development work. As use of the tool spreads, developers may have an interest in monitoring and assessing the consistency of its performance across projects.

The establishment of this repository from seven projects is a significant first step in facilitating these efforts.

Group-level aggregation of individual RMT responses was previously guided by intuition, tradition, and a practical need for presenting digestible, actionable assessment results to members of an organization. This study is the first to offer statistical evidence that the organizational-level subcomponent scores derived from averaging individual responses are a reliable and valid representation of that site's level of readiness. This is an important and necessary step for justifying the aggregation of individual-level data into higher level scores. This evidence is more robust for having been derived from multiple indices of interrater agreement and interrater reliability, rather than relying upon a single measure or dichotomized cut-points for determining the degree of agreement in each sample.

Limitations and Future Directions

This study has a few limitations that suggest areas for future study. First, this study was limited by the comparatively low number of respondents per site from existing RMT data sets. Both *ICC(2)* and r_{WG} index estimates are affected by the number of raters contained within the sample. When sample size is small and agreement between raters is not especially high, estimates of interrater agreement will be attenuated (James, Demaree, & Wolf, 1984) and evidence of group-level differences may not emerge (LeBreton & Senter, 2008). To prevent attenuation of this nature, Lindell, Brandt, and Whitney (1999) suggest that ratings should be collected from 10 or more respondents per site. Within the seven projects for which RMT data was collected for this study, only 2% of sites – both of which were contained within Project 2: FP – met this sample standard.

Future studies may compare interrater agreement estimates calculated in this study to results from using multi-item versions of these indices. In the present study, at the advice of a statistical consultant, a subscale score was calculated for each respondent by averaging their responses to items within a particular subscale. This subscale score was then used to represent a single variable when calculating single-item r_{WG} and AD_M indices, assuming a uniform null distribution. Comparing results from using multi-item $r_{WG(J)}$ and $AD_{M(J)}$ indices and running $r_{WG(J)}$ indices using multiple null distributions (e.g., triangular, normal, slight/moderate/heavy skew) to account for the influence of different forms of response bias (as recommended by LeBreton & Senter, 2008) may provide more accurate estimates of interrater agreement in these samples.

An additional opportunity for future study would be to replicate tests of interrater reliability and interrater agreement using data from a laboratory study, similar to the methodology used by Shea et al. (2014) in their psychometric assessment of the ORIC. By creating profiles of organizations that systematically differ in their level of readiness, investigators can test whether the measure itself is able to reliably and validly differentiate these organizations. Data from the field, while high in external validity, can sometimes obscure strong levels of interrater agreement when there is limited between-unit variance (e.g., all organizations are rated at a similar level of readiness) (LeBreton & Senter, 2008). Although the large number of subscales contained within the RMT may complicate the vignette design for such a laboratory study (Shea et al. (2014) used a 2 (high- versus low-change commitment) x 2 (high- versus low-change efficacy) between-subjects design for a readiness measure with two distinct constructs), it's an approach worthy of consideration for testing measure properties in a more controlled setting.

The limitations of secondary data analysis from deidentified data were clearest in the inability to further investigate or contextualize inconsistencies in results between sites. In each project, one site stood out as having results that did not support the aggregation of individual-level data into an organizational-level mean. Collecting information about the efforts taken by organizational leaders and administrators of the RMT to orient participants to the innovation and to the concept of readiness could be useful in understanding why results from sites, and some subscales, did not support aggregation or demonstrated lower interrater agreement. It's possible that some individuals were provided with more information about the innovation than others, which would have resulted in different views on the organization's readiness and different scores on the RMT. Additionally, although participants were instructed to complete the measure independently, remote administration of the measure prevented project managers from confirming that this was the case in practice. If participants from a site completed the measure as a group, yet submitted individual responses, estimates of interrater agreement and interrater reliability would be artificially inflated.

Working from a deidentified data set without additional information to contextualize the respondents also prevented consideration of the source of within-site variation in patterns of response. r_{WG} estimates are based on the assumption that there is one true score for the object being rated. Although identified as an area in need of continued research, prior RMT work has investigated the influence of participant role (e.g., leadership/administration vs. service providers) on readiness ratings. Low or out-of-range r_{WG} estimates can result from multiple subgroups of participants each assigning their site a different "true" readiness score. If future work on identifying sources of

within-group response variability determines that participant factors such as role, prior exposure to the innovation, or duration of experience in the setting create distinct subgroups that influence RMT scores, then it may be more appropriate to estimate interrater agreement for multiple subgroups using the r_{WGp} index, which accounts for multiple true scores (LeBreton, James, & Lindell, 2005).

