Evidence-based practices in community collaborative work: Implications of Relational Developmental Systems Concepts

A dissertation

submitted by

Jennifer Elise Prescott

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in

Child Study and Human Development

TUFTS UNIVERSITY

August 2015

COMMITTEE:

Richard M. Lerner, Ph.D. (chair)

Jonathan F. Zaff, Ph.D.

Sara K. Johnson, Ph.D.

Aaron Metzger, Ph.D.

Abstract

A youth system is a concept useful for discussing the alignment of contextual assets and individual strengths and needs. Such systems are adaptive when they support the positive, healthy development of both the person and the context. Comprehensive community initiatives (CCIs) have demonstrated efficacy at moving youth systems toward being more adaptively supportive by, at least in part, involving collaborations using evidence-based programs (EBPs). An integral part of the evidence base is examining how the collaboration itself is functioning. However, to date, most measures of collaboration functioning that have been validated are intended for use in a single collaborative approach, and those designed for use across different types of collaborative approaches have rarely been validated. In this dissertation, I examined measures of collaboration functioning – structure and process – to see if they performed similarly across two collaborative approaches. I found that items pertaining to collaborative structure loaded onto three factors in almost identical ways across collaborative approaches. The items in these three factors also behaved similarly across collaborative approaches. However, the items pertaining to collaborative process did not perform similarly across collaborative approaches. Future research should develop process items that maintain similar factor structure and behave similarly across collaborative approaches.

Acknowledgements

I am so incredibly grateful for the support of so many kind, generous, and wonderful people in my life. I cannot thank you all enough, but I will begin to try here.

First, I have to thank my absolutely amazing committee, Drs. Richard M. Lerner, Jonathan F. Zaff, Sara K. Johnson, and Aaron Metzger. Rich, thank you so much for your guidance and support throughout this process. I cannot express enough my appreciation for taking me on as your student. Your feedback and insight has proved invaluable, with this dissertation and beyond. I will carry the lessons you have taught me for years to come.

Jon, it has meant the world to me that you allowed me to work with you over the last four years. I doubt there is a doctoral program anywhere else that would have exposed me to the same types of experiences that I gained working with you, and I am truly grateful. Your mentorship and understanding were the bedrock of my doctoral experience.

Sara, thank you for being one of the most patient methodologists I have ever had the pleasure of working with. Your willingness to make time for me repeatedly and ensure that I fully understood all of the possible angles of my dissertation smoothed out what could have been a very rough process. Aaron, thank you so much for your thoughtful feedback. It was wonderful to hear your perspective and it provided important additions to this work.

A very sincere thank you to Dr. Sunho Jung, who provided the MATLAB code used in this dissertation for the regularized exploratory factor analysis. His expertise and willingness to help a complete stranger made parts of this dissertation possible, and for that I am incredibly grateful. I would also like to thank Dr. Francine Jacobs, for providing support and guidance throughout the first few years of my doctoral experience.

The remainder of my gratitude belongs to my friends and family, without whom this whole process would not have been possible. To my family, thank you. Thank you for helping me to become who I am today, for supporting me and loving me unconditionally, and being people that I like and also love. You are really great, all of you, but in particular I would like to thank Mom, Dad, Jeff, Alex, John, Linda, Lauren, Daniel, Nathan, Granddad, Fred, Rachel, Steven, Mark, Maris, Heather, Bryan, Granddad, Aunt Ruth, Uncle Nick, Jay, Laura, and Pam (there are a lot of you – and I could not have done it without every single one).

To my grad school family, I love you, and thank you for being in this with me. In particular, I would like to thank ERK, MMC, ESL, MCE, LMD, EAS, AMCP, MBW, JPA, ASG, ERM, SDRL, KA, AB, EPJ, AED, KLSC, FDC, TOB, JP, SEAA, LC, ZMS, and MCL. I owe a very large thank you to Brandon Foster for walking me through Rasch analyses. To my friends outside grad school, thank you for grounding me and being my anchors throughout this journey. I would like to thank MEJ, ARP, CJT, CB, MAOA, SDW, DGS, LSD, BM, DM, SEH, EL, BR, JH, GMW, and AB. Last and certainly least, I would like to thank Top Gear seasons 1 through 21 for providing the perfect break from writing.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: INTRODUCTION

Relational developmental systems (RDS) metatheories posit that development consists of mutually influential relationships between the individual and the context (Overton, 2013; 2015). When the strengths and needs of the young person are aligned with the assets in the context, adaptive development will occur for the youth and for his or her surrounding community (Lerner & Overton, 2008; Overton, 2015). However, youth in a variety of communities are not experiencing the developmental supports that they need and are subsequently less likely to be on positive developmental trajectories (e.g., Kim, 2012; Luthar & Barkin, 2012; Luthar, Barkin, & Crossman, 2013; Perna & Titus, 2005; Swanson, 2009). Youth-focused collaborations may provide a useful way to strengthen the connections among the assets in community contexts (e.g., school, after-school programs, health, public works) and move entire communities toward becoming more supportive of youth (e.g., Jenson, Alter, Nicotera, Anthony, & Forrest-Bank, 2013; Kubisch, Auspos, Brown, & Dewar, 2010). However, it is important to ensure that collaborations are working toward efficacy – or ensuring that they are impacting outcomes in the intended manner (e.g., Keast & Mandell, 2012). One of the important hallmarks of effective collaboration is the use of evidence-based practices and data-driven decision-making.

There are a variety of measures that address collaborative efficacy, including measures of collaboration functioning. Indeed, how the collaboration itself functions is predictive of improved youth and family outcomes (Feinberg, Bontempo, & Greenberg, 2008). Measures of collaboration functioning include measuring factors among collaboration members such as clarity of roles for collaboration members, trust among collaboration members, shared accountability, shared vision and mission, and collective efficacy (e.g., Fawcett, Francisco, Paine-Andrews, & Schultz, 2000; Goldsmith & Eggers, 2004; Lasker & Weiss, 2003; Lawson, Claiborne, Hardiman, Austin, & Surko, 2007; Roussos & Fawcett, 2000; Seldon, Jolin, & Schmitz, 2012).

However, measures that have been examined have traditionally been assessed for their psychometric characteristics within the context of specific collaborative approaches (e.g., Brown, Feinberg, & Greenberg, 2012; Kegler & Swan, 2011). Many of these measures have not been validated for use in collaborations that are not adhering to those specific approaches. Furthermore, most other measures that were developed for use across collaborative approaches were either not validated or the structures of the measures were not statistically tested (e.g., Bush, Dower, & Mutch, 2002; Kaye, 1993).

If few of the measures in use for collaborative functioning are validated, and those that are validated are not validated for use with other collaborative approaches (e.g., Brown, Feinberg, & Greenberg, 2012; Kaye, 1993; Kegler & Swan, 2011), then it will be useful to examine measures of collaborative functioning across collaborative approaches. Accordingly, the question I addressed in this research was: Can measures of collaborative functioning be used across different perceived collaborative approaches to improving youth developmental outcomes? To address this question, I conducted two types of analyses: a specialized form of an exploratory factor analysis to examine the underlying factor structure, and a Rasch analysis to examine how the questions performed across collaborative approaches.

The Analyses

Analysis 1 – Examining the underlying factor structure of the measures across collaborative approaches

The first analysis involved using a form of exploratory factor analysis (EFA) designed for small samples called regularized exploratory factor analysis (REFA). The EFA is designed to uncover the underlying structure of and relationships among the measured variables, as they emerge in collaborations focused on youth development outcomes (Brown, 2014).

Analysis 2 – Invariance of measures of collaborative functioning across collaborative approaches

The second analysis involved using Rasch analysis to test whether the measures that have been used for assessing collaborative functioning were invariant across multiple collaborative approaches. The Rasch analysis combines item response theory and validity testing and ascertains invariance of a measurement by items across groups. This analysis tested the invariance of the behavior of the items across the two perceived collaborative approaches.

This dissertation is organized into four chapters. In this first chapter, I present a literature review and an RDS-based theoretical model for the connection between youth development trajectories and the efficacy of youth-focused collaborations. I then focus empirically on one component of this theoretical model, collaborative functioning. Then I describe measures of collaborative functioning that have been used previously. In the following chapter, I present the method used for both sets of analyses. In the penultimate chapter, I present the results and brief discussion of the two analyses described above. In the final chapter, I will present the implications of this research for the collaboration literature, suggest directions for future research, and discuss practical applications.

An Introduction to Youth Systems

An individual does not develop within a vacuum; there are many layers of the ecology in which a person develops, with the individual at the center of the system (Bronfenbrenner, 1979; Bronfenbrenner & Morris, 2006). That is, there are many contexts that influence the development of an individual. Importantly, the person simultaneously influences his or her development by taking actions and making decisions in an attempt to regulate adaptively the interactions between himself or herself and the environment (e.g., Brandtstädter, 1998, 2006; Spencer, 2006).

Theories derived from the relational developmental systems (RDS) metatheories posit that development consists of mutually influential relationships between the individual and the context (Overton, 2013, 2015). In other words, a young person influences and is influenced by the many contexts in which she is developing (e.g., family, school, programs, the larger community, and culture; Lerner, 2012; Lerner & Overton, 2008). Thus, adaptive developmental regulations involve relations between agentic individuals and their contexts that result in positive youth development outcomes. That is, when the strengths and needs of the young person are aligned with the assets and resources in the context, adaptive development will occur.

There may be many individually cognized or expressed facets of the ecology that are relevant to adaptive developmental regulations (e.g., individual perception or stress; Spencer, 2006). When a young person is at the center of this system, one may use an heuristic termed the *youth system* (Zaff, 2011). A *youth system* is the interaction between a given young person and other individuals, organizations, physical settings, and larger societal norms and practices. This heuristic is an applied representation of Bronfenbrenner's (1979; Bronfenbrenner & Morris, 2006) bioecological systems model, action theoretical perspectives, and RDS-based theories more generally (Bronfenbrenner & Morris, 2006; Brandtstädter, 2006; Lerner, 2012; Overton, 2015; Spencer, 2006). The *youth system* explicitly places an individual young person at the center of the model in applied settings, rather than as an individual in a theoretical setting.

A *youth system*, in and of itself, has no valence. The *youth system* can be supportive or unsupportive, or fall anywhere between these extremes. Indeed, there is evidence that, in a variety of communities, *youth systems* are not currently supportive of young people (e.g., Kim, 2012; Luthar & Barkin, 2012; Luthar, Barkin, & Crossman, 2013; Perna & Titus, 2005; Swanson, 2009). However, when the resources in a community are aligned with the needs and strengths of the young people in the community (an alignment that should lead to adaptive developmental regulations; Brandtstädter, 1998), then the *youth system* could be a considered a supportive (or adaptive) youth system (Zaff, Donlan, Pufall Jones, & Lin, in press).

There is evidence to suggest that comprehensive community initiatives (CCIs) are an effective way to move entire community systems from being less supportive toward becoming more supportive of youth (Jenson et al., 2013; Kubisch et al., 2010). CCIs are designed with the intent to help align resources across contexts by taking a comprehensive approach to community change, involving residents in meaningful ways, and intentionally building community capacity to create and maintain positive change (Kubisch et al., 2010). Many CCIs have demonstrated impacts on individual participants (Kubisch et al., 2010), and some CCIs have demonstrated impacts at the community level (Brown et al., 2009; Hawkins Catalano, & Arthur, 2002; Spoth, Greenberg, Bierman, & Redmond, 2004).

One organizational instantiation of a CCI is a community collaboration, or a voluntary alliance of organizations from a variety of sectors (public, private, and/or non-profit) that is designed to build capacity to work toward common goals through collectively pooling resources, sharing responsibilities, and distributing both risks and rewards (Himmelman, 1992; Jenson et al., 2013; Keast

& Mandell, 2012; Kubisch et al., 2010). Because of the rising prominence of collaborative work (e.g., Promise Neighborhoods Initiative; United States Department of Education, 2013), I focus on community collaboration as a model of CCIs that may create and sustain transformative and positive changes in communities.

There is evidence that community collaborations can strengthen the connections among contexts and can have impacts on youth level outcomes (e.g., Brown et al., 2009; Hawkins et al., 2002; Jenson et al., 2013; Spoth et al., 2004). Collaborations may be most successful when they implement evidence-based programs (EBPs; e.g., Brown, Hawkins, Arthur, Briney, & Abbott, 2007; Feinberg, Jones, Greenberg, Osgood, & Bontempo, 2010; Hawkins, Catalano, & Arthur, 2002; Spoth et al., 2004). Indeed, the use of EBPs implemented with fidelity has been demonstrated to be integral to the success of several collaborative approaches (e.g., Brown et al., 2007; Feinberg et al., 2010: Hawkins, Catalano, & Arthur, 2002; Spoth et al., 2004).

An important part of the use of evidence to inform practices includes using evidence to monitor how the collaboration itself is functioning. There are many measures that have been used across a variety of collaborative structures and locations (e.g., Brown, Feinberg, & Greenberg, 2012; Kegler & Swan, 2011). However, these measures have not been validated for use across multiple collaborative approaches, and instead have traditionally only been used within a single collaborative approach (e.g., Brown, Feinberg, & Greenberg, 2012; Kegler

& Swan, 2011). In addition, little, if any work has been done uncovering the underlying structure of and relationship among the measures across multiple collaborative approaches.

In this chapter, I will first describe RDS-based theories and *youth systems*. Then, I will describe how CCIs and community collaborations, in particular, can help communities move toward more supportive *youth systems*. Next, I will describe how some collaborations have demonstrated positive impacts using EBPs. I will then describe how measures have been used to examine collaborative functioning thus far in the field. Finally, I will argue for the need for measures of collaborative functioning which examine similar constructs across different types of collaborative approaches.

Relational Developmental Systems Theories

Theories derived from the relational developmental systems (RDS) metatheories define development as a result of the mutually influential relationship between the person and context. Therefore, such models are an ideal frame through which to consider comprehensive community initiatives, which consider both the person and context in an integrative manner. In particular, RDS-based theories argue that the person cannot be separated from the context, and that the person both influences and is influenced by the contexts in which he or she develops (Lerner, 2012; Overton, 2015). One instance of RDS theories are action-theoretical perspectives (Lerner, 2012; Overton, 2013). According to action-theoretical perspectives, individuals simultaneously actively produce and are the product of their environments (e.g., Brandtstädter, 1998, 2006). That is, throughout their lives, individuals take actions and make decisions that direct and influence their own ontogeny (e.g., Brandtstädter, 1998, 2006; Lerner, 2012).

These actions and decisions are goal-directed; they have the aim of improving each individual's development (Brandtstädter, 1998, 2006). As individuals coact with their environments and attempt to optimize their own developmental outcomes, they are attempting to regulate the relation between themselves and their environments in an adaptive way. Thus, in action-theoretical perspectives, and in RDS theories more generally, adaptive developmental regulations are the relation between the agentic individual and her/his surrounding context that result in positive developmental outcomes for both the individual and the context (Brandtstädter, 1998, 2006; Lerner, 2012).

According to other RDS-based models, such as bioecological theory, the young person is embedded within multiple layers of his or her ecology (Bronfenbrenner & Morris, 2006) and takes actions to align his/her own strengths with the contextual assets within that ecology (Brandtstädter, 2006). Within RDS theories, the stress on relative plasticity across the life span (Lerner, 1984, 2002, 2012) means that all youth possess the capacity to systematically change and, as a consequence, have strengths that help them navigate and negotiate their world (Benson, Leffert, Scales, & Blyth, 2012). For example, a young person can be particularly tenacious in pursuing goals (e.g., Brandtstädter, 2006; Duckworth, Peterson, Matthews, & Kelly, 2007; Heckhausen, Wrosch, & Schulz, 2010), or can be adept at using and relying on his/her social networks to help her as she moves through her life (e.g., Antonucci, Akiyama, & Takahashi, 2004; Syed, Azmitia, & Cooper, 2011). High levels of these internal strengths have been tied to higher academic achievement, academic attainment, and employment (e.g., Duckworth et al., 2007; Gutman & Schoon, 2013; Heckman, Stixrud, & Urzua, 2006; Peck, Roeser, Zarrett, & Eccles, 2008).

Contextual assets include people, physical spaces, programs, or communities that promote positive outcomes (Bronfenbrenner & Morris, 2006; Scales, Benson, & Mannes, 2006; Theokas et al., 2005). Some examples of contextual assets include supportive parents (e.g., Jimerson, Egeland, Sroufe, & Carlson, 2000; Kim, 2012; Perna & Titus, 2005; Swanson, 2009), a neighborhood with low rates of violence (e.g., Leventhal, Dupéré, & Brooks-Gunn, 2009), or a community with collectively higher levels of social capital and cohesion (e.g., Lochner, Kawachi, Brennan, & Buka, 2003; Sampson, Raudenbush, & Earls, 1997; Scales, Benson, & Mannes, 2006). Both youth and their contexts may evidence change – an individual can improve on his or her internal strengths, and assets in a context can improve in quality and function, as well (Benson et al., 2012; Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Hamre & Pianta, 2005; Spencer, 2013; Spencer & Swanson, 2013). The inverse is also true – an individual may not adapt his or her strengths to new life circumstances, and a community can lose assets and deteriorate (e.g., Hollander, Pallagst, Schwarz, & Popper, 2009).

Youth Systems

To further operationalize the idea of the person as embedded within his or her context specifically for initiatives focused on youth, I return to the concept of the *supportive youth system* to express the key developmental supports in a community that young people need to thrive and the attributes that young people themselves possess and use to navigate through their lives (e.g., grit, self-efficacy; Damon, 1997; Duckworth et al., 2007; Zaff, 2011; Zaff et al., in press; Zaff & Smerdon, 2009). Across all types of communities, *supportive youth systems* may be unlikely to develop without intervention. On the one hand, there is strong evidence indicating that youth exposed to higher levels of poverty are more likely to have health challenges, behavioral problems, and poor academic performance (e.g., Kim, 2012; Perna & Titus, 2005; Swanson, 2009). On the other hand, these problems are not limited to youth in poor communities. Youth from affluent communities also struggle with emotional and behavioral issues, such as drug and alcohol use, depressive symptoms, and/or acting out (Luthar & Barkin, 2012; Luthar, Barkin, & Crossman, 2013). Therefore, researchers must look for diverse avenues to encourage positive youth development in all community settings.

Whereas diverse types of communities struggle with encouraging adaptive development in youth, low-income communities face additional barriers not often found in middle- or high-income communities (Carter & Welner, 2013; Houston & Ong, 2013). For example, low-income communities are often historically disenfranchised and struggle with a lack of resources (e.g., Crowder & South, 2003; Duncan & Murnane, 2011; Vlahov et al., 2007). Low-income communities often struggle as well with higher rates of violence and gang activity (e.g., Howell & Egley, 2005; Morenoff, Sampson, & Raudenbush, 2001; Wilson, 1987). As such, low-income communities often face substantially worse youth outcomes than middle- or high-income communities (e.g., Carter & Welner, 2013; Duncan & Murnane, 2011). Therefore, I focused on the improvement of *youth systems* in low-income communities. In addition, there are a variety of current initiatives underway that focus on low-income communities (e.g., Promise Neighborhoods Initiative; United States Department of Education, 2013) that provide insight into how to promote supportive *youth systems*.

Person-Context Alignment and Youth Outcomes

There exist individual programs and initiatives that are able to shift the youth system toward becoming more supportive without explicitly focusing on the embedded nature of the youth in their context. Individual programs that demonstrate significant and substantive effects do not necessarily explicitly incorporate alignment across contexts as a core component. For example, some out-of-school time (OST) programs demonstrate efficacy by targeting only individual characteristics (e.g., Durlak, Weissberg, & Pachan, 2010; Lauer, Akiba, Wilkerson, Apthorp, Snow, & Martin-Glenn, 2006). RDS-based models suggest that these programs may be able to have a more significant impact on outcomes when the intervention targets alignment with multiple facets of the context as well as the individual (Benson et al., 2012; Lerner, 2012; Overton, 2013, 2015).

Indeed, there is evidence that alignment in the relationship between individual and context results in positive youth development (PYD; e.g., Lerner, Lerner, Bowers, & Geldhof, 2015). For example, a component of many successful youth programs is the integration of services across contexts (e.g., schools, family, or neighborhood; e.g., Blum, 2003; Catalano, Berglund, Ryan, Lonczak, & Hawkins, 2004; Dryfoos, 1994; Eccles & Gootman, 2002; McKnight & Mretzmann, 1993). It may be that these programs with more alignment may have better results than programs that do not explicitly align across contexts. One example of an alignment that has been empirically examined is the one between in-school and OST activities. For example, The AfterSchool Corporation (TASC) completed an evaluation that demonstrated improved student outcomes when afterschool programs were more aligned with the school context, through methods such as employing teachers in afterschool programs and using the same buildings as the schools (Reisner et al., 2004).

However, this alignment between contexts does not occur naturally with great frequency (e.g., Kim, 2012; Luthar & Barkin, 2012; Luthar, Barkin, & Crossman, 2013; Perna & Titus, 2005; Swanson, 2009). Youth and their families often must piece together disjointed resources for themselves because many programs are focused on discrete, individual needs rather than on providing holistic assistance (Silverstein, Lamberto, DePeau, & Grossman, 2008). Indeed, there is evidence that, although a variety of resources may be available in a community (e.g., health, afterschool programs, tutoring, etc.), families in low-income communities in particular find it difficult to navigate the disparate services (Dryfoos, 1994; McKnight & Kretzmann, 1993; Silverstein et al., 2008). This situation raises the question of how to foster the development of supportive youth systems to align youth's individual strengths with the external resources around them.

Initiatives to build connections across individual contexts are helpful (e.g., family engagement programs in schools, connecting school and family), and these individual connections between contexts may have enough impact to change the larger *youth system* (e.g., a family engagement program in a school may get a critical mass of participants to change larger cultural attitudes in the community around parental participation in school). However, these individual initiatives are not specifically targeted at changing the way the *youth system* functions on a wider scale. Rather, in order to intentionally shift the *youth system* from a less supportive system to a more supportive one, strategic efforts to align youth strengths and needs with resources across contexts may be more effective. Comprehensive community initiatives are one such type of strategic effort and offer a potential pathway to creating conditions assuring that each young person in a community experiences the key developmental supports that she or he needs to thrive academically, socially, emotionally, vocationally, and civically.

Comprehensive Community Initiatives (CCIs)

CCIs take a comprehensive approach to community change, and incorporate engagement with residents as well as community capacity building as intentional parts of the change process (Kubisch et al., 2010). CCIs are designed with the intention to help to align supports in the community with the needs and strengths of the members of the community. There are several different ways to structure CCIs, including single organizations targeting multiple sectors (e.g., Harlem Children's Zone (HCZ); Harlem Children's Zone, 2012), as well as formal community collaborations. Single organizations like the HCZ may run multiple programs operating in several sectors. For example, HCZ operates charter schools, parent education programs, community centers, and several health programs (Harlem Children's Zone, 2012). A formal community collaboration would involve distinct, existing organizations from multiple sectors coming together to work toward a common goal. For example, a school district, non-profits, a health center, and a community center may work toward the goal of increasing the high school graduation rate in the community in which they operate (Kubisch et al., 2010).

To date, there is no evidence which structure of CCI works "best." Consistent with RDS-based models, which emphasize diversity across time and place, it is likely that no single CCI structure will be the "best" model for intervening to move systems towards becoming more supportive. Rather, it is likely that different formats and structures will work for different communities tackling different problems.

The use of formal collaboration as an instantiation of CCI has increased from the mid-1990s to the 2010s (e.g., Jenson et al., 2013; Kubisch et al., 2010; Wolff, 2001). The federal government has prioritized funding for formal collaborative efforts in a variety of domains (e.g., the Promise Neighborhoods Initiative; United States Department of Education, 2013). As a result, I focused on formal community collaborative efforts as a CCI vehicle through which communities can attempt to align individual strengths and contextual resources.

Formal Community Collaboration

To date, there has been some confusion in defining collaboration. There is a difference between the verb, to collaborate, and the noun describing a formal group, a collaboration. "To collaborate" is the process by which parties who have a stake in a particular problem come together to solve the problem in a mutually agreeable way (Gray, 1989). A collaboration is a group of organizations, and may be defined as a "voluntary, strategic alliance of public, private and [/or] non-profit organizations to enhance each other's capacity to achieve a common purpose by sharing risks, responsibilities, resources and rewards" (Himmelman, 1992, p. 3). Here, I focused on the noun, where collaborations have formalized structures and processes, contain member organizations, and are usually guided by a single leader or a governing board that makes decisions about the collaboration's activities.

Consistent with a view of the active, developing child as embodied within her or his context (Overton, 2013, 2015), these interorganizational efforts strive to build connections across the contextual elements in the lives of youth and encompass a wide range of individuals and sectors that support the complex developmental needs and strengths of young people (e.g., youth development programs, schools, health care providers, government agencies, and youth, their parents, and other residents of the community; Hawkins, Catalano, & Arthur, 2002; Keast & Mandell, 2012; Jenson, et al., 2013). It is possible to have multiple sectors working together in ways that are not collaborative as I have defined this term, but instead act as coordinating efforts.

For example, HCZ is a multi-service organization that provides numerous services (e.g., social service programs, education, community health, early childhood care, and afterschool enrichment programs) and offers supports throughout the first two decades of life, starting prenatally and continuing into college. Although it is often held up as the model informing collaborative work, HCZ is a CCI structured as a single organization. In this dissertation, I focused only on collaborative efforts that involved three or more separate organizations working together to solve problems in their communities. The literature usually focuses on collaborations involving two or more organizations or sectors working together (e.g., Kubisch et al., 2010). However, because this relationship could be considered a partnership, I focused on three or more organizations. These tripartite collaborations may have a single lead organization, but decisions and goals are arrived at collectively, rather than dictated by a single person or organization.

There are multiple approaches to collaboration across communities (Kubisch et al, 2010). Some communities have a consensus-building approach, where the collaboration engages with a broad constituency and decision-making is shared among the participating organizations. Other collaborations take an intermediary approach, where there is a lead organization, with a traditional board of directors or trustees who drive the strategy. Collaborative partners are then brought in to the intermediary approach based on the agreed upon strategy. Another collaborative approach is the centralized approach, which has one organization or individual that takes the lead and possesses most, if not all, of the decision-making responsibilities.

Each of these models of formal collaborations has the potential to create change in the communities in which they work (Foster-Fishman, Berkowitz, Lounsbury, Jacobson, & Allen, 2001; Zakocs & Edwards, 2006). However, not all collaborations are effective (Hawkins, Catalano, & Arthur, 2002; Jenson et al., 2013; Kubisch et al., 2010; Lasker & Weiss, 2003). Effective collaborations for youth are those that not only function well on an organizational level (i.e., between and within organizations), but also successfully create the conditions within a community so that children and youth can thrive. That is, effective collaborations for youth not only develop strategies to solve community problems; they also develop and maintain the relationships between participating members and organizations so that service delivery systems and infrastructure for implementing the strategies can be created or strengthened (Keast & Mandell, 2012).

Although there is evidence demonstrating that collaborative work can build connections across contexts (e.g., Jenson et al., 2013), there is also a plethora of evidence to indicate that collaborative work is difficult and time-consuming (e.g., Brown, Feinberg, & Greenberg, 2010; Feinberg, Bontempo, & Greenberg, 2008; Kadushin, Lindholm, Ryan, Brodsky, & Saxe, 2005) and too often does not have much impact (Huxam, 2003; Kadushin et al., 2005). Collaborative meetings take time out of staff members' days and collaborative tasks may divert organizational resources to actions or plans that may not be directly relevant for individual organizations (e.g., Kadushin et al., 2005). When disagreements or conflicts arise within the process of collaborative work, the resources organizations provide to a collaborative effort may seem to be wasted (e.g., Huxam, 2003; Kadushin et al., 2005).

But, then, why would multiple organizations work together to collectively solve problems? Broadly speaking, organizations work together because of the potential of a "collaborative advantage" to resolve deeply entrenched problems faced by communities (e.g., community violence, poverty; Gibson, Smyth, Nayowith, & Zaff, 2013; Gray, 1989; Huxam, 2003; Lasker, Weiss, & Miller, 2001). The concept of "collaborative advantage" (e.g., Gray, 1989; Huxam, 2003; Lasker, Weiss, & Miller, 2001) posits that organizations working together can achieve goals and outcomes that they would not have been able to achieve alone (Huxam, 2003; Lasker, Weiss, & Miller, 2001). There are additional reasons for collaboration, including power, trust, and common goals (Huxam, 2003). Whether the underlying reason for collaboration is opportunities for advantage or a desire to work collectively, collaboration among organizations has been increasing in frequency since the 1990s (Jenson et al., 2013). However, there remains a major challenge of ensuring that collaborations are effective, that they have impacts on the lives of the youth and families in the communities. In order to determine what collaborations are effective and why, I first define a standard of evidence by which to determine efficacy.

A Standard of Evidence

In discussing evidence of efficacy, it is important to define what types of research qualify as evidence. Currently, many fields and organizations have standards of evidence used for determining quality of an evidence base, and organize clearinghouses of EBPs that fit these standards of evidence to facilitate the selection of EBPs in communities (e.g., Blueprints for Healthy Youth Development, Coalition for Evidence-Based Policy, National Registry of Evidence-Based Programs and Practices, Office of Juvenile Justice and Delinquency Prevention, What Works Clearinghouse, and Society for Prevention Science; Center for the Study and Prevention of Violence, 2014; Coalition for Evidence-based Policy, 2014; Flay et al., 2005; National Institute of Justice, 2014; NREPP, 2014; OJJDP, 2014; United States Department of Education, 2014).

These clearinghouses of program evidence tend to use standards of evidence such that EBPs are those that fit one into of three categories: 1. programs that have a measurable relation to individual development; 2. a suggested impact on individual development; or 3. provide no evidence for measurable relation. Many of these clearinghouses have as their highest tier of EBPs those with two or more randomized control trials (RCTs) demonstrating efficacy (e.g., Coalition for Evidence-Based Policy, NREPP, and OJJDP; Coalition for Evidence-Based Policy, 2014; National Institute of Justice, 2014; NREPP, 2014; OJJDP, 2014). The selection and use of EBPs in communities is directly influenced by these clearinghouses' emphasis on RCTs as the "gold standard" of evidence (McCall & Green, 2004).

This default to RCTs as the best evidence available may be a result of many organizations or researchers using a prevention science model, which is often consistent with a variable-centered approach (Lich, Ginexi, Osgood, & Mabry, 2013). A prevention science model identifies risk factors that affect later outcomes and, as a result, focuses on programs that target the specific risk factors to ensure improved outcomes later in life (e.g., Coie et al., 1993; Kellam & Langevin, 2003). Whereas most fields have emphasized an approach using RCTs to demonstrate impacts of interventions, prevention science in particular has emphasized RCTs, in part because the approach focuses on individual levers and purported "mechanisms" for change (e.g., Lerner, Agans, DeSousa, & Hershberg, 2014; Lerner & Callina, 2014; Lerner, Lerner, & Zaff, in press; Lich et al., 2013).

However, researchers have also argued that prevention science frames, or the focus on individual levers for intervention, while useful, are not enough, particularly where understanding developmental outcomes is concerned (e.g., Guerra & Bradshaw, 2008; Lerner, 2012; Lerner et al., 2014; Lerner & Callina, 2014; Lerner, Lerner, & Zaff, in press; Lich et al., 2013; Overton, 2013; Urban, Osgood, & Mabry, 2011). CCIs using prevention science models have demonstrated impacts (e.g., Hawkins, Catalano, & Arthur, 2002; Spoth et al., 2004), and may also impact the larger *youth system.* However, in order to move less supportive *youth systems* toward becoming more supportive of youth, I argue that programs and evaluations should use a systems-oriented approaches (Guerra & Bradshaw, 2008; Lerner, 2012; Lich et al., 2012; Overton, 2013; Urban, Osgood, & Mabry, 2011). A systems-oriented approach also suggests an emphasis on different types of evidence, ones associated with methods other than RCTs (e.g., Heckman, Ichimura, & Todd, 1998; McCall & Green, 2004; Schorr, 2012; Urban, Osgood, & Mabry, 2011).

Indeed, there is some debate as to whether RCTs are necessary to demonstrate efficacy of programs in general, and community-wide efforts more specifically (Kubisch et al., 2010; Flay et al., 2005; Schorr, 2012; Urban, Osgood, & Mabry, 2011). For example, researchers have argued that RCTs are useful for understanding relatively simple interventions with small populations, but that this methodology is not useful for understanding large, complex issues with large populations (Schorr, 2012; Smyth & Schorr, 2009; Urban, Osgood, & Mabry, 2011). Moreover, there is little evidence that RCTs provide ecological validity (Freund & Isaacowitz, 2014; Lerner & Callina, 2014; Schorr, 2012; Smyth & Schorr, 2009; Urban, Osgood, & Mabry, 2011). Ecological validity occurs when research is representative of what exists in everyday life (Brewer, 2000). For example, when RCTs of evidence-based community programs are scaled up in real-world situations, the results are rarely replicable (Smyth & Schorr, 2009). This lack of replication may be because of failure to implement specific program features that were the underlying reasons for success, or it may be that the original RCT study created artificially constrained situations under which the study was conducted, situations that do not exist in the real world.

For example, in medical treatment RCTs, patients are constrained to participate based on a narrow profile of medical diagnoses (Olsen, Aisner, & McGinnis, 2007; Smyth & Schorr, 2009). Technically, then, when the treatment becomes “proven,” it is only proven for the small population under the very specific set of circumstances. However, people who will be prescribed the medicine have varying degrees of severity of diagnoses and may have additional diagnoses at one time, resulting in differential reactions to treatment. As a result, the treatment may not work in the larger population the same way it worked in the RCT that was implemented (Olsen et al., 2007; Smyth & Schorr, 2009).

An analogous example may be made for social programs, particularly those that are complex and community-based. It is difficult to constrain treatments for a community-based intervention to a particular group, and contamination across groups, even when treatments are effectively constrained, is still possible (Schorr, 2012; Urban, Osgood, & Mabry, 2011). For example, even if a community attempted to constrain their treatment group to a specific sample of adolescents, adolescents may use social media to discuss or share their treatment with adolescents in control conditions or in other treatment groups. In addition, if a community-based program were able to effectively constrain their intervention group and then make random assignment, the findings would still only be true for a particular group, in a particular context, at a particular historical time. As a result, it is unlikely that communities would be able to replicate these conditions (Smyth & Schorr, 2009).

One of the particular concerns in reliance on RCTs is the way many practitioners, clearinghouses, and researchers treat the resulting evidence. Often, when a program has conducted several RCTs and reported evidence in support of the theory of change for the program, it is treated as a "proven" program (e.g., Blueprints for Healthy Youth Development, What Works Clearinghouse; Center for the Study and Prevention of Violence, 2014; United States Department of Education, 2014). However, these programs are only proven to work for a particular set of youth at a specific time point in a specific community. That is, RCTs only demonstrate the programs work for a given set of people, times, and settings. This methodology does not help determine for whom else, where else, and when else interventions work, which is important to consider when applying an intervention in a different community. There is no guarantee that these programs, when applied to other youth at another time in another community, will be effective.

There are a variety of other methodologies from which communities can choose to evaluate programs and community-wide efforts. Appropriate methodologies to address RDS-based questions may include using multiple data sources, matching methods to the research questions posed, and identifying core components that work across programs (Schorr, 2012). To analyze such data, an assortment of statistical methods could be used including econometric methods, such as propensity score analysis, instrumental variable (IV) analysis, and regression discontinuity designs. These statistical methods enable a researcher to answer questions such as “for whom does an intervention work” instead of only “does an intervention work” (Lerner, Lerner, & Zaff, in press).

Propensity score analysis attempts to reduce bias by controlling for covariates that could predict receipt of or reaction to treatment (Heckman et al., 1998). IV analysis is used to control for confounding variables in observational data, where the instrument is a variable that predicts variation in the treatment variable, and only predicts the outcome variable indirectly through the treatment variable (Martens, Pestman, de Moer, Belister, & Klungel, 2006). Regression discontinuity designs are also used in situations where random assignment is not feasible. These designs assign participants to control or treatment groups based on a pre-test score cutoff. Under the assumption that the assignment to control or treatment is comparable to random assignment close to the cutoff point, the relationship among the variables can be examined between the control and treatment groups to explore the role of the intervention (Lerner et al., 2014; Trochim, 2006).

By consulting clearinghouses of EBPs that emphasize RCTs, clearinghouses that do not incorporate other types of data that may be more appropriate for determining whether these programs would be applicable in different types of communities for different populations, communities may be missing possibly useful programs. In order for communities to be able to more intentionally select programs that are evidence-based according to a systems-oriented models, these clearinghouses must adjust what types of data qualify as rigorously evidence-based. Once this change has been made, more communities will be able to access EBPs that have been identified by systems-oriented models and will then be able to select and implement programs based on RDS.

Accordingly, in considering the evidence base for community collaboration, I consider evidence from a variety of sources. In particular, I emphasize the appropriateness of the methodology used to answer the question, and accept evidence from a variety of methodologies. Using these criteria in succeeding sections, I define effective formal community collaboration, describe theoretical models of effective collaboration, and then describe some of the evidence of effective collaborations.

Effective Formal Community Collaborations

I define effective collaborations as those that have the capacity to solve problems, change interactions across organizations and sectors, and create change in the communities in which they work (Keast & Mandell, 2012; Roussos & Fawcett, 2000; Zakocs & Edwards, 2006). Effective collaborations develop strategies to solve community problems, as well as develop and maintain the relationships among participating members so that new systems for implementing strategies can be created (Emshoff, Darnell, Darnell, Erickson, Schneider, & Hudgins, 2007; Keast & Mandell, 2012; Spoth & Greenberg, 2011; Wolff, 2001; Zakocs & Edwards, 2006). Importantly, effective collaborations are those that not only function well on an interorganizational and cross-sector level, but also are able to identify measurable differences in their outcomes of interest (Lasker & Weiss, 2001; Roussos & Fawcett, 2000; Wolff, 2001).

Theories of Collaborative Efficacy

There are many approaches for developing effective community collaboration (e.g., Ansell & Gash, 2008; Baker, Wilkerson, & Brennan, 2012; Chinman et al., 2005; Downey, Ireson, Slavova, & McKee, 2008; Fawcett, Francisco, Paine-Andrews, & Schultz, 2000; Foster-Fishman, Berkowitz, Lounsbury, Jacobson, & Allen, 2001; Huxam, 2003; Kubisch et al., 2010; Lasker & Weiss, 2003; Wandersman et al., 2008; Zakocs & Guckenburg, 2007). These conceptions all converge around many similar ideas for what makes collaborations effective. Most tend to include themes of clarity of vision, trust, problem-solving, and empowerment to move communities toward building capacity, which leads to specific structures and processes, which in turn lead to improved outcomes (e.g., Lasker & Weiss, 2003; Lawson et al., 2007; Luke et al., 2010; Spoth et al., 2004).

In particular, all of these conceptions emphasize that specific organizational and interpersonal factors need to be in place before collaboration can be successful. For example, collaboration members must have established trust among each other in order for the collaboration to work effectively (e.g., Lasker & Weiss, 2003; Lawson et al., 2007; Luke et al., 2010; Spoth et al., 2004). Once collaboration members trust each other, they can begin to establish a clear vision and goals for the collaboration, as well as explicit roles for each collaboration member. In addition to established structures, like clear roles and responsibilities for members, transparent processes (e.g., decision-making, conflict resolution) are integral to the success of collaboration. Finally, once all of these factors are in place, the collaboration can operate effectively. Subsequently, collaborations will be more likely to have an impact on the outcomes of interest if they are operating effectively (e.g., Feinberg, Bontempo, & Greenberg, 2008; Lasker & Weiss, 2003; Lawson et al., 2007; Luke et al., 2010; Spoth et al., 2004). However, many of these approaches to collaboration remain untested (e.g., Ansell & Gash, 2008; Downey et al., 2008; Lasker & Weiss, 2003). In the section that follows, I discuss evidence of effective collaboration.

Evidence of Efficacy

The conceptions discussed above depict ideal collaborative situations. The evidence of the efficacy of collaborative work in practice is less certain (e.g. Hawkins, Catalano, & Arthur, 2002; Jenson et al., 2013; Lasker & Weiss, 2003). Relatively few formal collaborations have been able to demonstrate that they create measurable impacts on the lives of the youth and families in the broader community (Hawkins, Catalano, & Arthur, 2002; Jenson et al., 2013; Kubisch et al., 2010; Lasker & Weiss, 2003).

However, some collaborations have demonstrated impacts at the community level. Here I illustrate what effectiveness may look like by describing an exemplar approach to collaboration that has demonstrated impact, Communities that Care (CTC; Brown et al., 2009; Hawkins et al., 2002). Although not reflective of an RDS-based approach (e.g., it does not emphasize the person-context relation as the basis for development and therefore the basis for intervention), I chose to use CTC as an example because it provides an illustration of a collaboration that uses a prevention science frame with demonstrated community-level outcomes.

Specifically, CTC has a five-step process where communities first gather more information about the process of beginning a CTC collaboration and begin to build the capacity to collaborate. Then the community organizes and develops a vision statement and a timeline for implementing the rest of the CTC steps. Next, the community identifies needs of the youth in the community, current resources available for those needs, and gaps between the needs and current resources available (Hawkins et al., 2002). The community subsequently develops a community action plan that includes selecting evidence-based programs from the Blueprints for Healthy Youth Development website. Finally, the community implements the programs and continually evaluates the implementation to ensure fidelity (Hawkins et al., 2002).

Embedded throughout all five steps is the use of the social development model (SDM), which emphasizes the processes through which adults and children bond in the community (Hawkins et al., 2002). The social development model posits that children are socialized through relationships with peers and through opportunities in and out of school, their degree of engagement with these relationships and opportunities, the strengths and skills they possess to engage, and the reinforcement they receive as a result of involvement (Catalano & Hawkins, 1996). This model provides a useful frame through which to view youth risk behaviors. For example, youth who bond to peer groups with norms to engage in risk behaviors, and have such behaviors rewarded through popularity, would be more likely to engage in risk behaviors. An intervention would therefore be designed to create positive social norms and create incentives for engaging in positive behaviors.

There is evidence demonstrating that the CTC approach to community collaboration is effective at moving community level indicators toward better outcomes (Hawkins et al., 2009; Hawkins et al., 2012). In addition, there is evidence demonstrating that use of the CTC model leads to sustained, quality implementation of EBPs (Brown et al., 2007; Brown, Hawkins, Arthur, Briney, & Fagan, 2011). For example, CTC communities demonstrated significantly lower levels of delinquent behavior, alcohol use, and cigarette use than non-CTC communities five years removed from the initial intervention (Hawkins et al., 2012).