Lastly, continued work on refining the RMT's factor structure is recommended. Although the specificity with which the subscales identify relative strengths and weaknesses in an organization's readiness is incredibly useful for directing targeted capacity-building efforts, when subscales perform inconsistently across sites, the ability to make generalizable statements about expected psychometric performance of the RMT as a whole is limited.

Implications

Perhaps most importantly, this study has important practical implications for the administration and analysis of data from the Readiness Monitoring Tool concerning modification of the measure, minimum sample size, and interpretation of RMT scores.

The RMT is designed to be adaptable for use with a variety of settings and innovations. Projects using the RMT have reported varying levels of modification to the original measure. This study has shown that internal consistency of subscales can vary across RMT adaptations. Although some degree of project-specific item adaptation is necessary to include setting- or innovation-specific terminology, projects that decide to make more substantive modifications to subscale content (e.g., removing or adding multiple items to subscales) are advised to check the internal consistency of those subscales to ensure that they remain unidimensional.

Projects using the RMT have often emphasized the necessity of multiple respondents from the same organization to create an accurate assessment of the organization's collective readiness. In early projects, subscale scores were only computed for sites represented by at least three respondents (Hartley, 2016). However, this study has shown that estimates of interrater agreement and interrater reliability (necessary for justifying group-level aggregation and reliably demonstrating differences in group-level means) require an even larger per site sample. Although a balance must be struck between the quality and quantity of potential participants to ensure that respondents function as essentially parallel raters, future users of the RMT are encouraged to make a concerted effort to recruit at least five to ten participants per site. Creating standardized administration procedures -- including 1) recommendations on the amount of information given to sites about the innovation and measure before initial RMT administration, and 2) reaffirmation of the importance of participants responding individually based on their perception of the group's collective readiness -- should be a priority of measure developers before encouraging more widespread RMT use.

Finally, this study has shown that there is general statistical support for group-level aggregation of individual responses into site-specific subscale scores and readiness profiles. The majority of sites where the RMT was administered demonstrated sufficient interrater agreement and interrater reliability for these means to offer a valid representation of the organization's collective readiness. However, this agreement was not always guaranteed: one site in each project lacked agreement to the point where group-level means may have offered an inaccurate picture of the organization's perceived collective readiness for an innovation.

Practically speaking, community-based sites using the RMT may lack the statistical capacity to confirm adequate interrater agreement in their sample using the indices computed within this study. However, it is important that sites or the individuals supporting them through readiness assessments familiarize themselves with the data before computing organizational-level subscale mean scores. Checking more accessible measures of response variability, such as standard deviation and range, and taking the opportunity to openly discuss results where agreement seems lacking are practical steps for determining whether means accurately reflect the group perspective. Discussion guides for collectively interpreting RMT scores and using them to strategize priorities for capacity-building have been developed by early RMT users familiar with empowerment evaluation and the readiness heuristic (Hartley, 2016).

This study provides preliminary evidence that the RMT is able to reliably and validly represent an organizational-level construct from the aggregation of individual-level responses. Although more work is required to further refine the Readiness Monitoring Tool as a psychometrically-sound instrument, its potential for helping organizations successfully implement innovations through readiness assessment, monitoring, and targeted capacity-building support is significant and worthy of continued effort.

REFERENCES

Aarons, G. A., Ehrhart, M. G., & Farahnak, L. R. (2014). The implementation leadership scale (ILS): development of a brief measure of unit level implementation leadership. *Implementation Science, 9*(1), 45.

Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research, 38*, 4–23.

Armenakis, A. A., Harris, S. G., & Mossholder, K. W. (1993). Creating readiness for organizational change. *Human relations, 46*(6), 681-703.

Backer, T. E. (1995). Assessing and enhancing readiness for change: Implications for technology transfer. NIDA research monograph, 155, 21-41.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349-381). San Francisco, CA, US: Jossey-Bass.