One hypothesized way that effective formal collaborations are able to begin to link actions to outcomes is by having very specific strategic plans that incorporate EBPs (e.g., CTC and PROSPER; e.g., Hawkins, Catalano, & Arthur, 2002; Spoth et al., 2004). However, the current collaboration approaches that have demonstrated impact using EBPs do so using a prevention science model, both in informing the collaborative approach of choice and in the selection and prioritization of EBPs (Hawkins, Catalano, & Arthur, 2002; Spoth et al., 2004). Indeed, there is evidence that, although changes in these communities may be widespread, outcomes are not being optimized for all youth within the community (Oesterle, Hawkins, Fagan, Abbott, & Catalano, 2010). That is, these approaches are predicated on determining risk factors that have been tied to later impacts at the community level (Coie et al., 1993; Kellam & Langevin, 2003). As a result, the individuals within the community may not necessarily be gaining the multiple experiences that they need to show positive youth development. In other words, addressing risk factors on the community level does not guarantee that the strengths of individuals are better aligned with the assets of the communities overall.

Consequently, the results of these efforts are likely not resulting in optimization for all youth. For example, whereas it is possible that, by targeting individual indicators with EBP, the larger *youth system* is shifted toward becoming more supportive of youth in the community, the RDS metamodel suggests that an approach that more intentionally accounts for the embedded nature of the youth in context may be more equipped to move youth outcomes closer to optimization. Thus, in order to optimize youth development, move toward a *supportive youth system*, and achieve even greater substantive impacts, I posit that a collaboration's strategic plan should use EBPs predicated on RDS-based models that emphasize the embodied, mutual exchanges involved in a *supportive youth system*.

Evidence-Based Programs (EBPs)

Current research on the efficacy of collaborations indicates that the use of evidence-based practices and programs is important for determining measurable impacts on outcomes of interest (e.g., Brown, Feinberg, & Greenberg, 2010; Hawkins, Catalano, & Arthur, 2002; Spoth et al., 2004). Here, I will refer to EBPs, with the understanding that these programs are the combined results of inseparable and, in some cases, as of yet inextricably tangled combinations of evidence-based practices that come together to create EBPs (e.g, Fixsen et al., 2005). Evidence-based practices are defined as "skills, techniques, and strategies that can be used by a practitioner" that are based on research-derived data (Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005, p. 26). In contrast, EBPs are defined as "collections of practices that are done within known parameters (philosophy, values, service delivery structure, and treatment components)" (Fixsen et al., 2005, p. 26). Therefore, EBPs are intentional collections of evidence-based practices, where the evidence is derived from research (Fixsen et al., 2005). In the following section, I review the use of EBPs in community collaboration.

Evidence-Based Programs in Community Collaboration

There are several approaches to collaboration focused on improving youth and family outcomes that currently have demonstrated impact using evidence-based programs (Baker, Wilkerson, & Brennan, 2012; Kubisch et al., 2010; Ladd, Muschkin, & Dodge, 2014). One of the major tenets of both CTC is that each community is provided with a wide range of EBPs from which to choose, which allows communities to select services and programs that they believe will best address the needs of the community members (e.g., Brown et al., 2009; Spoth & Greenberg, 2011; Spoth et al., 2004). In addition to selecting programs for implementation, a substantial part of incorporating evidence into collaborative work includes collecting data on how the collaborative itself is functioning (e.g., Brown, Feinberg, & Greenberg, 2012; Kegler & Swan, 2011). In the section that follows, I will discuss the process several collaborations have undertaken to measure their collaborative functioning.

Measuring Collaborative Functioning

Collaborations have used a variety of different measures to examine how they function. The vast majority of these measures examine factors that have been tied to improved collaborative functioning, such as clarity of roles for collaboration members, trust among collaboration members, shared accountability, shared vision and mission, and collective efficacy (e.g., Brown, Feinberg, & Greenberg, 2010; Fawcett et al., 2004; Lasker & Weiss, 2003; Lawson et al., 2007; Marek, Brock, & Salva, 2015; Roussos & Fawcett, 2000; Seldon, Jolin, & Schmitz, 2012). For example, CTC has measures of collaboration functioning that they use with each of their collaborations in different communities (e.g., Brown, Feinberg, & Greenberg, 2012). These measures ask questions about leadership, interpersonal relationships, task focus, participation benefits/costs, sustainability planning, and community support (Brown, Feinberg, & Greenberg, 2012). These constructs were selected because of previous theoretical and empirical work that provided evidence of their importance to collaborative functioning, as well as extensive scale testing and factor analysis (Brown, Feinberg, & Greenberg, 2012).

Indicators of particular importance for collaborative functioning fall under one of two categories: the structure of the collaboration and the collaborative processes (Kubisch et al., 2010; Marek, Brock, & Salva, 2015). Structural indicators of importance include governance, leadership, roles and responsibilities, whereas process indicators include organizational data use and shared vision (Alexander, Christianson, Hearld, Hurly, & Scanlon, 2010; Brown et al., 2007; Butterfoss, 2006; Feinberg, Bontempo, & Greenberg, 2008; Feinberg et al., 2010; Hawkins, Catalano, & Arthur, 2002; Kubisch et al., 2010; Spoth et al., 2004). In the sections that follow, I discuss the importance of each of these indicators.

Structure. Several aspects of the structure of collaborations have been demonstrated to be important to the long-term success of the collaboration and its functioning (e.g., Lasker & Weiss, 2003; Lawson et al., 2007; Luke et al., 2010; Spoth et al., 2004). For example, structural factors such as the governance, leadership, and roles and responsibilities have been tied to the longevity of the collaborative work, and in turn, its efficacy (e.g., Alexander et al., 2010; Butterfoss, 2006; Feinberg, Bontempo, & Greenberg, 2008).

Governance. Agreement on how the collaboration is managed, including decision-making processes, membership, and resources, has been shown to be important for collaborative functioning (Alexander et al., 2010; Lasker & Weiss, 2003). For example, Alexander and colleagues (2010) found that determining the governance structure was a formative challenge to collaborations. In addition, Ansell and Gash (2008) found that establishing these guidelines collectively was an important part of the collaborative process, and facilitated the building of trust and common ground among collaboration members.

Leadership. Having the “right” leaders is a subject that has been very prevalent in collaboration literature (e.g., Baker, Wilkerson, & Brennan, 2012; Kubisch et al., 2010). As a result, determining whether the leadership that is currently part of the structure of the collaborative has the skills, knowledge, and support to succeed is incredibly important (e.g., Brown, Feinberg, & Greenberg, 2010; Kubisch et al., 2010; Wolff, 2001). As a result, here I examine measures of leadership concerning whether the current leadership structure is set up such that those in power are able to adeptly help their collaboration navigate the community and its needs (e.g., Brown, Feinberg, & Greenberg, 2010; Wolff, 2001).

Roles and Responsibilities. Clearly defined roles and responsibilities are integral to effective collaboration functioning (e.g., Kubisch et al., 2010; Lasker & Weiss, 2003). Explicit roles and expectations help increase accountability among collaboration members, which creates a culture of motivation in collaborative work (Feinberg, Bontempo, & Greenberg, 2008; Kubisch et al., 2010).

Process. In addition to structural elements, there are integral parts of the process of collaborative work that have been tied to collaborative success (e.g., Kubisch et al., 2010). For example, both the use of data and evidence-based practices and programs and the development and implementation of a shared vision among collaborative members have been tied to collaborative success (Brown et al., 2007; Feinberg et al., 2010; Hawkins, Catalano, & Arthur, 2002; Kubisch et al., 2010; Spoth et al., 2004).

Organizational Data Use. As discussed above, the use and implementation of data and evidence-based practices and programs is demonstrably important in the success of several collaborative approaches (e.g., Brown et al., 2007; Feinberg et al., 2010; Hawkins, Catalano, & Arthur, 2002; Spoth et al., 2004). As a result, I will examine how collaborative members report their collaboration's gathering and use of data as part of their collaborative process.

Shared Vision. A shared vision between collaboration members has been demonstrated to impact the long-term sustainability of a collaborative effort (Kubisch et al., 2010). That is, collaboration members have discussed and agreed upon short- and long-term goals for the collaborative effort, as well as the theory of change (Kubisch et al., 2010). The process of developing the shared vision for the collaboration has been demonstrated to be an important part of the process of building trust among collaboration members and impacts the later collaboration functioning (Ansell & Gash, 2008; Kubisch et al., 2010).

These indicators of collaborative functioning have been tied to the sustainability of the collaboration itself, and in turn to the long-term impact of the collaborative effort (Feinberg, Bontempo, & Greenberg, 2008). However, as discussed above, there exist a plethora of collaborative approaches that have been used in different communities (Kubisch et al., 2010). Not all of these collaborations have developed their own measures to monitor their collaborative functioning, although monitoring may be necessary (Kubisch et al., 2010). Indeed, many of these collaborations may not have the resources with which to develop measures specific to their collaborative. Thus, measures of collaborative functioning that apply across many different approaches to collaborative work would contribute to this growing field of work.

As a result, in this dissertation, I examined the underlying structure of the collaborative functioning measures across multiple collaborative approaches in multiple communities. Because I found that the structure of the measures was the same across collaborative approaches, I was able to then examine how the items behaved.

Measures of Collaborative Functioning across Collaborative Structures

There is a lack of validated measures of collaborative functioning, which has impeded research on this topic. I addressed this gap in the extant literature by conducting a series of analyses to assess the invariance of collaborative functioning measures for youth-focused collaborations (e.g., collaborations focused on high school graduation, career readiness, health and wellness). Importantly, these collaborations did not all adhere to one specific collaborative approach, and instead each drew from different perceived approaches. Therefore, I examined whether measures of collaborative functioning can validly and reliably be used more broadly by collaborations, not only within specific approaches.

First, however, I must address the part-whole problem inherent in measuring collaborative functioning and collaborative work more generally. Specifically, researchers use individual-level measures to infer something about the whole, or in this case, the functioning of the collaboration. Because collaborations are comprised of individuals, the behavior and perspective of each individual contributes to the behavior of the collaboration as a whole, but the individual measures may not be enough to measure collaborative functioning more generally. Neighborhood research encounters this same issue (e.g., in regard to indexing collective efficacy or neighborhood effects; Sampson, Raudenbush, & Earls, 1997; Small & Supple, 2001). Generally, for neighborhood-level measures, best practices recommend measuring neighborhood-level indicators with a separate sample from the sample with individual-level indicators (Leventhal, Dupéré, & Shuey, 2015; Raudenbush & Sampson, 1999). This differential neighborhood-level sampling can be accomplished with a separate survey, by interviewing key informants, or through observation protocols.

In this dissertation, I was interested in collaborative-level functioning. As a result, my sample could be considered the separate sample intended for measuring collaborative functioning (Radenbush & Sampson, 1999). As such, I used individual responses to examine collaborative-level functioning, an approach consistent with best practices in other fields that deal with similar part-whole dilemmas (Leventhal et al., 2015; Raudenbush & Sampson, 1999).

Here, I considered measures of process and structure that have been tied to the ability of the collaboration to function. These measures include governance, leadership, roles and responsibilities, organizational use of data, and shared vision and goals. Other researchers have examined these factors in a variety of ways (e.g., Bush, Dower, & Mutch, 2002; Kaye, 1993; Mancini & Marek, 2004). For example, Mancini and Marek (2004) developed a measure examining community-based programs for families and factors related to their sustainability, called the Program Sustainability Index (PSI). These scales assessed within organization processes and between organization processes, and contained seven factors measured by 53 items. These items were then rated on a three-point Likert scale from not at all to very much. Preliminary examination of these scales indicated that six of the factors were retained, including leadership competence, effective collaboration, demonstrating program results, strategic funding, staff involvement and integration, and program responsivity (Mancini & Marek, 2004).

Another series of measures examining collaboration functioning has been created by Kaye (1993). These measures included the clarity of the collaboration's goals and vision, collaboration structure, outreach and communication, meeting procedures, member responsibility and growth, project efficacy, use of research, sense of community, costs and benefits of participation, and relationship with larger institutional structures (Kaye, 1993). Similar to the PSI, these measures were designed to be used across a variety of collaboration types in a variety of different communities (Kaye, 1993). However, these measures have not been statistically validated.

Still other researchers have framed measures of collaboration functioning as community capacity, developing an index in an attempt to measure how well a community was able to recognize, evaluate, and address problems (Bush, Dower, & Mutch, 2002). The community capacity index examined four domains:

network partnerships, knowledge transfer, problem solving, and infrastructure (Bush, Dower, & Mutch, 2002). Each domain had three levels of capacity that were measured. This index has also not been statistically tested, but it has been used in the field across different collaborations focused on a wide range of outcome goals (Bush, Dower, & Mutch, 2002).

Other measures include the work done by Communities that Care (CTC; Brown, Feinberg, & Greenberg, 2012). These measures were developed to be more approach-specific and, as a result, many of the questions pertain to how well the collaboration is adhering to the collaborative approach. However, there are some questions contained in these measures of collaborative functioning that can be applied more extensively to other types of approaches to collaboration. These measures were statistically validated, and had their structure examined across communities using the same collaborative approach (i.e., a centralized approach; Brown, Feinberg, & Greenberg, 2012). Thus, my research question and hypotheses addressed the use of measures of collaborative functioning across collaborative approaches.

Research Question and Hypotheses

My overarching research question concerned whether measures of collaborative functioning can be used across collaborative approaches. Specifically, I examined the following question: Are measures of collaborative process and structure invariant across perceived collaborative approaches? I conducted two sets of analyses (an exploratory factor analysis and a Rasch analysis) to help answer this question, using data that were derived from a larger survey designed to measure collaborative functioning. This survey was conducted with members of a diverse group of community collaborations focused on improving high school graduation rates, among other youth development outcomes.

Analysis 1: Underlying structure of measures of collaborative functioning used across collaborative models.

Whereas the underlying structure of the several measures used have been examined within-individual collaborative approaches (e.g., Brown, Feinberg, & Greenberg, 2012; Mancini & Marek, 2004), the underlying structure of collaboration functioning measures have not been examined across collaborative approaches. Thus, I examined the underlying structure of measures of collaboration functioning as they were used across collaborative approaches.

Despite the small sample size, the REFA enabled me to investigate the underlying structure of and relationships among the measured variables of the collaborative survey. A REFA is designed for use with small data sets. The size of the data set available to me was not large enough to split the sample in half and conduct an EFA on one sample and a CFA on another. I hypothesized that a consistent factor structure would be found across collaborative approaches. In addition, I hypothesized that these factor structures would be reflective of the designed measures (i.e., for structure: governance, leadership, and roles; for process: shared vision and data use).

Analysis 2: Invariance of measures of collaborative functioning across collaborative approaches

To account for variation in measurement and variation in collaborative approach, I conducted a Rasch analysis. The Rasch analysis combines item response theory and validity testing and ascertains invariance of a measure by items across groups. That is, a Rasch analysis tests whether the use of a measure varies significantly across groups. This analysis permitted me to statistically clarify which items in the collaborative survey were invariant measures of collaborative functioning across different collaborative approaches. I hypothesized that the items would be invariant across collaborative approaches.

Data Source

To examine the above hypotheses, I used data from a larger project focused on how communities can best serve the needs of youth and families within those communities. A subset of this project was the collaboration survey, which included measures of collaborative functioning. In the next chapter, I will describe the methodology of the larger study.

CHAPTER 2: METHOD

In this chapter, I describe the method of the overarching study. From 2011 through 2014, the Center for Promise (CFP; the research center of America's Promise Alliance; APA) engaged in a research project with 15 communities. These communities were recruited to the research project through their partnership with APA. Each of these communities were offered the opportunity to engage in multiple phases of a research project aimed at examining how community collaborations functioned within these communities.

Sample

The sample consisted of 183 participants from seven communities. Each community had only one collaboration, and the seven communities represented a wide range of urban centers in the United States. The communities and their basic demographic information are described in Table 1.

The community with the lowest poverty rate was Sonoma County, CA (11.9%), while the community with the highest poverty rate was Jackson, MS (30.2%). The community with the smallest population was Jackson, MS (172,638), while the community with the highest population was Tucson, AZ (526,116). The majority of the collaborations served entire cities or metropolitan areas (e.g., New Orleans, Mobile), and only one focused on small, geographic areas within larger cities (i.e., Durham).

Each collaborative had on average 16.21 ($SD = 4.04$) survey participants, with the number of participants per community ranging from five to 69. The average age of participants was 44.54 (SD = 12.56), and the youngest participant was 22 years old, whereas the oldest was 75. The sample was 58.47% female, and 26.22% male (15.30% missing gender information). The sample was highly educated, with 90.15% having obtained at least a bachelor's degree. The roles held by each participant in their individual organization varied widely both within and across collaborations. Participants included executive directors of organizations, school district staff members, mayor's office representatives, front line providers (e.g., after-school providers, teachers), and members of local businesses.

Procedure

The three phases of the project included a document review, a collaborative level survey, and a site visit. Protocols and measures for each of these phases were based on existing literature on collaborative development, functioning, and efficacy (e.g., Brown, Feinberg, & Greenberg, 2012; Bush, Dowers, & Mutch, 2002; Fawcett, Francisco, Paine-Andrews, & Schultz, 2000; Goldsmith & Eggers, 2004; Kaye, 1993; Lasker & Weiss, 2003; Lawson et al., 2007; Mancini & Marek, 2004; Roussos & Fawcett, 2000; Seldon, Jolin, & Schmitz, 2012), as well as results from a pilot of the larger project. The pilot took place from 2010 to 2011, and included interviews from key informants in four of the communities, as well as instrument piloting.

The current analyses only used data from the collaborative survey. The collaborative survey was administered in seven of the 15 cities. Each community was offered the chance to administer the survey to their collaboration members free of charge. In addition, the collaboration was considered the owner of the data, provided that the CFP was allowed access to analyze and publish on the data. The communities were also invited to add questions at the end of the survey to personalize the data for each community. The CFP agreed to provide a report summarizing the results of the survey to each community, which they could use in grant applications, annual reports, or reporting out to the larger community. Finally, each community was informed that they could repeat the survey administration on a yearly basis free of charge if they wished to track their data longitudinally.

The survey was administered electronically and hosted on Qualtrics. The CFP provided suggested language in a draft email for a key informant to send out to collaborative members to invite them to participate in the survey. Each community selected people to participate in the survey who were actively involved in the decision-making processes of the collaboration. For some collaborations, this included the governing body (e.g., governing board, executive board or council). For other collaborations, this included the governing body as well as many of the collaboration participants, as the collaboration participants had a voice in the decision-making processes of the larger collaboration.

Surveys were kept open from between two weeks to a month to allow members enough time to complete the survey. During this time, collaborative members were provided with up to five reminders to take the survey. Once the survey was closed, duplicate participants had their first survey responses kept and their second discarded. No participants had more than one duplicate response.

Measures

Participants were asked to identify their collaborative approach as well as measures of collaborative functioning. These measures used in this dissertation were developed by the CFP research team and partially adapted from previous work measuring collaboration functioning described above (e.g., Brown, Feinberg, & Greenberg, 2012; Bush, Dowers, & Mutch, 2002; Kaye, 1993; Mancini & Marek, 2004). These items were selected from the larger set in part because of their hypothesized importance in the literature, as well as on the basis of the constructs' perceived importance by participants in the pilot study (Brown, Feinberg, & Greenberg, 2012; Bush, Dowers, & Mutch, 2002; Kaye, 1993; Mancini & Marek, 2004). Appendix A presents the full list of items from each of the scales.

Collaboration Type

Participants were asked to select which of the three types of collaboration best described their collaborative approach: consensus-building, intermediate, or centralized. These typologies of collaboration format were developed based on the pilot interviews that were conducted with the executive directors of the collaborations. In addition, these typologies were reflected in the literature. Although they were not explicitly described using these labels, many types of collaboration described in the literature fell into one of these three categories (e.g., Jenson et al., 2013; Kubisch et al., 2010)

The consensus-building approach was defined as a collaborative structure such that the collaboration *engage[s] with a broad constituency within the governance structure and decision-making is shared among those organizations.* The intermediary approach was defined as a collaboration with *a lead organization, with a traditional board of directors or board of trustees to drive the strategy and to bring in collaborative partners based on the agreed upon strategy*. The centralized approach was defined as a collaboration that had *one organization or individual take the lead and possess most if not all of the decision-making responsibilities.* The frequencies are presented in Table 2. Consensus-building was the most frequently selected collaborative approach, followed by intermediate, and then centralized.

It is important to note here that these responses only constitute the perceived collaborative type for each of the participants. That is, participants who belonged to the same collaborative effort could still select different collaborative types that they believed best described the collaborative effort. Thus, it is likely that these categories do not represent an objective collaboration "type," but might be linked to the participants' perceived role and power in the collaboration.

Structure

Based on prior research (Alexander et al., 2010; Butterfoss, 2006; Feinberg, Bontempo, & Greenberg, 2008), the survey contained measures of three aspects of collaborative structure: governance, leadership, and roles and responsibilities.

Governance. The governance scale was composed of four items concerning the governance of the collaboration. These items included statements such as, "decision-making processes are agreed upon" and "how leadership is determined is agreed upon," and are rated on a six-point Likert scale from strongly agree to strongly disagree (scored from one to six). The mean was 4.69 (SD = .98). The scale was adapted from Mancini and Marek (2004), where all four items were originally written as a single statement.

Leadership. The leadership scale was comprised of seven items, which were developed by CFP, and in part adapted from the community capacity index (Bush, Dower, & Mutch, 2002). The items included statements such as, "the collaboration leaders are able to build consensus across the community" and "the collaboration leaders are able to develop the capacity of the organization as a whole." These statements were rated on a six-point Likert scale, scored from one to six. The mean was 4.88 (SD = .73).

Roles and responsibilities. The roles and responsibilities scale consisted of five statements such as "collaboration member roles are defined" and "responsibilities of collaboration members are formalized." Participants were asked to indicate how much they agreed or disagreed with the statements based on a six-point Likert scale rated from strongly disagree to strongly agree, scored from one to six. The roles and responsibilities scale was adapted from the Mancini and Marek (2004) PSI. The average score was 4.54 (SD = .85).

Process

Based on prior research (Brown et al., 2007; Feinberg et al., 2010; Hawkins, Catalano, & Arthur, 2002; Kubisch et al., 2010; Spoth et al., 2004), the survey contained measures of two aspects of collaborative process: organizational data use and shared mission and vision.

Organizational data use. The organizational data use scale was adapted based on the pilot work with the four communities from Mancini and Marek's (2004) PSI. This scale included seven statements (one reverse-coded), such as "my collaboration conducts formal evaluations of its programs" and "in my collaboration, it is important to have data to show whether programs are achieving their goals." These statements were rated on a six-point Likert scale from strongly disagree to strongly agree, scored from one to six. The average scale score was 3.63 (SD = 1.44).

Developing shared vision. The developing shared vision scale was adapted from the PSI (Mancini & Marek, 2004) as well as the "clarity of your coalition's goals and visions" scale (Kaye, 1993). This scale consisted of seven items, such as "collaboration members have developed and support a common vision" and "explicit goals have been agreed upon." These statements were rated on a six-point Likert scale from strongly disagree to strongly agree, scored from one to six. The average scale score for this sample was 4.64 (SD = .89; α = .92).

Analysis Plan

To address my research question, I chose to do two separate sets of analyses for each analytic approach (one for the structure scales and one for the process scales). This decision was made because of the problems posed by the small sample size of this data set. By separating the analyses, I was able to examine more facets of collaborative functioning. However, this separation did not allow me to examine whether there was overlap between the items in the structure and process sections of the collaborative survey.

To compare the measures across collaborative approaches, I divided the sample based on their response to the perceived type of collaborative approach to which they ascribed. As described above, participants could select whether they belonged to a consensus-building, intermediary, or centralized approach. Because of the role of "lead" organizations in both the intermediary and centralized approaches as compared to a lack of a "lead" organization in the consensus-building approach, I compared the consensus building approach to the intermediary and centralized approaches combined. Dividing the samples based on identified collaborative approach allowed me to compare the measures across perceived types of collaborative approaches, which to my knowledge, had not yet been conducted. Structuring the data in this way allowed me to compare measures of collaborative functioning in collaborations that participants believed were consensus-building models against models that had more of a centralized approach.

As displayed above in Table 2, combining the centralized and intermediary structures resulted in comparison groups of 81 (consensus) and 53 (lead organization) participants, and a total sample size of 134. The consensus-building and centralized responses represented a wide range of the communities surveyed. The comparison of self-identified perceived collaborative type across communities is presented in Table 3. In the following sections, I describe my analysis plans for each of the two analyses.

Analysis 1: Examining the underlying factor structure of the measures across collaborative approaches

I analyzed the data to determine underlying factors by computing a regularized exploratory factor analysis (REFA) in MATLAB for Windows. Because of the lack of certainty that these items measured distinct constructs that form "structure" and "process" more generally, I used Geomin rotation to allow the factors to correlate in the REFA. I chose to use a REFA, rather than a traditional EFA, as a result of the small size of my sample. A traditional EFA is designed to test the underlying structure of the phenomena being studied by fitting a factor model to the sample covariance matrix. The REFA fits a factor model to a regularized covariance matrix, which is obtained through constraining the unique variances in the covariance matrix to be proportional. This constraint reduces the number of parameters to estimate the unique variances to one (Jung & Lee, 2011). Thus, by reducing the number of parameters to be estimated, the REFA allows researchers to use EFA techniques with small samples. REFA has been shown to outperform other approaches to EFAs with small sample sizes (e.g., maximum likelihood factor analysis and principal component analysis; Jung & Lee, 2011). That is, the REFA produced EFA results with an artificially constrained sample size that were closest to the results seen with the full sample as compared to maximum likelihood factor analysis and principal component analysis (Jung & Lee, 2011).

Determining the number of factors to extract in an EFA can be difficult, and as a result, there are several guidelines to assist in the selection of the appropriate number of factors (Crawford et al., 2010). The Kaiser-Guttman rule indicates that factors with an eigenvalue greater than one should be retained (Guttman, 1954; Kaiser, 1960). However, the Kaiser-Guttman rule may not be the most accurate measure for factor retention (Velicer & Jackson, 1990). As a result, I also examined the results of a parallel analysis (using both principal component analysis and principal axis factoring; Crawford et al., 2010; Horn, 1965) and a minimal average partial correlation test (Velicer, 1976) for each of the models to indicate which number of factors best accounted for the most variation in the items.

The parallel analyses compared model fit of randomly generated data that match the means and sample sizes of the sample data to the eigenvalues of the sample data. The mean eigenvalues for the random data are compared to the mean eigenvalues for the sample data, and the number of factors where the eigenvalues for the sample data exceed the eigenvalues of the random data is determined to be the best fit (Crawford et al., 2010). Parallel analyses using principal component analysis (PA-PCA) examines the factor fit for the variance in all of the data as it compares to random data generated that matches the specifications of the data set in use (Crawford et al., 2010). Parallel analysis through principal axis factoring (PA-PAF) similarly compares the sample data eigenvalues to a randomly generated data set that matches the specifications of the sample data; however, the analysis only examines the variance accounted for by the factors, and not all of the variance in the data. To date, methodologists have not agreed on which of these approaches is most effective to determine the appropriate number of factors. Specifically, the PA-PCA has demonstrated a tendency to overfactor, and the PA-PAF has demonstrated a tendency to underfactor (Crawford et al., 2010). As a result, for these analyses, I used both tests to help inform my decision about the most appropriate number of factors.

The minimal average partial correlation (MAP) test examines the average partial correlation between the variables after the effect of the factors has been accounted for (Velicer, 1976). The number of factors selected should be where the average partial correlation between the variables is lowest after the effect of the factors has been removed. The MAP can over-extract factors, so here I used it in conjunction with the PA-PCA, PA-PAF, and the Kaiser-Guttman rule (Crawford et al., 2010; Guttman, 1954; Kaiser, 1960; Warne & Larsen, 2014).

After I determined the appropriate number of factors, I examined the item loadings on each of the factors. Consistent with methodologist recommendations, I considered an item with a factor loading greater than .45 as loading onto that factor (Comrey & Lee, 1992). If the items loaded onto similar factors across the perceived collaborative approaches, I tested how similar these factor loadings were by calculating Tucker's congruence coefficient (Lorenzo-Seva & Ten Berge, 2002). This comparison of item loadings is preferred over Pearson's *r* because it tests how far away the loadings are from zero, whereas Pearson's *r* tests how far away the item loadings are from the mean of the two loadings for each factor. As a result, Pearson's *r* may give misleading results when comparing factor loadings (Lorenzo-Seva & Ten Berge, 2002). A Tucker's coefficient of .95 or higher indicates that the items are loading onto the factors virtually identically (Lorenzo-Seva & Ten Berge, 2002).

Finally, I calculated the internal consistency reliability for each of the resulting scales using only the items that loaded onto the factors (Cronbach, 1951). Internal consistency reliability tests whether responses to items that are purported to examine the same construct behave similarly in this sample (Cronbach, 1951). Scales with a Cronbach's alpha over .70 may be regarded as having adequate internal consistency reliability (Cronbach, 1951).

Analysis 2: Invariance of measures of collaborative functioning across collaborative approaches

For the items that had consistent factor structures across perceived types of collaborative approach, I conducted a Rasch analysis in R version 3.1.2 for Mac. The purpose of this analysis was to examine whether the measures of collaborative functioning were invariant across perceived collaborative approaches. In most social science research, the logical step after completing an EFA is to test the factor structure and item invariance by conducting a confirmatory factor analysis (CFA; Brown, 2014). Conducting a CFA on the same sample that provided the results of the EFA does not provide additional information about underlying structure, as I would be testing whether the data fit the model that the data produced. In order to use both an EFA and a CFA, I would have had to split the sample. However, the sample size was too small to split the sample and conduct a CFA on the other half of the sample to examine whether the factors found in the REFA would meet the stricter test of a CFA. Thus, in order to test whether the factors produced in the REFA reflected cohesive measures, I tested the data using Rasch analysis.

Rasch analysis combines item response theory and validity testing and ascertains invariance of items across groups (Bond & Fox, 2013; Rasch, 1960; Wright, 1977). Rasch analysis is a form of sample non-specific, item-response theory analysis, which is used to examine the latent constructs measured by tests, questionnaires, or survey instruments. These analyses help examine whether each item is an indicator of the larger, unidimensional latent constructs (Bond & Fox, 2013; Rasch, 1960; Wright, 1977). Specifically, if an item fits the Rasch model, it is an indicator of the latent construct (Bond & Fox, 2013).

The Rasch analysis uses the items to define the measure's latent scale in log odds units. Then the responses of each individual and each item are placed on this shared scale so that the response scales for the items and individuals can be examined. Specifically, each individual is represented in regard to what is referred to as their “person ability” (Bond & Fox, 2013). The higher the individual average response on the latent scale, the higher the person ability. For statements with Likert-type response patterns, this placement along the latent scale reflects, on average, how many of the items the individual positively endorsed. The items are then examined by what is termed item difficulty (Bond & Fox, 2013). That is, the higher the item is on the latent scale, the more difficult the item. For statements with Likert-type response patterns, this placement along the latent scale is interpreted as how likely the statement was to be endorsed. For example, an item from the governance scale stated that decision-making processes are agreed upon. If that item appears toward the top of the latent scale defined by the Rasch analysis, it indicates that this item was difficult to endorse, or that fewer participants indicated that they agreed with that statement. In sum, Rasch analysis provides a nuanced examination of how people use the response scale of the items and whether those items are good indicators of a unidimensional latent construct (Bond & Fox, 2013; Rasch, 1960; Wright, 1977).

For this dissertation, I used a specific type of Rasch testing known as the polytomous Rasch model or the partial credit model (Masters, 1982). This type of Rasch model allows the response scale to vary across the items within the factor. Although all of the questions used in the collaborative survey had the same response scale, it was possible that the meaning or difficulty of that scale varied from item to item. The partial credit model tests for that possibility, and it provides results that indicate whether the scale is the same across items within the factor (Masters, 1982). It was important to test whether the response scale for each item remained the same for items within the factor or varied because "strongly agree" for one item may be interpreted differently than "strongly agree" on another item. Often, this equality of scale across items is assumed, and as a result, it is important to examine it explicitly (Master, 1982).

The Rasch analysis treated the item response scale as ordinal, which is in contrast to the previous analyses, where the REFA treated the response scale as continuous. Although this differential treatment of the variables may appear to be at cross-purposes, each type of analysis provided different information about the behavior and structure of the items. Because these analyses were preliminary and in the interest of developing measures of collaborative functioning, gathering as much information as possible about the items and how they behave was of paramount importance. Therefore, in my analyses, I first determined what factors emerged by using a REFA. Then, in order to examine how well the items indicated each factor, I conducted the Rasch analysis to examine how the item and the person measures fell onto the latent scale for each factor.

In interpreting Rasch model results, I first examined the item fit statistics in regard to whether the items fit the Rasch model. The null hypothesis in a Rasch model is that the data fit the model, so a non-significant chi-square indicates that an item is Rasch-conforming (Linacre, 2003). Next, the Rasch model reports the infit and outfit statistics using both mean squares and standardized fits. An infit statistic examines how well the inlier data fit the model (i.e. not outliers; Linacre, 2002), and outfit statistic examines how well the outlier data fit the model (Linacre, 2002). Mean square information for each of these statistics indicates the randomness in the model, or whether there is distortion in the measurement system. A mean square value for both infit and outfit should be between 0.5 and 1.5 to indicate that the item is productive for construction of the measurement, meaning it acts as a useful indicator for the latent construct (Linacre, 2002). The standardized fit statistics are *t*-tests of the question of whether the data fit the model perfectly. A standardized fit value should be between -1.9 and 1.9 to indicate that the data have reasonable predictability. Data that are too predictable may have other dimensions or factors that are constraining the responses, and data that are too unpredictable are not well described by the current model (Linacre, 2002). Thus, I used the standardized fit value to examine whether the data were reasonably predictable. Scores higher than these values indicate that the data are too unpredictable, and scores below these values indicate that the data are too predictable (Linacre, 2002).

Finally, I examined the person-item maps, also known as Wright maps (Bond & Fox, 2013), and the threshold parameters associated with the Wright maps. These maps place the item measures and person measures visually onto the same logit scale. The parameter estimates and threshold locations (i.e., where each category in the Likert scale fell on the latent scale) are provided for each item across the categories on the logit map. The Wright map permits the visual comparison of difficulty of the items, as well as the nuanced examination of how the participants used the scale. In conjunction with the item fit and threshold measures, these maps can be used to determine whether the scale for each item should be altered.

After running the Rasch analyses using all of the data, I then used a differential item response test to examine whether there were differences in the person and item estimates across perceived collaborative approaches (Choi, Gibbons, & Crane, 2011). The differential item functioning (DIF) test can be conducted on any scale with five or more items, and examines overall model and group specific differential item functioning (Choi et al., 2011). For each item, the DIF fits three ordinal logistic nested models. The first model examines the relative ability of the person on the item, the second examines the ability of the person and the ability of the group, and the third examines the ability of the person, of the group, and of the interaction between the person and group. The analysis then provides three chi square statistics. The first chi-square statistic compares Models 1 and 3, the second compares Models 1 and 2, and the third compares Models 2 and 3. If all three chi square statistics for each item are non-significant, there are no differences between the models, and thus, no differences between the groups compared (Choi et al., 2011). This analysis enabled me to clarify which items in the collaborative survey were invariant between the two collaborative models.

There are other types of invariance testing. For example, the ideographic filter (IF) tests for invariance across people of different measures (Molenaar & Nesselroade; 2012, 2015). Specifically, the IF allows the researcher to deal with heterogeneity in responses to surveys by allowing for person-specific factor loadings (Molenaar & Nesselroade, 2015). In addition, IF allows for the detection of homogeneity at the latent level, not just the observed variable level (Molenaar & Nesselroade, 2015). However, in order to conduct this type of analysis, the data need to include multiple time points per participant. Thus, this type of analysis was not appropriate for the current data set.

Missing Data

The rate of missing data ranged from 3.7% to 17.2% across the items. For the first analysis, missing data were estimated using multiple imputation in R version 3.1.2 for Mac. Because of the rate of missingness (as noted, 17.2% at the highest), I conducted 20 imputations (Graham, Olchowski, & Gilreath, 2007). MATLAB does not allow for the use of multiply-imputed data sets in analyses and, as a result, I conducted the analyses separately on each imputed data set and pooled the results according to procedures given by Rubin (1987).

In the second analysis, missing data were estimated using conditional maximum likelihood, which is a simplified version of maximum likelihood estimation (Linacre, 2004). Using multiple imputation for handling missing data is not recommended for use with Rasch analysis because the Rasch model is not sample-specific (Linacre, 2004). Instead, conditional maximum likelihood estimation is recommended for Rasch models, where the factors involve a finite number of items (Linacre, 2004). In the following chapter, I describe the results of the analyses outlined above.

CHAPTER 3: RESULTS OF ANALYSES 1 AND 2

In the sections that follow, I present the results of the regularized exploratory factor analysis (REFA) and the Rasch analysis.

Results of Analysis 1: Underlying structure of measures of collaborative functioning used across collaborative approaches

Descriptive analyses with the means, standard deviations, and ranges for each of the items are presented in Table 4. As shown in Table 4, the items had relatively high means, where the range was generally the full response set with a few exceptions. Within this sample, it appeared that participants tended to rate the items using the response options of slightly agree or higher.

Tables 5 displays the correlations among the structure items, and Table 6 displays the correlations among the process items. As shown in Table 5, the structure items had relatively high correlations within the hypothesized factors. In other words, the items pertaining to governance were highly correlated with each other, the items pertaining to leadership were highly correlated with each other, and the items pertaining to roles and responsibility were highly correlated with each other. The items between these hypothesized factors were still correlated, but not as highly as the items within the hypothesized factors. As shown in Table 6, the process items displayed a similar pattern to the structure items. Items within the hypothesized factors of organizational data use and shared vision were highly correlated to each other and slightly correlated to items outside of those hypothesized factors.

Structure Items in the Collaborative Survey

The first set of items I examined were the items pertaining to the structure of the collaboration. These items contained indicators of three hypothesized factors: governance, leadership, and roles and responsibilities.

Factor selection analysis. The Kaiser-Guttman rule, PA-PCA, PA-PAF, and MAP analyses all indicated that three factors provided the best model fit for both the consensus and centralized collaborative structures. That is, the model with three factors had the last eigenvalue over one and had the last eigenvalue above the eigenvalues generated for the random data in the PA-PCA and PA-PAF. In addition, the three factor model had the lowest minimum average partial correlation. Thus, I chose the three-factor model within both perceived types of collaborations.

Item loadings. I pooled the results from the imputed data sets in Microsoft Excel using Rubin's (1987) guidelines. The pattern matrices for both the centralized and consensus structures are presented in Table 7. These matrices display the item loadings by factor across both perceived collaborative approaches.

Items with a loading greater than 0.45 were considered loading onto that factor (Comrey & Lee, 1992). Items loading onto a particular factor are highlighted in bold and italicized. For example, the first leadership item loaded onto Factor 1 for the consensus collaborations, but not for the centralized collaborations. Items that did not load on to any factors are underlined. As shown in Table 7, Factor 1 could be interpreted as the leadership factor. All of the leadership items loaded onto Factor 1 for at least one of the collaboration structures. Items 1 and 2 in leadership only loaded onto the factor for consensus-building collaborative approaches. Although the second leadership item did load onto factors for the centralized collaborations, it loaded almost equally onto Factors 1 and 2, and as a result, was not considered as loading onto either factor for further analyses. These items addressed leadership ability to build consensus and manage conflict within their collaboration, skills which may be more important in a collaboration structured around consensus-building.

Factor 2 can be interpreted as the roles and responsibilities factor, as shown in Table 7. All the items in roles and responsibilities loaded onto Factor 2 for both consensus and centralized approaches. Finally, Factor 3 can be interpreted as the governance factor. All four governance items loaded onto Factor 3 for both the centralized and consensus-building approaches.

Factor similarity across collaborative approaches. To examine the similarity of the item loadings onto the factors across perceived collaborative approaches, I calculated Tucker's congruence coefficient. Table 8 displays the results of the Tucker's coefficient between the factor loadings for the centralized and consensus-building approach. I was interested in the similarity of the item loadings onto the same factors; in other words, I was interested in examining how the items loaded onto Factor 1 across centralized and consensus-building approaches. This information is displayed in the diagonal of Table 8, which compares the loadings of items onto each factor across the perceived collaborative approaches. As displayed in the table, the Tucker's coefficients between all of the pairs of factors were higher than .95, which indicated that the items were loading onto the factors almost identically (Lorenzo-Seva & Ten Berge, 2002). Thus, across the two perceived collaborative approaches, these items had a virtually identical underlying structure in regard to the number of factors, as well as the pattern and magnitude of the loadings.

Internal consistency reliability. Using only the items that loaded onto each factor for each perceived type of collaborative approach, I calculated internal consistency reliability scores for each of the resulting scales. These results are presented in Table 9.

As displayed in Table 9, the Cronbach's alphas for Factor 1 (leadership) in the consensus-building and centralized approaches were equal at .93. This value indicated that scores on these subscales had high levels of internal consistency across both perceived collaborative approaches. In addition, for Factor 1 across the perceived collaborative approaches, the inter-item correlations were the same, and well within the recommended range. There was little variance in the inter-item correlations within the items for the first factor.

Factor 2 (roles and responsibilities) demonstrated similar results to Factor 1 across both the centralized and consensus-building approaches. That is, Factor 2 had alpha values that indicated high internal consistency of the scores on those factors, with acceptable inter-item correlations that were similar to each other.

Factor 3 (governance) showed slightly different results across the perceived collaborative approaches, with an alpha of .77 for the centralized approach and an alpha of .92 for the consensus-building approach. The average inter-item correlations were also different between the two perceived approaches. The centralized approach had an average inter-item correlation of .54, whereas the consensus-building approach had an average inter-item correlation of .81. The variance of the inter-item correlations within both perceived approaches was similar within Factor 3. This pattern of findings may indicate that scores on Factor 3 were more consistently reliable in consensus-building approaches than in centralized approaches.

Process Items in the Collaborative Survey

In the second portion of the REFA analysis, I examined the structure of the measures of collaborative process across both the centralized and consensus-building approaches. These items contained indicators from two hypothesized factors: organizational data use and developing a shared vision. Unlike the results from the model-testing for the structure, the results for the process across perceived collaborative approaches were less clear. I present below the results separately for the centralized and consensus-building approaches.