Burke, M. J., & Dunlap, W. P. (2002). Estimating interrater agreement with the average deviation index: A user's guide. *Organizational research methods, 5*(2), 159-172.

Burke, M. J., Finkelstein, L. M., & Dusig, M. S. (1999). On average deviation indices for estimating interrater agreement. *Organizational Research Methods, 2*(1), 49-68.

Choi, M., & Ruona, W. E. (2011). Individual readiness for organizational change and its implications for human resource and organization development. *Human Resource Development Review, 10*(1), 46-73.

D'Amico, J. J., & Corbett, H. D. (1988). Developing Readiness for School Improvement: Lessons of Context. Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 5-9, 1988).

Davis, H. R. (1978). Management of innovation and change in mental health services. *Psychiatric Services, 29*(10), 649-658.

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

Drzensky, F., Egold, N., & Van Dick, R. (2012). Ready for a change? A longitudinal study of antecedents, consequences, and contingencies of readiness for change. *Journal of Change Management, 12*(1), 95–111.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American journal of community psychology, 41*(3-4), 327-350.

Ehrhart, M. G., Aarons, G. A., & Farahnak, L. R. (2014). Assessing the organizational context for EBP implementation: the development and validity testing of the Implementation Climate Scale (ICS). *Implementation Science, 9*(1), 157.

Fetterman, D. & Wandersman, A. (Eds.). (2005). *Empowerment evaluation: Principles in practice.* New York, NY: Guilford Press.

Flaspohler, P., Duffy, J., Wandersman, A., Stillman, L., & Maras, M. A. (2008). Unpacking prevention capacity: An intersection of research-to-practice models

and community-centered models. *American journal of community psychology, 41*(3-4), 182-196.

Fox, D. G., Ellison, R. L., & Keith, K. L. (1988). Human resource management: An index and its relationship to readiness for change. *Personnel Administration, 17*(3), 297-302.

Greenhalgh, T., Robert, G., MacFarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: Systematic review and recommendations. *The Milbank Quarterly, 82*(4), 581–629.

Hall, G., & Hord, S. (2015). *Implementing change: Patterns, principles, and potholes* (4th ed.). Boston: Pearson.

Hartley, S.N. (2016). R=MC2 projects follow-up: Highlights from summer 2016 interviews. Prepared for the Institute for Healthcare Improvement.

Hoagwood, K., Schoenwald, S. K., & Chapman, J. E. (2003). Dimensions of organizational readiness—revised (DOOR-R). Unpublished instrument.

Holt, D. T., & Vardaman, J. M. (2013). Toward comprehensive understanding of readiness for change: The case for an expanded conceptualization. *Journal of Change Management, 13*(1), 9-18.

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of applied psychology, 69*(1), 85.

James, L. R., Demaree, R. G., & Wolf, G. (1993). rwg: An assessment of within-group interrater agreement. *Journal of applied psychology, 78*(2), 306.

Jansen, K. J. (2000). The emerging dynamics of change: Resistance, readiness, and momentum. *People and Strategy, 23*(2), 53.

Klein, K. J., & Kozlowski, S. W. (2000). From micro to meso: Critical steps in conceptualizing and conducting multilevel research. *Organizational research methods, 3*(3), 211-236.

Klein, K. J., Dansereau, F., & Hall, R. J. (1994). Levels issues in theory development, data collection, and analysis. *Academy of Management review*, *19*(2), 195-229.

Kotter, J. P. (1996). *Leading change*. Harvard business press.

LeBreton, J. M., Burgess, J. R., Kaiser, R. B., Atchley, E. K., & James, L. R. (2003). The restriction of variance hypothesis and interrater reliability and agreement: Are ratings from multiple sources really dissimilar?. *Organizational Research Methods, 6*(1), 80-128.

LeBreton, J. M., James, L. R., & Lindell, M. K. (2005). Recent issues regarding rWG, rWG, rWG (J), and rWG (J). *Organizational Research Methods*, *8*(1), 128-138.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational research methods, 11*(4), 815-852.

Lehman, W. E., Greener, J. M., & Simpson, D. D. (2002). Assessing organizational readiness for change. *Journal of Substance Abuse Treatment, 22*(4), 197-209.