Centralized approach.

Factor selection analysis. For the centralized models, the Kaiser-Guttman rule suggested three factors. The PA-PCA and the PA-PAF both recommended one factor for the process items in the collaborations taking a perceived centralized approach. The MAP recommended two factors; however, as discussed above, the MAP can over-extract (i.e., it can overestimate the number of factors), and the Kaiser-Guttman rule is not always reliable for determining the number of factors (Crawford et al., 2010; Guttman, 1954; Kaiser, 1960; Warne & Larsen, 2014). Therefore, based on the results from all four tests, I chose to extract one factor.

Item loadings. Items with a loading over 0.45 were considered loading onto that factor, and are highlighted in bold and italicized (Comrey & Lee, 1992). As displayed in Table 10, Items 1 through 5 in the "organizational data use" items and all of the "shared vision" items loaded onto the single factor. These items in "organizational data use" addressed the use of formal evaluation, the use of data for feedback, the use of outside evaluators, setting aside time for the collaboration to discuss data, and the presence of staff trained in evaluation. The items that did not load addressed the collaboration's resources for evaluation and emphasized the importance of data. It may be in collaborations that participants perceive as taking a centralized approach, the use of formal evaluation and outside evaluators, data for feedback, and discussing data were part of the development of the shared vision of the collaboration, which may help explain why these items loaded onto one factor. The pattern matrix for the factor is presented in Table 10.

Internal consistency reliability. Using only the items that loaded onto the factor, I calculated the internal consistency reliability of the resultant scale. The alpha for the factor was .92, with an inter-item correlation mean of .56, and a variance of .01. This alpha indicated high internal consistency, with acceptable levels of inter-item correlations, and minimal variance in the inter-item correlations among the items.

Consensus-building approach.

Factor selection analysis. For the consensus models, the Kaiser-Guttman rule suggested four factors. The PA-PCA and PA-PAF both suggested the two-factor model as the best solution. The MAP was equally split between the imputed data sets and suggested two and three factor models as the best fit (10 data sets results suggested two factors, 10 data sets results suggested three factors). Based on the unreliability of the Kaiser-Guttman rule and the fact that the MAP can over-extract, I chose to use the two factor model.

Item loadings. Items with a loading greater than 0.45 were considered loading onto that factor, and are highlighted in bold and italicized (Comrey & Lee, 1992). Items that did not load onto either factor are underlined. As displayed in Table 11, the two factors fell along the lines of the organizational data use items and the shared vision items. Factor 1 could be described as the shared vision factor. It appeared that the measures of "shared vision" all reflected a factor of the process of collaborative functioning. Interestingly, one of the "organizational data use" items loaded onto this factor. The fourth data question addressed whether the collaboration sets aside time to discuss the program data available. It is possible that within consensus-building collaborations, group discussion of data and its implications were a part of the shared vision and goal development process for the collaboration. The pattern matrix of the item loadings is presented in Table 11.

Factor 2 could be described as the organizational data use factor. Factor 2 contained Items 1, 2, 3, and 5 from the "organizational data use" items. It is interesting to note that Items 6 and 7 for the "organizational data use" items did not load onto the factors for either the centralized or the consensus-building perceived approaches. These items addressed the resources available to the collaboration for evaluations and the belief in the importance of the use of data. It may be that these items were substantively different from questions on the use of data and the staff available to collect data.

Internal consistency reliability. Using only the items that loaded onto the factors, the alpha for the Factor 1 scale was .91, with an inter-item correlation mean of .56, and a variance of .02. The Factor 2 resultant scale had an alpha of .84, with an inter-item correlation mean of .60, and a variance of .03. These results indicated that the scores on items in both Factor 1 and Factor 2 had high internal consistency, with acceptable levels of inter-item correlation, and minimal variance in the inter-item correlation.

Results of Analysis 2: Invariance of measures of collaborative functioning across collaborative models

As discussed above, in the first set of analyses I used a REFA to examine whether measures of collaborative functioning can be used across multiple collaborative models in different communities. I found that, for this sample, items that addressed the structure of collaborative functioning had the same factor structure across collaborative models, whereas the items that addressed the process of collaborative functioning did not have the same factor structure across the two collaborative models. Therefore, I examined the behavior of the structure items in the Rasch model, but not the process items. Within the structure items measuring collaborative functioning, I found three factors, as expected: governance, leadership, and roles and responsibilities. The findings for the Rasch analyses for all three factors are presented below.

Governance Items

In the REFA, all four of the governance items loaded onto one factor for both the centralized and the consensus-building perceived approaches. As a result, I included all four of these items in the Rasch analyses.

Item fit statistics. All of the items had non-significant chi square statistics, which indicated that the data fit the model and that these items comprised a unidimensional measure of governance in collaborative functioning. The mean square infit and outfit statistics were all between 0.5-1.5, which indicated that the data reasonably fit the model. In other words, the responses were not overly predictable (which would indicate model overfit), but there was also not much unexplained randomness left in the model (which would indicate model underfit). However, the standardized scores had different results. Items 1 and 3 both had infit and outfit statistics below the recommended values (i.e., below negative two). This finding may indicate that the data were too predictable, and that these items may have been redundant. Item 2 had an infit statistic that was close to the recommended value (at -1.95) but an outlier standardized statistic below -2, which may indicate that there were not many outliers for that item. This finding can imply that the participants responses were too similar to each other, and did not provide useful variation. Item 4 had standardized statistics that fell well within the recommended range, indicating that the data from that item were neither too predictable nor had much unexplained randomness in the model. Table 12 displays the Rasch item fit statistics for the governance items.

Item parameter estimates and Wright map. The ascending locations of each item indicated that the items were increasing in difficulty or, in other words, the fourth governance item was less likely to be endorsed than the third, which was less likely than the second, which was less likely than the first. Because this survey was not one concerning ability, this finding is interesting, but reflects non-essential information. If the survey had been designed to measure ability, it would be important to have the items in ascending order of difficulty, such that participants would be more challenged with each subsequent question. However, the survey was not designed with participant ability in mind, and as a result, the order of difficulty of the items is not integral to the function of the scale. The threshold estimates provide useful information about the nuanced use of the scales by displaying how participants used the Likert scales for each item. These estimates are displayed in the Wright map in Figure 1. The threshold parameter estimates for the governance items are displayed in Table 13.

Along the top of the Wright map there is a bar labeled “Person Parameter Distribution” that appears similar to a bar graph. The spikes in this bar graph indicate where individuals fall on the log scale created by the governance items, which ranged from negative two to six for the governance items. The distribution of this graph demonstrated a somewhat normal distribution, with a spike of participants that appeared around three on the latent dimension. This spike indicates that many of the participants rated the items quite highly on the observed six point Likert scale for each indicator item. This finding was corroborated by the placement of the lines, the values of which are displayed in Table 14. If the Likert scale for each of the items was used in an ordinal manner by the participants, the thresholds would be in ascending order. As is displayed in both Table 13 and in Figure 1, this pattern was not the case for any of the items in the governance scale. The wide gaps between the top three possible ratings (three through five) demonstrated that most of the variability in useful information was in the top of the scale. That is, most of the variation was in degree of agreement (i.e., slightly agree, agree, strongly agree options) rather than disagreement, with slight differences from item to item.

For example, in regard to the second governance item (which involved membership decisions), there was almost no discernible difference in the Wright map between Categories 1, 2, and 3. The variation in the responses for that item came between one of those options and Categories 4 and 5 (agree and strongly agree). This high degree of overlap between Categories 1, 2, and 3 could indicate that these categories were not used as frequently. This finding was also reflected in the high means for the items, as demonstrated in Table 4. These items were not necessarily controversial items, and many of the participants agreed with the statements. The useful variation for these items, therefore, is in the degree of agreement rather than disagreement.

A similar issue was demonstrated with Item 4 (concerning resource and expense decisions). The Likert-type scale accounted for the most variation between Categories 2, 4, and 5, with 1 and 3 placed very close to each other. This finding indicates that the six-category Likert scale did not usefully differentiate beyond three categories for the governance items. The fact that some of the categories are out of order is a reflection of this lack of useful differentiation.

Differential item functioning test. The differential item response test can only be conducted on scales with five or more items (Choi et al., 2011). Therefore, I was unable to conduct a test of whether the item and person parameter estimates for the governance items differed between the two perceived collaborative approaches.

Leadership Items

The results of the Rasch analysis for the leadership items are presented below.

Item fit statistics. All seven of the leadership items had non-significant chi square statistics, which indicated the items were Rasch-conforming. For the mean square statistic, only two items had infit and outfit statistics that were not within the recommended range. Item 3 (the ability to develop the capacity of the organization) had a slightly lower than recommended outfit mean square statistic (.47), which indicated that this item may not be productive for measurement development (Linacre, 2002). Item 4 (the ability to develop the capacity of individuals) had slightly lower than recommended outfit and infit statistics (.39 and .37, respectively), which also indicated that this item may not be productive for measurement (Linacre, 2002). In addition, it is possible that this item could contribute to artificially inflated reliability coefficients by being overly predictable (Linacre, 2002). Table 14 contains the item fit statistics for the leadership items.

Similar to the governance items, the standardized scores had slightly different results. Items 1, 2, 6, and 7 had outfit and infit standardized statistics close to or within the recommended range. These items addressed the ability of leaders to build consensus, handle conflict, motivate members, and their understanding of the problems in the community. However, Items 3 through 5 had outfit and infit standardized statistics outside of the recommended range, all below negative two, which indicated that the data for these items were too predictable. It is possible that these items were redundant, or that there was not enough variation in the responses.

Item parameter estimates and Wright map. As the location estimates of each of the items indicated, this scale was not as neatly ordered as the governance scale. The items did not move from "easiest to endorse" to "hardest to endorse." Instead, it appeared that Item 7 (leader understanding of community problems) was the easiest to endorse, whereas Item 2 (leader ability to manage conflict) was the most difficult. As discussed above, the scale was not organized so that the items would be in ascending order of difficulty; but these results inform where items within the leadership scale fall in regard to difficulty to endorse. In addition, the item response scales differed across the leadership items. Indeed, as shown in Table 4, Leadership Items 1, 2, 3, and 7 had a range from 2-6, which indicated that only five categories were used. As a result, the Rasch model was not able to estimate the category thresholds for the sixth category. The threshold parameter estimates for the location items are displayed in Table 15. Figure 2 contains the Wright person-item map displays the threshold information contained in Table 15.

As was demonstrated with the governance items, the person parameter distribution showed a generally normal distribution, with a spike of participants that appeared at the higher end of the scale. This finding corroborated what was shown in Table 4, with relatively high averages for each of the items. In addition to the difference of scale (i.e. five categories versus six categories), a similar problem with differentiation appeared in the leadership items that emerged for the governance items. That is, it appears that the majority of the variation was in how strongly participants agreed with each of the statements, and that the six-point Likert-type scale may have included too many options.

Differential item functioning test. I then conducted the DIF test on the leadership items to examine whether there were differences in the Rasch model estimates across collaborative structures. The DIF test returned non-significant chi-square statistics for each of the three tests (i.e., between the nested Models 1 and 3, 2 and 3, and 1 and 2, shown in Table 16). This result indicated that there were no significant differences in the Rasch estimates between the centralized and consensus-building collaborative approaches (i.e., that respondents from both collaborative approaches used the response scales in a similar way).

Roles and Responsibilities Items

The results of the Rasch analysis for the roles and responsibilities items are presented below.

Item fit statistics. Items 1, 2, 4, and 5 in roles and responsibilities had non-significant chi square statistics, meaning that those items fit the Rasch model. However, Item 3 (roles subject to renegotiation) had a significant chi square value, indicating that this item was not Rasch-conforming. Thus, the unidimensional measure of roles and responsibilities would consist of only Items 1, 2, 4, and 5. Table 17 contains the item fit statistics for the roles and responsibilities items.

Of the items that fit the Rasch model, the mean square outfit and infit statistics indicated that these items were productive for measurement (Linacre, 2002). However, of the items that fit the Rasch model, only Item 4 (whether roles are formalized) had outfit and infit standardized statistics within the recommended range. Items 1 (roles defined), 2 (roles understood), and 5 (work assignments given) had outfit and infit standardized statistics below the recommended value (Linacre, 2002). That is, the data from these items may have been too predictable, and it is possible that the model was overfitted.

Item parameter estimates and Wright map. As with the leadership items, the roles and responsibilities items were not in order of easiest to endorse to hardest to endorse. The easiest to endorse item was Item 1 (roles defined), but the hardest to endorse Rasch-conforming item was Item 4 (whether roles were formalized). Although Item 3 had fewer used response categories than the other items in roles and responsibilities, that item was not Rasch-conforming, and as a result the threshold differences should not be interpreted. Table 18 contains the threshold parameter estimates for the roles and responsibilities items. The Wright person-item map of the threshold differences displayed in Table 18 is shown in Figure 3.

The person parameter distribution again demonstrated a fairly normal distribution of persons across the top, with a spike higher up on the scale. As with the governance and leadership scales, this spike indicated that a greater proportion of participants had higher responses on the scale. This finding was consistent with the high item average scores for the roles and responsibilities items displayed in Table 4.

The threshold estimates for the Rasch-conforming items in roles and responsibilities indicated that the response scales for these items were used more like ordinal scales by participants than the response scales for the items in either governance or leadership. For example, Item 5 (work assignments given) had all of the categories in order, which indicated that the order of the Likert items was used by participants in a manner consistent with its design. However, the distances between categories were not commensurate for Item 5, which indicated that the distance between categories did not carry equal weight. In other words, the distance between disagree and slightly disagree (Categories 1 and 2) was not the same as the distance between slightly agree and agree (Categories 3 and 4). Item 4 (whether roles are formalized) was also in ordinal order and had more even spacing between the categories, which indicated an almost perfect ordinal scale use by the participants. Items 1 and 2, however, had threshold statistics that indicated the presence of the same problems I observed in the governance and leadership scales. The categories were not necessarily in order, and there were categories that were very close together or on top of one another, which indicated that they did not provide meaningful variation in responses. Thus, for Items 1 and 2, the responses were most strongly positive for the participants, and Items 4 and 5 contained more meaningful variation across the full Likert-type scale (strongly disagree to strongly agree).

Differential item functioning test. I then conducted the DIF test on the leadership items to examine whether there were differences in the Rasch model estimates across the two perceived types of collaborative structure. The DIF test returned non-significant chi square results for each of the three tests (as shown in Table 19). This result indicated that there were no significant differences in the Rasch estimates for the roles and responsibilities items between the centralized and consensus-building collaborative approaches.

Summary

I conducted a REFA to examine the underlying structure of the items assessing collaborative structure and process. I hypothesized that the items that addressed collaborative structure would have three underlying factors, of governance, leadership, and roles and responsibilities, across perceived collaborative approaches. I hypothesized as well that the items that addressed collaborative process would have two underlying factors, of organizational data use and shared vision, across perceived collaborative approaches. I found that the items had a consistent three-factor structure across perceived collaborative approaches for collaborative structure, but no consistent factor structure across items pertaining to collaborative process.

The results for the governance scale indicated that all four items were Rasch-conforming, meaning that the items were good indicators of the latent factor of governance. The mean square infit and outfit statistics indicated that each item was useful for measurement development and acted as a good indicator of governance. However, the standardized infit and outfit statistics indicated that some of the items may have not had enough variation in responses.

This finding for the governance items was supported by the results of the threshold and person-item estimates for each of the items. The spike of persons on the Wright map higher up on the latent scale indicated that a large group of participants scored fairly high on the governance scale. The threshold estimates for each item corroborated this finding, and indicated that the useful differentiation in the scales occurred in degrees of agreement. That is, participants generally agreed with the statements in the governance scale. Variation among responses occurred within only a portion of the response scale.

The results for the leadership scale were very similar to the governance scale. All of the items had item fit statistics that indicated that they were Rasch conforming. The mean square infit and outfit statistics were acceptable for all but two items (pertaining to the development of capacity for individuals and the organization as a whole). The standardized infit and outfit statistics were acceptable for the four items that addressed the ability of leaders to build consensus, handle conflict, motivate members, and their understanding of the problems in the community. However, the three remaining items had outfit and infit standardized statistics outside of the recommended range, which indicated that the data for these items were too predictable. That is, the model fit too well, and thus was not optimal for the generation of useful information. It is possible that these items were redundant, or that there was not enough variation in the responses. Again, similar to the governance scale, the leadership items threshold estimates corroborated this finding. The DIF test indicated that these items behaved similarly across perceived collaborative approaches.

The roles and responsibilities scale had slightly different findings from both the governance and leadership scales. Only four of the items had fit statistics that indicated the items were Rasch-conforming, meaning that the final scale should likely only contain four items. The item that would be dropped pertains to whether roles were subject to renegotiation. It may be that the ongoing flexibility of the roles is not salient to the collaborative functioning. What may be salient is that the roles are formalized and are embedded within the structure of the collaboration.

Of the items that fit the Rasch model, the mean square outfit and infit statistics indicated that these items were useful for measurement development. However, three of the items had standardized infit and outfit statistics below the recommended values, which may have indicated that these items were too predictable. The DIF test indicated that these items behaved similarly across perceived collaborative approaches. The implications of these results for the collaboration literature and for research and practice are discussed in the following chapter.

CHAPTER 4: DISCUSSION

As previously discussed, RDS-based models posit that individual development cannot be separated from the surrounding context; the individual influences and is influenced by his or her environment (Overton, 2013, 2015). When the assets in the context are aligned with the individual strengths and needs of the young person, adaptive development will occur for the youth and for his or her surrounding community (Lerner, 2004; Lerner et al., 2005; Lerner & Overton, 2008; Overton, 2015). However, there is evidence that youth in some communities are not experiencing the developmental supports that they need (e.g., Kim, 2012; Luthar & Barkin, 2012; Luthar, Barkin, & Crossman, 2013; Perna & Titus, 2005; Swanson, 2009). Youth-focused community collaborations may provide a useful way to move entire communities toward becoming more supportive of youth (e.g., Jenson et al., 2013; Kubisch et al., 2010). However, it is important to ensure that collaborations are working toward efficacy.

There are multiple measures that address collaborative efficacy, including measures of collaboration functioning, which have been tied to improved youth and family outcomes (Feinberg, Bontempo, & Greenberg, 2008). Measures of collaboration functioning include measuring factors such as clarity of governance, efficacy of leadership, clarity of roles for collaboration members, processes of organizational data use, and shared vision and mission (e.g., Fawcett et al., 2000; Goldsmith & Eggers, 2004; Lasker & Weiss, 2003; Lawson et al., 2007; Roussos & Fawcett, 2000; Seldon et al., 2012).

Measures of collaborative functioning have traditionally been assessed for their psychometric characteristics within the context of specific collaborative approaches (e.g., CTC; Brown, Feinberg, & Greenberg, 2012; Kegler & Swan, 2011). However, the behavior of many of these measures had not been examined in collaborations that did not use those specific approaches. Furthermore, the psychometric properties of most other measures that were developed for use across collaborative approaches was not examined, and the structures of the measures were not statistically tested across collaborative approaches (e.g., Bush, Dower, & Mutch, 2002; Kaye, 1993).

Thus, examining the measures of collaborative functioning across perceived approaches is a vital step toward advancing the conduct and evaluation of collaborative work. Accordingly, the question I addressed in this dissertation was: Can measures of collaborative functioning be used across different collaborative approaches of improving youth developmental outcomes? To answer this question, I conducted two types of analyses: a REFA to examine the underlying structure of the items across perceived collaborative approaches, and a Rasch analysis to examine how the items behaved across perceived collaborative approaches.

In the sections that follow, I discuss the results of each of the analyses organized by type of item. First, I discuss the results for the structure items, and then I discuss the results for the process items. It is important to note that across all of the analyses, the sample sizes were smaller than is considered ideal (Brown, 2014; Linacre, 2002). Consequently, these results should be interpreted with caution. I then address the implications of this research for the collaboration literature, suggest directions for future research, and discuss practical applications.

Structure Items in the Collaborative Survey

I had hypothesized that the items in the collaborative survey would display similar factor structures consistent with the measures used across perceived collaborative approaches. In addition, I hypothesized that these items would be invariant across perceived collaborative approaches. For the structure items, these hypotheses were supported.

Analysis 1: REFA

For these items, the three factors could be considered the *governance*, *leadership*, and *roles and responsibilities* factors. According to Tucker's congruence coefficient, the items loaded onto these factors almost identically across the perceived collaborative approaches. This result indicated that the these items had a virtually identical underlying structure in regard to the number of factors, as well as the pattern and magnitude of the loadings across the two perceived collaborative approaches.

Thus, the REFA provided promising results for the structure items and the development of a survey that can be used across collaborative approaches. That is, there were distinct factors of the structure of collaboration functioning, and these factors were consistent across centralized and consensus-building collaborative approaches for this sample. This finding may imply that collaborative structures can be measured across collaborative approaches using the same items, and that these items will perform in similar ways across the perceived approaches.

Analysis 2: Rasch

I then conducted a partial credit model Rasch analysis to examine how each factor of the structure items behaved across perceived collaborative approaches (Master, 1982). More specifically, this analysis tested whether the participants response patterns were the same across items, and whether these items were good indicators of the unidimensional latent factors identified by the EFA.

Governance. All of the governance scale items fit the Rasch model. However, some of the threshold indices indicated that the participants only varied in regard to their degree of agreement with the items. A possible consequence of this finding would be considering collapsing the response categories for the governance scales. One possible solution would be to reduce the number of Likert options available to participants, but still not provide participants with a neutral category (e.g., a four point scale associated with strongly disagree, disagree, agree, strongly agree). Another possible solution would be to reduce the Likert options and provide a neutral category (e.g., disagree, neither agree nor disagree, agree). Future research could examine which type of response option provides the most useful variation in responses for the governance items.

Another approach to improving the use of the response scale would be to increase the "difficulty" of the items themselves. That is, future research could consider editing the statements such that they are more difficult to agree with. This change might yield more variation in the full use of the response scales. In order to better inform this approach, the use of qualitative data and key informants across perceived collaborative approaches would be useful.

I was unable to conduct a DIF test for the governance items, as the scale did not have enough items to satisfy the minimum requirements for the analysis in the program that I used. In future research, I would consider testing this scale with a different sample that would allow for the analytic comparison of the item results across samples.

Leadership. Similar to the governance items, all of the leadership items were Rasch model conforming. However, the items also had threshold indices that indicated that for almost all of the items, participants only differed in the scale to the extent that they agreed. As was true with governance, a possible consequence would be collapsing the categories. Future research should examine what response scale provides the most useful differentiation in participant responses for the leadership items.

The findings of the DIF test indicated that the Rasch analysis results were not significantly different across perceived collaborative approaches. This finding supported the results from the REFA, which indicated that the factor structure for the leadership factor was almost identical across the perceived collaborative approaches. Thus, it appeared that the leadership scale is one that may be used across perceived collaborative approaches, but that the response sets for these scales could possibly be reduced.

Roles and responsibilities. Four of the roles and responsibilities items were Rasch-conforming, and one item was not. As a result, for future research, users of the scale should discard the item that was not Rasch-conforming and use the four-item roles and responsibilities scale. As was true with both the governance and leadership scales, the person-item estimates indicated that many of the participants agreed with many of the statements, and much of the variance in the responses came from the degree of agreement. However, it appeared that the six-point Likert type response set was appropriate for at least two of the roles and responsibilities items (these items addressed whether roles were formalized and whether work assignments were given). For these items, participants used more of the full Likert scale.

The remaining two items in the roles and responsibilities scale had threshold estimates that indicated that a full, six-point Likert scale did not usefully differentiate across the responses for the participants for all of the items. Consequently, I would consider collapsing the response categories for those items. However, collapsing only some of the items within a scale is not ideal. Alternatively, those items could be increased in difficulty. The DIF test indicated no significant differences in item behavior between groups, lending support for the ability to use the measure across perceived collaborative approaches.

Research Implications

These preliminary findings on a small sample size produced some promising findings for the items measuring collaborative structure. Considering the diversity of roles of the individuals both within and across the collaborations, it is impressive that I found consistencies in the way the structure measures behaved. This finding implies that these structural items address systems properties, and do not measure properties of the individuals. Across both perceived collaborative approaches, the items (where testable) demonstrated similar underlying structure and behavior, indicating that these items can be used for examining the structural elements of collaborative functioning across perceived approaches. Future research should work to further validate and replicate these findings with larger sample sizes, and with multiple analytic approaches (e.g., traditional EFAs and CFAs).

In particular, these findings support the idea that there are measurable consistencies in the structure of effective collaborations. These structural pieces have been proposed and examined by others, but their quantitative measurement across perceived collaborative approaches had yet to be examined (Brown, Feinberg, & Greenberg, 2012; Fawcett et al., 2000; Goldsmith & Eggers, 2004; Kegler & Swan, 2011; Lasker & Weiss, 2003; Lawson et al., 2007; Roussos & Fawcett, 2000; Seldon et al., 2012). However, the results did have some implications for changes to the items.

Future research should examine whether a six-point Likert response scale is necessary, as these results suggested a condensed response scale would be more effective. For example, Marek and Mancini (2004) used a three-point scale from “not at all” to “very much” in their PSI. Items with this type of response set should be tested to see if they maintain the same factor structure and are invariant across perceived collaborative approaches. In addition, future research should focus on whether the factor structure for these items is consistent across additional perceived collaborative approaches beyond the two measured in this study.

Process Items in the Collaborative Survey

As previously discussed, I had hypothesized that the factor structure of the items would be consistent with the measures used, similar across perceived collaborative approaches, and that the items would be invariant across perceived approaches. These hypotheses were not supported for the items concerning process in collaborative functioning. Indeed, the process items loaded onto different factors for each perceived collaborative approach. For the centralized approach, the items loaded onto only one factor, which included items from both the hypothesized organizational data use scale and the shared mission and vision scale. For the consensus-building approach, the items loaded onto two factors, which were mostly consistent with the hypothesized two factors (organizational data use and shared mission and vision). As a result of the factor inconsistency across approaches, I did not conduct Rasch analyses on the process items.

The findings on the differential performance of items concerning collaborative process provide some insight into the behavior of measures of collaborative process, but also leave several questions. First, what are the implications of the items loading onto differential numbers of factors? It is possible that in approaches that are perceived as centralized, organizational data use and the mission and vision are seen as things that get developed as part of a top-down approach, and less something that the individual respondents participate in. Accordingly, in approaches that are perceived as consensus-building, it is possible that respondents feel more like they are participating in a bottom-up approach, and therefore separately helped develop the mission and vision, and simultaneously are aware of the data use practices. A third possibility is that the perceived categories of the collaborations are not necessarily meaningful in the intended ways and, as a result, the invariance of the process items across categories was not manifested.

A second question raised by these results is whether it is reasonable to expect that items measuring collaborative process should behave similarly across perceived collaborative approaches. It may be that because the collaborative processes by definition differ across perceived collaborative approaches, these items will not perform in a consistent manner. Indeed, items on collaborative process that had been previously examined across communities were only examined within particular collaborative approaches (e.g., Brown, Feinberg, & Greenberg, 2012; Kegler & Swan, 2011). Thus, it could be argued that it is not feasible to develop measures of process that will perform consistently across perceived collaborative approaches.

However, previous research indicates that there are common features of effective collaborative processes that are tied to improved functioning across collaborative approaches (e.g., Kubish et al., 2010). Some of these common features of effective collaborative process across approaches include trust among collaboration members, shared accountability, and collective efficacy (e.g., Fawcett et al., 2000; Goldsmith & Eggers, 2004; Lasker & Weiss, 2003; Lawson et al., 2007; Roussos & Fawcett, 2000; Seldon et al., 2012). As a result, it should be possible to develop survey measures that address whether these common features of effective collaborative processes are present in a collaboration, and thus also feasible to develop these across collaborative approaches such that the measures behave in similar ways across those approaches.

Given these findings, it is important to again note the small sample sizes used in this dissertation. It is possible that the differential findings across perceived collaborative approach for the items pertaining to process may have been a result of the small sample size. It is possible that if I had more participants, the results would have been different.

Assuming that the finding does reflect the inadequacy of the items pertaining to collaborative process, it is possible that these items may be too specific to the development of these features of effective collaborative processes. For example, some of the current survey items asked specifically if the collaboration reevaluates its goals and theory of change periodically, and if it evaluates itself in relation to those goals and theory of change. The salient part for the sake of quantitative measurement of the collaborative process may be whether a shared vision exists. In addition, items on the current survey may have been too “double-barreled” for participants to answer in meaningful ways across collaborative approaches. For example, the first item under shared vision asks whether participants have developed and support a shared vision. Thus, the question asks them both if the shared vision has been developed and also if they support it. It is possible that participants answered the question as it pertained to only one of those responses, and they had no way of indicating which portion of the statement it was that they endorsed. As a consequence, the responses from this item may have been in fact measuring different portions of the item, thus obscuring the results.

Research Implications

What is clear from these results is that for this small sample, these particular survey items concerning the collaborative process did not adequately address these common features of collaborative process that other researchers have noted in effective collaborative functioning (e.g., Fawcett et al., 2000; Goldsmith & Eggers, 2004; Lasker & Weiss, 2003; Lawson et al., 2007; Roussos & Fawcett, 2000; Seldon et al., 2012). As a result, the items for collaborative process need to be developed further such that they address the larger constructs of process that are salient across the contexts of different perceived collaborative approaches. Future research could work to develop these measures through ensuring that the items are not double-barreled and that the measures address larger constructs and not the process of developing those constructs.

Implications for Practice

I found that the items examined here that measured collaborative structure could be used across perceived collaborative approaches, but that the items measuring collaborative process cannot be used meaningfully in their current form across perceived approaches. This finding has immediate implications for collaborative practice. For example, the items measuring collaborative structure can be put into use in collaborations more broadly. Indeed, measuring collaborative functioning has been posited to be a key part of collaborative success (Feinberg, Bontempo, & Greenberg, 2008; Kubisch et al., 2010). Through intentionally examining the functioning of the collaborative as it pertains to the structure, collaborations can begin to monitor their functional progress (Kubisch et al., 2010).

With free, online survey-hosting widely available (e.g., Survey Monkey, Google Forms, Zoomerang, SurveyGizmo), collaborations could begin to longitudinally track their data as it pertains to their collaborative structure. If collaborations more broadly began to use this survey tool, it would be possible to examine structural collaborative functioning over time across perceived collaborative approaches and communities. This intentional use of data to inform the progress of the collaborative functioning can help researchers begin to better understand the development of collaborations and effective collaborative functioning, particularly as it pertains to the key structural elements of effective collaborative functioning.

For example, collaborations in the early stages may not demonstrate strong leadership, or clearly defined roles and responsibilities. It is also possible that these factors emerge differently across collaborative approaches, such that consensus-building approaches have clearly defined roles and responsibilities very early on with leadership showing a clear growth trajectory, whereas centralized approaches show the opposite.

There are theories on collaborative development phases (e.g., Butterfoss, Goodman, & Wandersman, 1993; Downey et al., 2008); however, these approaches could be further supported and examined through widespread use of similar measures. In addition, the widespread use of a common set of measures would allow collaborations to administer this survey themselves, then possibly compare their results to collaborations that have similar features.

Limitations

The studies presented in this dissertation suffered from several limitations. First and foremost, these analyses were conducted with relatively small sample sizes. One of the consequences of the small sample sizes was the necessary split of analyses between structure and process items in the collaborative survey. It is possible that the items would have loaded differently if I had the sample size to include both the structure and process items in the EFA. As a result, I cannot be sure that the factors found here are entirely unidimensional, as it is possible that the items could have loaded differently given analyses that used more items. Consequently, the results should be considered exploratory and interpreted with caution.

If I had access to a larger data set, I would have split the data set and conducted an EFA and a confirmatory factor analysis (CFA). This analysis would have allowed me to examine the factor structure of the data and compare it to how the data fit the factor structure in a more robust framework. As a result of the nature of the data available to me, I was unable to conduct these analyses, and instead had to use techniques better suited to smaller sample sizes.

In addition, the data were collected over an extended period of time in each of the communities (over a year and a half for the wave of data used here). This differential timing may have impacted the responses from the participants in each of the communities, and may have influenced the results. In addition, each of these collaborations was at a different point in the "life-span development" of the collaboration itself (Urban, Hargraves, & Trochim, 2014; Urban, Osgood, & Mabry, 2011). Thus, the differential timing, both in terms of chronologically in history and in the life-span of the collaboration may have impacted the findings.

This dissertation also examined the differences in measures across perceived collaboration types, not within the communities as the data were collected. By analyzing the data across self-reported perceived collaboration type, I lost the variation across communities. In addition, the perceived collaboration types used here were only validated by pilot interviews and supported in existing research. However, it is clear from the current results that asking the individuals to select a collaboration type does not result in the selection of an objectively present collaboration type. This possibility was demonstrated in members from each collaboration selecting different collaboration types to describe their collaborative effort. It is possible that what these individuals are describing in selecting the type is a personal experience – that they personally experience the collaboration to have more top-down functioning or bottom-up functioning. However, more research is needed to determine what the full meanings are behind the collaboration type, and to elucidate the individual experiences of participation in collaboration. In addition, future research should work to examine measures of collaborative functioning with larger sample sizes and across additional contexts, not just perceived collaboration type.

Conclusions

RDS-based theories posit that development consists of mutually influential relationships between the individual and the context (Overton, 2013, 2015). When the strengths and needs of the young person are aligned with the assets in the context, adaptive development will occur for the youth and his or her community (Lerner & Overton, 2008). Youth-focused collaborations may provide a useful way to strengthen the connections between the assets in community contexts (e.g., school, after-school programs, health, public works) and move entire communities toward becoming more supportive of youth (e.g., Jenson et al., 2013; Kubisch et al., 2010).

There is a lack of validated measures of collaborative functioning, which has impeded research on this topic. I addressed this gap in the literature by conducting a series of analyses to assess the validity of collaborative functioning measures for youth-focused collaborations (e.g., collaborations focused on high school graduation, career readiness, health and wellness). It is important to note, these collaborations did not all adhere to one specific collaborative approach, and instead each draw from their own preferred approach.

I found that items pertaining to collaborative structure loaded onto the three factors in almost identical ways across perceived collaborative approaches. The items in these three factors also behaved similarly across perceived collaborative approaches, but it is possible that the number of response options need to be reduced. However, the items pertaining to collaborative process did not perform similarly across perceived collaborative approaches. As a result, future research should work to determine what response set works best with the structure items, and work to develop process items that maintain similar factor structure and behave similarly across perceived collaborative approaches.

Tables

Table 1: Demographic Characteristics of the Communities

Community Name	Population	Poverty Rate
New Orleans, LA	378,715	27.3%
Durham, NC	245,475	20.0%
Jackson, MS	172,638	30.2%
Tucson, AZ	526,116	25.2%
Mobile, AL	194,899	23.4%
MS Gulf Coast*	382,516	18.2%
Sonoma County, CA	495,025	11.9%

Source: United States Census Bureau, 2013
*Determined by summing the population for Hancock, Harrison, and Jackson counties in Mississippi. Poverty rate is the average of the same three counties.

Table 2: Sample Sizes of Collaborative Approaches

Collaborative Type	Sample Size
Consensus	81
Intermediary	39
Centralized	14
Other	5
Missing	44

Table 3: Collaboration Type by Community Membership

Site Name	Collaboration Type				
	Consensus-building	Intermediary	Centralized	Other	Total
Mobile	10	60	2	1	19
NOLA	15	5	2	1	23
MS Gulf Coast	2	2	2	2	8
Durham	0	3	1	0	4
Jackson	36	12	2	0	50
Sonoma	4	4	1	1	10
Tucson	14	7	4	0	25
Total	81	39	14	5	139

Table 4: Descriptives of Items Measuring Collaborative Structure and Process

Variable	Mean	SD	Range
Gov 1: Decision-making	4.85	1.02	1-6
Gov 2: Membership decisions	4.77	1.08	1-6
Gov 3: Leadership decisions	4.69	1.17	1-6
Gov 4: Resource and expense decisions	4.49	1.21	1-6
Lead 1: Build consensus	4.78	.90	2-6
Lead 2: Manage conflicts	4.66	.81	2-6
Lead 3: Develop capacity – organization	4.90	.81	2-6
Lead 4: Develop capacity – individuals	4.77	.95	1-6
Lead 5: Right strategies for right levers	4.86	.91	1-6
Lead 6: Motivate members	4.94	.85	1-6
Lead 7: Understand problems	5.29	.81	2-6
Roles 1: Roles defined	4.61	1.03	1-6
Roles 2: Roles understood	4.54	.99	1-6
Roles 3: Role renegotiation possible	4.60	.95	2-6
Roles 4: Responsibilities formalized	4.40	1.09	1-6
Roles 5: Work assignments given	4.50	1.04	1-6
Data 1: Formal evaluations conducted	4.51	1.20	1-6
Data 2: Data used for feedback	4.72	1.11	1-6
Data 3: Outside evaluators used	4.31	1.36	1-6
Data 4: Discuss program data	4.58	1.09	1-6
Data 5: Staff trained in evaluation	4.40	1.28	1-6
Data 6: Lack internal resources for evaluation	3.37	1.44	1-6
Data 7: Important to have data for outcomes	5.26	1.02	1-6
Vision 1: Developed and support common vision	5.10	.91	1-6
Vision 2: Partnership goals and collaborative goals	4.88	.96	2-6
Vision 3: Goals are agreed upon	4.83	1.08	1-6
Vision 4: Goals and theory of change informed by community	4.89	.97	1-6
Vision 5: Residents and institutions aware of goals	3.93	1.27	1-6
Vision 6: Goals and theory of change are updated	4.45	1.05	2-6
Vision 7: Collab evaluated relative to goals and theory of change	4.45	1.13	2-6

Table 5: Correlations among Items Measuring Collaborative Structure

Variable	Gov 1	Gov 2	Gov 3	Gov 4	Lead 1	Lead 2	Lead 3	Lead 4	Lead 5	Lead 6	Lead 7	Roles 1	Roles 2	Roles 3	Roles 4	Roles 5
Gov 1	1															
Gov 2	.73**	1														
Gov 3	.65**	.72**	1													
Gov 4	.65**	.55**	.69**	1												
Lead 1	.45**	.35**	.48**	.48**	1											
Lead 2	.46**	.37**	.40**	.47**	.49**	1										
Lead 3	.50**	.30**	.41**	.49**	.47**	.64**	1									
Lead 4	.59**	.39**	.47**	.54**	.49**	.68**	.85**	1								
Lead 5	.53**	.33**	.40**	.58**	.54**	.62**	.76**	.80**	1							
Lead 6	.45**	.27**	.34**	.45**	.58**	.49**	.78**	.79**	.79**	1						
Lead 7	.44**	.30**	.42**	.42**	.45**	.44**	.62**	.66**	.59**	.64**	1					
Roles 1	.51**	.39**	.41**	.54**	.42**	.57**	.49**	.50**	.52**	.42**	.44**	1				
Roles 2	.53**	.43**	.43**	.53**	.53**	.59**	.54**	.53**	.57**	.46**	.42**	.89**	1			
Roles 3	.27**	.23**	.28**	.33**	.33**	.28**	.26**	.21**	.35**	.20*	.17	.46**	.47**	1		
Roles 4	.38**	.43**	.36**	.40**	.40**	.45**	.39**	.37**	.42**	.35**	.25**	.67**	.69**	.44**	1	
Roles 5	.33**	.28**	.31**	.45**	.45**	.46**	.44**	.44**	.54**	.38**	.19**	.67**	.70**	.56**	.71**	1

*p<.05 **p<.01

Note: Gov = Governance items; Lead = Leadership items; Roles = Roles and responsibilities items

Table 6: Correlations among Items Measuring Collaborative Process

Variable	Data 1	Data 2	Data 3	Data 4	Data 5	Data 6	Data 7	Vision 1	Vision 2	Vision 3	Vision 4	Vision 5	Vision 6	Vision 7
Data 1	1													
Data 2	.85**	1												
Data 3	.69**	.55**	1											
Data 4	.67**	.67**	.45**	1										
Data 5	.58**	.64**	.29**	.54**	1									
Data 6	-.17	-.27**	.03	-.20*	-.43**	1								
Data 7	.23*	.18*	.21*	.26**	.15	.10	1							
Vision 1	.34**	.38**	.11	.45**	.33**	-.21*	.13	1						
Vision 2	.49**	.50**	.32**	.55**	.45**	-.25**	.17	.67**	1					
Vision 3	.40**	.40**	.26**	.54**	.31**	-.14	.20*	.64**	.63**	1				
Vision 4	.51**	.54**	.28**	.64**	.47**	-.19*	.25**	.65**	.70**	.71**	1			
Vision 5	.44**	.38**	.22*	.52**	.33**	-.05	.08	.50**	.59**	.52**	.57**	1		
Vision 6	.51**	.56**	.37**	.56**	.49**	-.24**	.09	.52**	.60**	.51**	.64**	.58**	1	
Vision 7	.49**	.56**	.36**	.61**	.46**	-.15	.15	.55**	.71**	.59**	.72**	.69**	.78**	1

*p<.05 **p<.01
Note: Data = Organizational data use items; Vision = Shared vision item

Table 7: Factor Loadings for the Three-Factor Model of Measures of Collaborative Structure for Centralized and Consensus Collaborations

	Factor 1		Factor 2		Factor 3	
	Cen	Co	Cen	Co	Cen	Co
Gov 1	0.252	0.160	0.025	0.131	***0.633***	***0.707***
Gov 2	0.216	0.090	0.025	0.064	***0.832***	***0.824***
Gov 3	0.024	0.103	0.061	0.070	***0.820***	***0.796***
Gov 4	0.217	0.199	0.181	0.165	***0.576***	***0.493***
Leader 1	0.433	***0.745***	0.343	0.092	0.130	0.091
Leader 2	***0.452***	***0.601***	***0.473***	0.126	0.042	0.101
Leader 3	***0.907***	***0.830***	0.071	0.047	0.128	0.100
Leader 4	***0.931***	***0.840***	0.018	0.053	0.021	0.153
Leader 5	***0.714***	***0.897***	0.239	0.065	0.074	0.109
Leader 6	***0.881***	***0.906***	0.042	0.047	0.039	0.101
Leader 7	***0.733***	***0.569***	0.093	0.104	0.164	0.158
Roles 1	0.046	0.120	***0.829***	***0.800***	0.061	0.116
Roles 2	0.049	0.165	***0.823***	***0.767***	0.089	0.074
Roles 3	0.222	0.116	***0.523***	***0.604***	0.064	0.061
Roles 4	0.107	0.127	***0.809***	***0.741***	0.035	0.065
Roles 5	0.041	0.237	***0.839***	***0.728***	0.084	0.183