Lindell, M. K., Brandt, C. J., & Whitney, D. J. (1999). A revised index of interrater agreement for multi-item ratings of a single target. *Applied Psychological Measurement, 23*(2), 127-135.

Lynne, D., Finnerty, M., & Boyle, P. (2005). Implementing EBPs: Early results and the use of the General Organizational Index and other tools in implementing EBPs. Paper presented at the Fifteenth Annual NRI Conference on State Mental Health Agency Services Research, Program Evaluation, and Policy. Baltimore, MD.

Murphy, K. R., & Myors, B. (2004). *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests* (2nd ed.). Mahwah, NJ: Erlbaum.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Pasmore, W. A., & Fagans, M. R. (1992). Participation, individual development, and organizational change: A review and synthesis. *Journal of Management, 18*(2), 375-397.

Patterson, M., et al. (2005). Validating the organizational climate measure: links to managerial practices, productivity and innovation. *Journal of Organizational Behavior, 26*, 379-408. Retrieved from https://wiki.uio.no/admin/amunder/images/a/a0/OCM-1-.pdf

Pond III, S. B., Armenakis, A. A., & Green, S. B. (1984). The importance of employee expectations in organizational diagnosis. *The Journal of applied behavioral science, 20*(2), 167-180.

Powell, B. J., McMillen, J. C., Proctor, E. K., Carpenter, C. R., Griffey, R. T., Bunger, A. C., . . . York, J. L. (2012). A compilation of strategies for implementing clinical innovation in health and mental health. *Medical Care Research and Review, 69*(2), 123–157.

Robert Wood Johnson Foundation. (2015, April 27). Communities Receive Funding to Accelerate and Deepen Efforts to Improve Residents' Health. Retrieved from http://www.rwjf.org/en/library/articles-and-news/2015/04/communities-receive-funding-to-accelerate-and-deepen-efforts-to-improve-residents-health.html

Rogers, E. M. (2003). *Diffusion of Innovations* (5th ed.). New York: Free Press.

Scaccia, J. P., Cook, B. S., Lamont, A., Wandersman, A., Castellow, J., Katz, J., & Beidas, R. S. (2015). A practical implementation science heuristic for organizational readiness: R= MC2. *Journal of Community Psychology, 43*(4), 484-501.

Schriesheim, C. A., Powers, K. J., Scandura, T. A., Gardiner, C. C., & Lankau, M. J. (1993). Improving construct measurement in management research: Comments and a quantitative approach for assessing the theoretical content adequacy of paper-and-pencil survey-type instruments. *Journal of Management, 19*(2), 385-417.

Scott, V.C., Kenworthy, T., Godly-Reynolds, E., Bastien, G., Scaccia, J.P., McMickens, C., Rachel, S., Cooper, S., Wrenn, G., Wandersman, A. (2017). The Readiness for Integrated Care Questionnaire (RIC-Q): A New Tool to Assess Readiness to Integrate Behavioral Health and Primary Care. *American Journal of Orthopsychiatry*. doi: 10.1037/ort0000270

Shea, C. M., Jacobs, S. R., Esserman, D. A., Bruce, K., & Weiner, B. J. (2014). Organizational readiness for implementing change: a psychometric assessment of a new measure. *Implementation Science, 9*(1), 7.

Simpson, D. D. (2002). A conceptual framework for transferring research to practice. *Journal of Substance Abuse Treatment, 22*, 171–182.

Wandersman, A. & Scaccia (2017). Organizational readiness: Measurement and as a predictor of progress: Final report. Prepared for The Robert Wood Johnson Foundation.

Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L., ... & Saul, J. (2008). Bridging the gap between prevention research and practice: The interactive systems framework for dissemination and implementation. *American journal of community psychology, 41*(3-4), 171-181.

Weiner, B. J. (2009). A theory of organizational readiness to change. *Implementation Science, 4*(67).

Weiner, B. J., Amick, H. & Lee, S.-Y. D. (2008) Conceptualization and measurement of organizational readiness for change: a review of the literature in health services research and other fields. *Med. Care Res. Rev. 65*, 379–436.

Weiner, B. J., Belden, C. M., Bergmire, D. M., & Johnston, M. (2011). The meaning and measurement of implementation climate. *Implementation Science, 6*(1), 1-12.