Note: Cen = Centralized approach; Co = Consensus-building approach; Gov = Governance items; Leader = Leadership items; Roles = Roles and responsibilities items

Table 8: Tucker's Congruence Coefficient Matrix Comparing Item Loadings onto Factors for Centralized and Consensus-Building Approaches

		Centralized Approach		
		Factor 1	Factor 2	Factor 3
Consensus-building Approach	Factor 1	.97	.40	.28
	Factor 2	.22	.96	.24
	Factor 3	.37	.26	.99

Table 9: Internal Consistency Reliability for the Factors of the Structure Items

	Factor 1		Factor 2		Factor 3	
	Cen	Co	Cen	Co	Cen	Co
α	.93	.93	.89	.90	.77	.92
Inter-Item Correlation Mean	.73	.69	.67	.65	.54	.81
Variance	.01	.01	.01	.01	.00	.03

Note: Cen = Centralized approach; Co = Consensus-building approach

Table 10: Item Loadings for the One-Factor Model of Items Measuring Collaborative Process for Centralized Collaboration Approaches

	Factor 1
Data 1	***0.813***
Data 2	***0.830***
Data 3	***0.563***
Data 4	***0.824***
Data 5	***0.566***
Data 6	0.239
Data 7	0.285
Vision 1	***0.664***
Vision 2	***0.729***
Vision 3	***0.637***
Vision 4	***0.744***
Vision 5	***0.665***
Vision 6	***0.689***
Vision 7	***0.842***

Note: Data = Organizational data use items; Vision = Shared vision items

Table 11: Item Loadings for the Two-Factor Model of Items Measuring Collaborative Process for Consensus-Building Collaboration Approaches

	Factor 1	Factor 2
Data 1	0.061	***0.900***
Data 2	0.163	***0.787***
Data 3	0.137	***0.711***
Data 4	***0.489***	0.353
Data 5	0.174	***0.601***
Data 6	0.135	0.261
Data 7	0.075	0.194
Vision 1	***0.645***	0.179
Vision 2	***0.741***	0.123
Vision 3	***0.689***	0.078
Vision 4	***0.786***	0.146
Vision 5	***0.735***	0.086
Vision 6	***0.759***	0.119
Vision 7	***0.869***	0.083

Note: Data = Organizational data use items; Vision = Shared vision items

Table 12: Rasch Analysis Item Fit Statistics for the Governance Items

	Chi Square	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t
Gov 1	62.227	1.00	0.53	0.66	-3.72	-2.34
Gov 2	74.207	0.99	0.64	0.71	-2.66	-1.95
Gov 3	81.401	0.99	0.70	0.66	-2.35	-2.47
Gov 4	111.282	0.52	0.98	0.89	-0.11	-0.67

Note: MSQ = Mean square statistic; Gov = Governance items

Table 13: Threshold Estimates for Governance Items

	Location	Threshold 1	Threshold 2	Threshold 3	Threshold 4	Threshold 5
Gov 1	0.59	0.35	-1.73	-1.05	0.88	4.48
Gov 2	0.75	-0.76	-0.68	-0.78	1.18	4.80
Gov 3	0.87	-1.37	-0.03	-0.33	1.49	4.57
Gov 4	1.26	-1.08	0.67	-0.93	2.08	5.55

Note: Gov = Governance items

Table 14: Rasch Analysis Item Fit Statistics for the Leadership Items

	Chi Square	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t
Leader 1	92.65	0.95	0.79	0.83	-1.38	-1.15
Leader 2	115.60	0.47	1.00	0.91	0.03	-0.58
Leader 3	55.87	1.00	0.47	0.50	-4.14	-3.83
Leader 4	46.28	1.00	0.39	0.37	-4.82	-4.59
Leader 5	68.55	1.00	0.58	0.56	-2.91	-2.64
Leader 6	98.57	0.90	0.83	0.59	-0.93	-2.08
Leader 7	123.28	0.35	1.04	0.86	0.22	-0.76

Note: MSQ = Mean square statistic; Leader = Leadership items

Table 15: Threshold Estimates for Leadership Items

	Location	Threshold 1	Threshold 2	Threshold 3	Threshold 4	Threshold 5
Leader 1	1.46	-0.47	-0.95	1.50	5.78	NA
Leader 2	1.84	-1.61	-0.98	1.96	7.99	NA
Leader 3	0.98	-0.88	-1.83	1.24	5.40	NA
Leader 4	0.94	0.06	-2.04	-0.55	1.29	5.94
Leader 5	0.74	-0.00	-1.38	-1.38	1.17	5.27
Leader 6	0.62	-0.07	0.02	-2.81	0.76	5.18
Leader 7	0.41	0.92	-2.22	-0.35	3.31	NA

Note: Leader = Leadership items

Table 16: Chi Square Results for Tests One through Three of DIF in Leadership Items

	Chi Square 1	Chi Square 2	Chi Square 3
Leader 1	0.78	0.96	0.98
Leader 2	0.27	0.39	0.40
Leader 3	0.17	0.08	0.08
Leader 4	0.58	0.85	0.92
Leader 5	0.57	0.34	0.18
Leader 6	0.68	0.68	0.44
Leader 7	0.40	0.64	0.66

Note: Chi Square 1 and 3 *df*=1; Chi Square 2 *df* = 2; Leader = Leadership items
*p<.05

Table 17: Rasch Analysis Item Fit Statistics for the Roles and Responsibilities Items

	Chi Square	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t
Role 1	62.76	1.00	0.54	0.53	-3.50	-3.61
Role 2	61.74	1.00	0.53	0.53	-3.62	-3.75
Role 3	144.88	0.03	1.25	1.27	1.53	1.89
Role 4	87.94	0.97	0.77	0.78	-1.77	-1.60
Role 5	72.80	1.00	0.64	0.65	-2.11	-2.63

Note: MSQ = Mean square statistic; Role = Roles and responsibilities item

Table 18: Threshold Estimates for Roles and Responsibilities Items

	Location	Threshold 1	Threshold 2	Threshold 3	Threshold 4	Threshold 5
Role 1	1.05	0.24	-2.09	0.20	1.61	5.28
Role 2	1.09	-1.12	-1.54	0.20	1.87	6.02
Role 3	1.43	-2.35	0.72	1.51	5.84	NA
Role 4	1.27	-1.67	-0.64	0.61	2.43	5.64
Role 5	1.14	-2.63	-0.57	0.76	1.44	6.69

Note: Role = Roles and responsibilities item

Table 19: Chi Square Results for Tests One through Three of DIF in Roles and Responsibilities Items

	Chi Square 1	Chi Square 2	Chi Square 3
Role 1	0.90	0.97	0.83
Role 2	0.48	0.33	0.19
Role 3	0.44	0.28	0.16
Role 4	0.33	0.52	0.56
Role 5	0.56	0.84	0.93

Note: Chi Square 1 and 3 *df*=1; Chi Square 2 *df* = 2; Roles = Roles and responsibilities item
*p<.05

Figure 1: Wright Person-Item Map of Governance Items

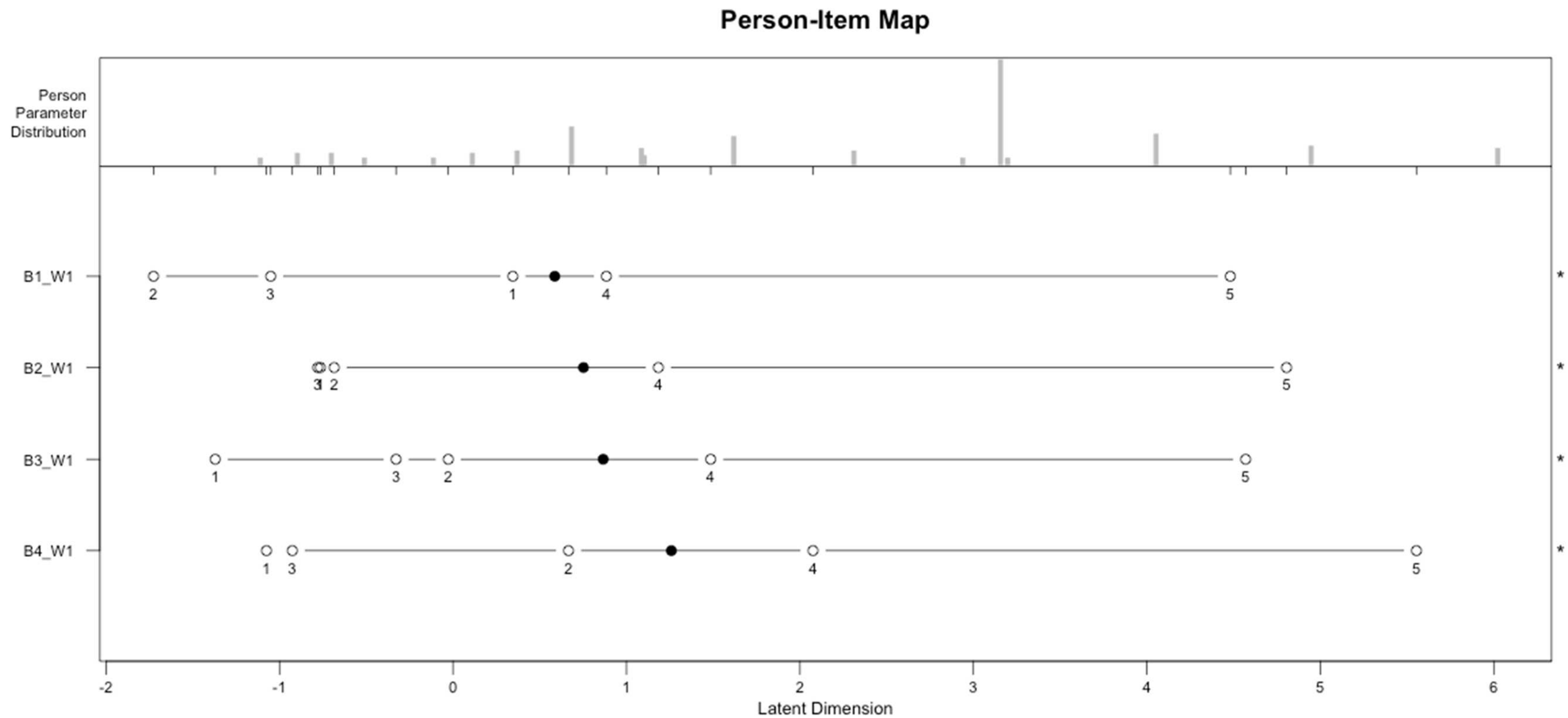


Figure 2: Wright Person-Item Map of Leadership Items

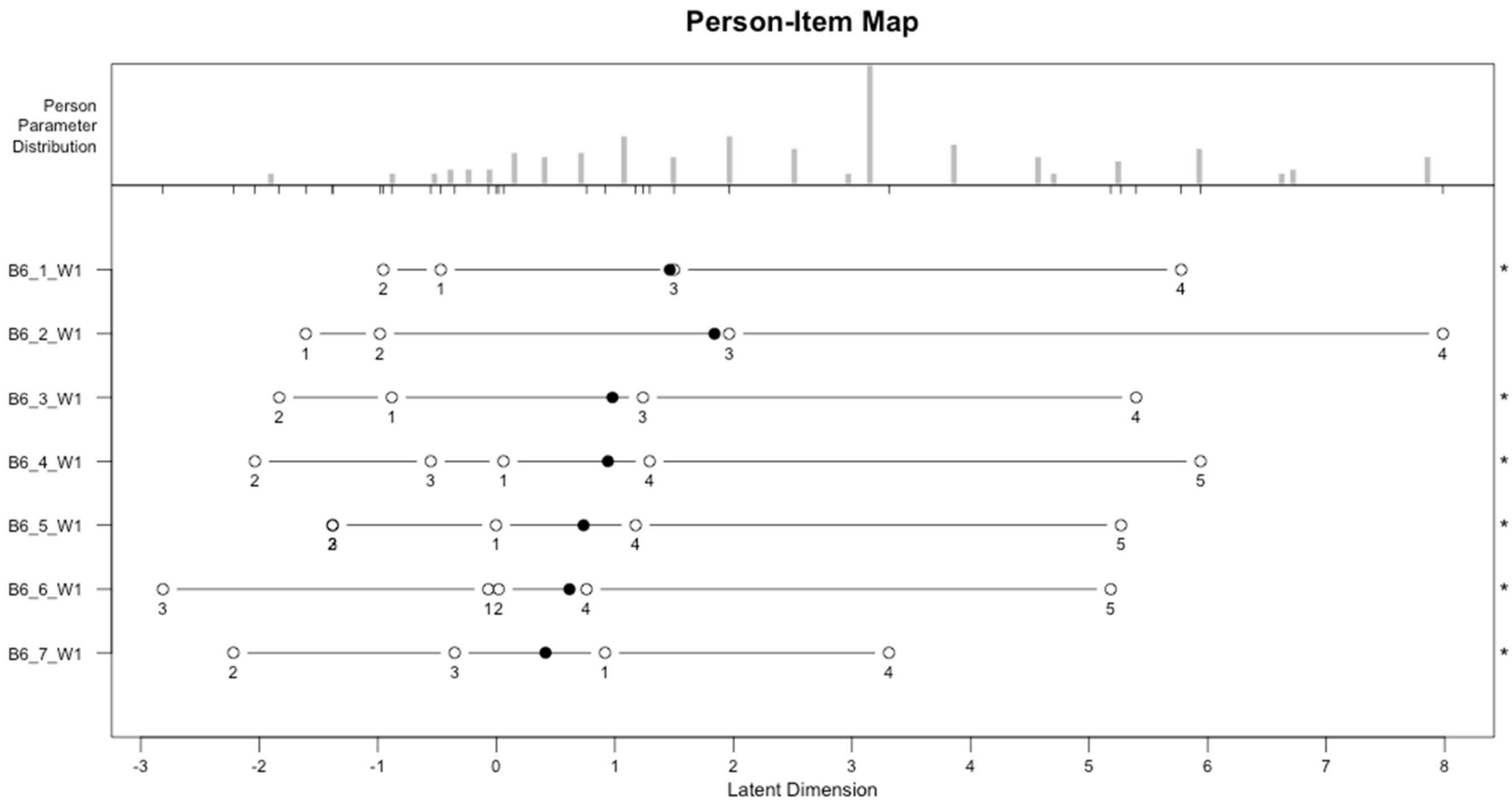
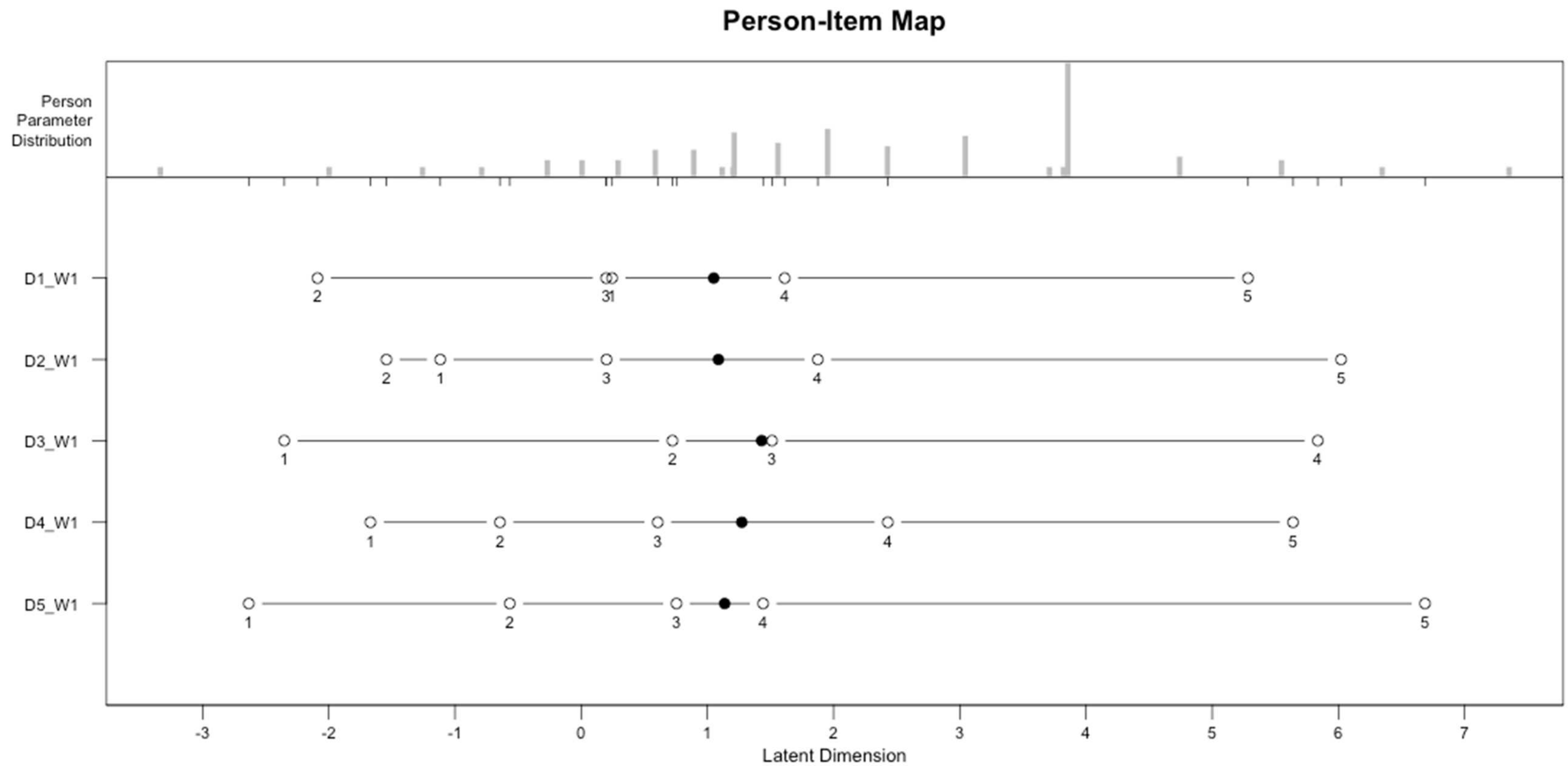


Figure 3: Wright Person-Item Map of Roles and Responsibilities Items



Appendix A: Scales from Community Collaborative Survey

STRUCTURE:

Governance Scale

1. Decision-making processes are agreed upon
2. Matters of how collaboration membership is decided are agreed upon.
3. How leadership is determined is agreed upon.
4. How resources and expenses are shared are agreed upon.

Leadership

1. The collaboration leaders are able to build consensus across the community.
2. The collaboration leaders are able to manage conflicts between different groups within the community.
3. The collaboration leaders are able to develop the capacity of the organization as a whole.
4. The collaboration leaders are able to develop the capacity of individuals within the organization.
5. The collaboration leaders are able to articulate strategies that pull on the "right" levers.
6. The collaboration leaders are able to motivate members of the organization to take collective action.
7. The collaboration leaders understand the problems facing youth in our community today.

Roles and Responsibilities

1. Collaboration member roles are defined.
2. Collaboration members understand their respective roles.
3. Roles are subject to renegotiation.
4. Responsibilities of collaboration members are formalized.
5. Responsibilities of collaboration members are given the significance of work assignments.

PROCESS:

Organizational Data Use

1. My collaboration conducts formal evaluations of its programs.
2. My collaboration uses data for feedback and continuous quality improvement.
3. My collaboration contracts with someone from the outside when we need to evaluate a program.

4. In my collaboration, we set aside time to discuss program data and what it means regarding program implementation and effectiveness.
5. My collaboration has specific staff trained in evaluation and data collection/analysis, who coordinate our evaluation efforts.
6. My collaboration lacks the resources (people, software, skills) to conduct our own program evaluations or impact assessments.
7. In my collaboration, it is important to have data to show whether programs are achieving their goals.

Developing a Shared Vision

1. Collaboration members have developed and support a common vision.
2. Collaboration members have "partnership" goals, as well as goals that are particular to their respective organizations.
3. Explicit goals have been agreed upon.
4. The collaboration's goals and theory of change takes into account what is happening in the community.
5. Residents and institutions are all aware of the goals of the collaboration.
6. The collaboration periodically reevaluates and updates its goals and theory of change.
7. The activities of the collaboration are evaluated in relation to the goals and theory of change of the collaboration.

References

Albright, J.J. & Park, H.M. (2009). *Confirmatory Factor Analysis Using Amos, LISREL, Mplus, and SAS/STAT CALIS*. Working Paper. The University Information Technology Services (UITS) Center for Statistical and Mathematical Computing, Indiana University.

Alexander, J.A., Christianson, J.B., Hearld, L.R., Hurley, R. & Scanlon, D.P. (2010). Challenges of capacity building in multisector community health alliances. *Health Education & Behavior, 37*(5), 645-664.

Antonucci, T., Akiyama, H., & Takahashi, K. (2004). Attachment and close relationships across the life span. *Attachment & Human Development*, *6*(4), 353-370.

Ansell, C., & Gash, A. (2008). Collaborative governance in theory and practice. *Journal of public administration research and theory*, *18*(4), 543-571.

Baker, E.A., Wilkerson, R., & Brennan, L.K. (2012). Identifying the role of community partnerships in creating change to support active living. *American Journal of Preventive Medicine, 43*(5 Supplement 4), S290-S299.

Benson, P. L., Leffert, N., Scales, P. C., & Blyth, D. A. (2012). Beyond the “village” rhetoric: Creating healthy communities for children and adolescents. *Applied Developmental Science*, *16*(1), 3-23.

Blum, R. W. (2003). Positive youth development: A strategy for improving adolescent health. *Handbook of applied developmental science: Promoting* *positive child, adolescent, and family development through research, policies, and programs*, *2*, 237-252.

Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch model: Fundamental measurement in the human sciences*. (2nd ed.). New York: Routledge.

Brandtstädter, J. (1998). Action perspectives on human development. In W. Damon & R.M. Lerner (Eds.), *Handbook of child psychology. Vol. 1: Theoretical models of human development* (5th ed., pp. 807–863). New York: Wiley.

Brandtstädter, J. (2006). Action perspectives on human development. In R. M. Lerner (Ed.). *Theoretical models of human development*. Volume 1 of *Handbook of Child Psychology* (6th ed.). Editors-in-chief: W. Damon & R. M. Lerner. Hoboken, NJ: Wiley.

Brewer, M.B. (2000). Research design and issues of validity. In H. Reis and C. Judd (Eds.) *Handbook of Research Methods in Social and Personality Psychology.* Cambridge: Cambridge University Press.

Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.

Bronfenbrenner, U. & Morris, P.A. (2006). The bioecological model of human development. In R.M. Lerner (Ed.) *Theoretical Models of Human Development.* Volume 1 of the *Handbook of Child Development* (6th ed.) Editors-in-chief: W. Damon & R.M. Lerner. (pp.793-828). Hoboken, NJ: Wiley.

Brown, E.C., Graham, J.W., Hawkins, J.D., Arthur, M.W., Baldwin, M.M., Oesterle, S. et al. (2009). Design and analysis of the Community Youth Development Study longitudinal cohort sample. *Evaluation Review, 33*(4), 311-334.

Brown, E.C., Hawkins, J.D., Arthur, M.W., Briney, J.S., & Abbott, R.D. (2007). Effects of *Communities That Care* on prevention services systems; Findings from the Community Youth Development Study at 1.5 years. *Prevention Science, 33*, 180-191.

Brown, E. C., Hawkins, J. D., Arthur, M. W., Briney, J. S., & Fagan, A. A. (2011). Prevention service system transformation using Communities That Care. *Journal of community psychology*, *39*(2), 183-201.

Brown, L.D., Feinberg, M.E., & Greenberg, M.T. (2010). Determinants of community coalition ability to support evidence-based programs. *Prevention Science, 11(*3), 287-297.

Brown, T. A. (2014). *Confirmatory factor analysis for applied research*. Guilford Publications.

Bush, R., Dower, J., & Mutch, A. (2002). Community capacity index: Version 2. Centre for Primary Health Care, The University of Queensland.

Butterfoss, F. D., Goodman, R. M., & Wandersman, A. (1993). Community coalitions for prevention and health promotion. *Health Education Research, 8*(3), 315-330.

Carter, P. L., & Welner, K. G. (Eds.). (2013). *Closing the opportunity gap: what America must do to give every child an even chance*. Oxford University Press.

Catalano, R.F., Berglund, M.L., Ryan, J.A.M., Lonczak, H.S., & Hawkins, J.D. (2004). Positive youth development in the United States: Research Findings on evaluations of positive youth development programs. *The ANNALS of the American Academy of Political and Social Science, 591*(1), 98-124.

Catalano, R. F., & Hawkins, J. D. (1996). The social development model: A theory of antisocial behavior. In J. D. Hawkins (Ed.), *Delinquency and crime: Current theories* (pp. 149–197). New York: Cambridge University Press.

Center for the Study and Prevention of Violence (2014, January). *Matrix of programs as identified by various federal and private agencies.* Retrieved from: http://www.blueprintsprograms.com/resources/Matrix_Criteria.pdf.

Chinman, M., Hannah, G., Wandersman, A., Ebener, P., Hunter, S. B., Imm, P., & Sheldon, J. (2005). Developing a community science research agenda for building community capacity for effective preventive interventions. *American Journal of Community Psychology*, *35*(3-4), 143-157.

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software*, *39*(8), 1.

Coalition for Evidence-Based Policy (2014). *Coalition for Evidence-Based Policy: Our Mission.* Retrieved from: http://coalition4evidence.org/

Coie, J.D., Watt, N.F., West, S.G., Hawkins, J.D., Asamow, J.R., Markman, H.J., et al. (1993). The science of prevention. *American Psychologist 48*(10), 1013-1022.

Comrey, A.L. & Lee, H.B. (1992). *A first course in factor analysis.* Hillsdale, NJ: Erlbaum.

Crawford, A.V., Green, S.B., Levy, R., Lo, W.J., Scott, L., Svetina, D., et al. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement, 70*(6), 885-901.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Crowder, K., & South, S. J. (2003). Neighborhood distress and school dropout: The variable significance of community context. *Social Science Research*, *32*(4), 659-698.

Damon, W. (1997). *The youth charter: How communities can work together to raise standards for all our children.* Free Press: New York, NY.

Downey, L.M., Ireson, C.L., Slavova, S., & McKee, G. (2008). Defining elements of success: A critical pathway of coalition development. Health Promotion Practice, 9(2), 130-139.

Dryfoos, J. G. (1994). *Full-service schools: A revolution in health and social services for children, youth, and families*. Jossey-Bass: San Francisco, CA.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087-1101.

Duncan, G. J., & Murnane, R. J. (Eds.). (2011). *Whither opportunity?: Rising inequality, schools, and children's life chances*. Russell Sage Foundation.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405-432.

Durlak, J. A., Weissberg, R. P., & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American journal of community psychology*, *45*(3-4), 294-309.

Eccles, J. S., & Gootman, J. (Eds.). (2002). *Community programs to promote youth development.* Washington, DC: National Academy Press.

Emshoff, J.G., Darnell, A.J., Darnell, D.A., Erickson, S.W., Schneider, S., & Hudgins, R. (2007). Systems change as an outcome and a process in the work of community collaboratives for health. *American Journal of Community Psychology, 29*(3-4), 255-267.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*(3), 430-457.

Fawcett, S.B., Francisco, V.T., Paine-Andrews, A., & Schultz, J.A. (2000). A model memorandum of collaboration: A proposal. *Public Health Reports,* 115(2-3), 174-179.

Feinberg, M.E., Bontempo, D.E., & Greenberg, M.T. (2008). Predictors and level of sustainability of community prevention coalitions. *American Journal of Preventive Medicine, 34*(6), 495-501.

Feinberg, M.E., Jones, D., Greenberg, M.T., Osgood, D.W., & Bontempo, D. (2010). Effects of the Communities That Care model in Pennsylvania on change in adolescent risk and problem behaviors. *Prevention Science, 11*(2), 163-171.

Fixsen, D.L., Naoom, S.F., Blase, K.A., Friedman, R.M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature.* Tampa, FL: National Implementation Research Network at University of South Florida.

Flay, B.R., Biglan, A., Boruch, R.F., Castro, F.G., Gottfredson, D., Kellam, S. et al. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6*(3), 151-175.

Foster-Fishman, P.G., Berkowitz, S.L., Lounsbury, D.W., Jacobson, S., & Allen, N.A. (2001). Building collaborative capacity in community coalitions: A review and integrative framework. *American Journal of Community Psychology, 29*(2), 241-261.

Freund, A. M., & Isaacowitz, D. M. (2014). Beyond age comparisons: A plea for the use of a modified Brunswikian approach to experimental designs in the study of adult development and aging. *Human Development*, *56*(6), 351-371.

Gibson, C., Smyth, K., Nayowith, G., & Zaff, J.F. (2013, September 9). *Nonprofit management: To get to the good, you gotta dance with the wicked.* Retrieved from: http://www.ssireview.org/blog/entry/to_get_to_the_good_you_gotta_dance_with_the_wicked

Goldsmith, S., & Eggers, W.D. (2004). *Governing by network: The new shape of the public sector.* Washington, D.C.: Brookings Institution Press.

Graham, J.W., Olchowski, A.E., & Gilreath, T.D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science, 8*, 206-213.

Gray, B. (1989). *Collaborating: Finding common ground for multiparty problems*. San Francisco, CA: Jossey-Bass Publishers.

Guerra, N. G., & Bradshaw, C. P. (2008). Linking the prevention of problem behaviors and positive youth development: Core competencies for positive youth development and risk prevention. In N. G. Guerra & C. P. Bradshaw (Eds.), *Core competencies to prevent problem behaviors and promote positive youth development. New Directions for Child and Adolescent Development, 122,* 1–17.Gutman, L., & Schoon, I. (2013). *Non-cognitive skills: evidence and intervention.* Education Endowment Fund.

Gutman, L. M., & Schoon, I. (2013). The impact of non-cognitive skills on outcomes for young people. *Education Endowment Foundation. Retrieved from: http://educationendowmentfoundation.org/uploads/pdf/Non-cognitive_skills_literature_review.pdf*.

Guttman, L . (1954). Some necessary conditions for common factor analysis. *Psychometrika, 19*, 149– 162.

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development*, *76*(5), 949-967.

Harlem Children's Zone (2012). *Harlem Children's Zone Quick Facts 2012*. Retrieved from http://www.hcz.org

Hawkins, J.D., Catalano, R.F., & Arthur, M.W. (2002). Promoting science-based prevention in communities. *Addictive Behaviors, 27*(6), 951-976.

Hawkins, J. D., Oesterle, S., Brown, E. C., Arthur, M. W., Abbott, R. D., Fagan, A. A., & Catalano, R. F. (2009). Results of a type 2 translational research trial to prevent adolescent drug use and delinquency: A test of

Communities That Care. *Archives of Pediatrics & Adolescent Medicine*, *163*(9), 789-798.

Hawkins, J. D., Oesterle, S., Brown, E. C., Monahan, K. C., Abbott, R. D., Arthur, M. W., & Catalano, R. F. (2012). Sustained decreases in risk exposure and youth problem behaviors after installation of the Communities That Care prevention system in a randomized trial. *Archives of Pediatrics & Adolescent Medicine*, *166*(2), 141-148.

Heckhausen, J., Wrosch, C., & Schulz, R. (2010). A motivational theory of life-span development. *Psychological Review*, *117*(1), 32-60.

Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies*, *65*(2), 261-294.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

Hollander, J. B., Pallagst, K., Schwarz, T., & Popper, F. J. (2009). Planning shrinking cities. *Progress in Planning*, *72*(4), 223-232.

Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 20*, 179-185.

Houston, D., & Ong, P. (2013). Arts accessibility to major museums and cultural/ethnic institutions in Los Angeles: can school tours overcome neighborhood disparities? *Environment and Planning A*, *45*(3), 728-748.

Howell, J. C., & Egley, A. (2005). Moving risk factors into developmental theories of gang membership. *Youth Violence and Juvenile Justice*, *3*(4), 334-354.

Himmelman, A.T. (1992). *Communities working collaboratively for a change.* Chris Huxam, Editor. London: Sage Publications, Limited. (Revised 1996).

Hu, L.T. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Huxham, C. (2003). Theorizing collaboration practice. *Public Management Review, 5*(3), 401-423.

Jenson, J.M., Alter, C.F., Nicotera, N., Anthony, E.K., & Forrest-Bank, S.S. (2013). *Risk, resilience, and positive youth development: Developing effective community programs for at-risk youth: Lessons from the Denver Bridge Project.* New York: Oxford University Press.

Jimerson, S., Egeland, B., Sroufe, L. A., & Carlson, B. (2000). A prospective longitudinal study of high school dropouts examining multiple predictors across development. *Journal of School Psychology*, *38*(6), 525-549.

Jung, S., & Lee, S. (2011). Exploratory factor analysis for small samples. *Behavior research methods*, *43*(3), 701-709.

Kadushin, C., Lindholm, M., Ryan, D., Brodsky, A., & Saxe, L. (2005). Why is it so difficult to form effective community coalitions? *City & Community, 4*(3), 255-275.

Kaiser, H.F . (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*(1), 141– 151.

Kaye, G. (1993). *Diagnosing your coalition: Risk factors for participation*. Brooklyn, NY: Community Development Consultants.

Keast, R. & Mandell, M. (2012). The collaborative push: Moving beyond rhetoric and gaining evidence. *Journal of Management and Governance,* 1-20.

Kegler, M.C. & Swan, D.W. (2012). Advancing coalition theory: The effect of coalition factors on community capacity mediated by member engagement. *Health Education Research, 27*(4), 572-584

Kellam, S.G. & Langevin, D.J. (2003). A framework for understanding "evidence" in prevention research and programs. *Prevention Science, 4*(3), 137-153.

Kim, J. (2012). Exploring the relationship between state financial aid policy and postsecondary enrollment choices: A focus on income and race differences. *Research in Higher Education*, *53*(2), 123-151.

Kubisch, A.C., Auspos, P., Brown, P., & Dewar, T. (2010, September 10). Voices from the field III: Lessons and challenges from the last two decades of community change efforts. Retrieved from:

http://www.aspeninstitute.org/publications/voices-field-iii-lessons-challenges-two-decades-community-change-efforts.

Ladd, H. F., Muschkin, C. G., & Dodge, K. A. (2014). From birth to school: Early childhood initiatives and third-grade outcomes in North Carolina. *Journal of Policy Analysis and Management*, *33*(1), 162-187.

Lasker, R.D. & Weiss, E.S. (2003). Broadening participation in community problem solving: A multidisciplinary model to support collaborative practice and research. *Journal of Urban Health,* 80(1), 14-47.

Lasker, R.D., Weiss, E.S., & Miller, R. (2001). Partnership synergy: A practical framework for studying and strengthening the collaborative advantage. *The Millbank Quarterly, 79*(2), 179-205.

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of educational research*, *76*(2), 275-313.

Lawson, H.A., Claiborne, N., Hardiman, E., Austin, S., & Surko, M. (2007). Deriving theories of change from successful community development partnerships for youths: Implications for school improvement. *American Journal of Education, 114*(1), 1-40.

Lerner, R. M. (1984). *On the nature of human plasticity*. New York: Cambridge University Press.

Lerner, R. M. (2002). *Concepts and theories of human development* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Lerner, R. M. (2012). Essay review: Developmental science: Past, present, and future. *International Journal of Developmental Science, 6,* 29-36.

Lerner, R. M., Agans, J. P., DeSouza, L., & Hershberg, R. (2014). Developmental science in 2025: A review! *Research in Human Development, 11*(4), 255-272.

Lerner, R. M., & Callina, K. S. (2014). Relational developmental systems theories and the ecological validity of experimental designs. *Human Development, 56*(6), 372-380.

Lerner, R. M., Lerner, J. V., Bowers, E., & Geldhof, G. J. (2015). Positive youth development and relational developmental systems. In W. F. Overton & P. C. Molenaar (Eds.), *Theory and Method*. Volume 1 of the *Handbook of Child Psychology and Developmental Science* (7th ed.). Editor-in-chief: R. M. Lerner. (pp. 607-651). Hoboken, NJ: Wiley.

Lerner, R.M., Lerner, J.V., & Zaff, J.F. (in press). Evaluating programs aimed at promoting positive youth development: A relational development systems-based view. *Applied Developmental Science.*

Lerner, R.M. & Overton, W.F. (2008). Exemplifying the integrations of the relational developmental system: Synthesizing theory, research, and application to promote positive development and social justice. *Journal of Adolescent Research, 23*(3), 245-255.

Leventhal, T., Dupéré, V., & Brooks-Gunn, J. (2009). Neighborhood influences on adolescent development: Contextual influences on adolescent development. In R.M. Lerner & L. Steinberg (Eds.) *Handbook of Adolescent Psychology (Volume III)*. New York: John Wiley and Sons.

Leventhal, T., Dupéré, V., & Shuey, E.A. (2015). Children in neighborhoods. In M.H. Bornstein & T. Leventhal (Eds.), *Ecological Settings and Processes*. Volume 4 of the *Handbook of Child Psychology and Developmental Science* (7th ed.). Editor-in-chief: R. M. Lerner. (pp. 493-533). Hoboken, NJ: Wiley.

Lich, K. H., Ginexi, E. M., Osgood, N. D., & Mabry, P. L. (2013). A call to address complexity in prevention science research. *Prevention Science*, *14*(3), 279-289.

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean. *Rasch Measurement Transactions*, *16*(2), 878.

Linacre, J. J. (2003). Rasch power analyses: Size vs. significance: Infit and outfit mean-square and standardized Chi square fit statistic. *Rasch Measurement Transactions, 17*(1), 918.

Linacre, J. M. (2004). Rasch model estimation: Further topics. *Journal of Applied Measurement*, *5*(1), 95-110.

Lochner, K.A., Kawachi, I., Brennan, R.T., & Buka, S.L. (2003). Social capital and neighborhood mortality rates in Chicago. *Social Science & Medicine, 56*(8), 1797-1805.

Luke, D.A., Harris, J.K., Shelton, S., Allen, P., Carothers, B.J., & Mueller, N.B. (2010). Systems analysis of collaboration in 5 national tobacco control networks. *American Journal of Public Health, 100*(7), 1290-1297.

Luthar, S.S., & Barkin, S.H. (2012). Are affluent youth truly “at risk”? Vulnerability and resilience across three diverse samples. *Development and Psychopathology, 24*(2), 429-449. doi:10.1017/S0954579412000089.

Luthar, S. S., Barkin, S. H., & Crossman, E. J. (2013). "I can, therefore I must": Fragility in the upper-middle classes. *Development and Psychopathology, 25th Anniversary Special Issue, 25*, 1529-1549.

Mancini, J. A., & Marek, L. I. (2004). Sustaining community-based programs for families: conceptualization and measurement. *Family Relations*, *53*(4), 339-347.

Marek, L.I., Brock, D.P., & Salva, J. (2015). Evaluative collaboration for effectiveness: Conceptualization and measurement. *American Journal of Evalution, 36*(1), 67-85.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149-174.

McCall, R. B., & Green, B. L. (2004). Beyond the methodological gold standard of behavioral research: Considerations for practice and policy. *SRCD Social Policy Reports, 18*, 1-19.

McKnight, J.L. & Kretzmann, J.P. (1993). Mapping community capacity. *Michigan State University Community & Economic Development Program Community News,* pp.1-4.

Molenaar, P. C., & Nesselroade, J. R. (2012). Merging the idiographic filter with dynamic factor analysis to model process. *Applied Developmental Science*, *16*(4), 210-219.

Molenaar, P.C.M. & Nesselroade, J.R. (2015). Systems for developmental research. In W. F. Overton & P. C. Molenaar (Eds.), *Theory and Method*. Volume 1 of the *Handbook of Child Psychology and Developmental Science* (7th ed.). Editor-in-chief: R. M. Lerner. (pp. 652-682). Hoboken, NJ: Wiley.

Morenoff, J. D., Sampson, R. J., & Raudenbush, S. W. (2001). Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. *Criminology*, *39*(3), 517-558.

National Institute of Justice (2014). *About CrimeSolutions.gov: Program Review and Rating from Start to Finish*. Retrieved from: http://www.crimesolutions.gov/about_starttofinish.aspx.

NREPP (2014, July 10). *SAMHSA's National Registry of Evidence-based Programs and Practices.* Retrieved from http://www.nrepp.samhsa.gov/ReviewQOR.aspx.

Oesterle, S., Hawkins, J. D., Fagan, A. A., Abbott, R. D., & Catalano, R. F. (2010). Testing the universality of the effects of the Communities That

Care prevention system for preventing adolescent drug use and delinquency. *Prevention Science*, *11*(4), 411-423.

OJJDP (2014). *Model Programs Guide.* Retrieved from: http://www.ojjdp.gov/mpg/Home/About

Olsen, L., Aisner, D., & McGinnis, J. M. (2007). *Institute of Medicine Roundtable on Evidence-Based Medicine: The Learning Healthcare System, Workshop Summary.* Washington DC, National Academies Press.

Overton, W. F. (2013). Relationism and relational developmental systems: A paradigm for developmental science in the post-Cartesian era. In J. B., Benson (Series Ed.) & R. M., Lerner & J. B. Benson (Vol. Ed.), *Advances in Child Development and Behavior: Volume 44. Embodiment and Epigenesis: Theoretical and Methodological Issues in Understanding the Role of Biology within the Relational Developmental System* (1st ed., pp. 21-64). Waltham, MA: Elsevier.

Overton, W.F. (2015). Process and relational-developmental systems. In R.M. Lerner (Series Ed.) & W.F. Overton & P.C. Molenaar (Vol. Ed.), *Theory and Method: Volume 1 of the Handbook of Child Psychology and Developmental Science* (7th ed., pp. 9-62). Hoboken, NJ: Wiley.

Peck, S. C., Roeser, R. W., Zarrett, N., & Eccles, J. S. (2008). Exploring the roles of extracurricular activity quantity and quality in the educational resilience of vulnerable adolescents: Variable ‐ and pattern ‐ centered approaches. *Journal of Social Issues*, *64*(1), 135-156.

Perna, L. W., & Titus, M. A. (2005). The relationship between parental involvement as social capital and college enrollment: An examination of racial/ethnic group differences. *The Journal of Higher Education*, *76*(5), 485-518.

Ponocny, I. (2001). Nonparametric goodness-of-fit tests for the Rasch model. *Psychometrika, 66*, 437-460.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research.

Reisner, E. R., White, R. N., Russell, C. A., & Birmingham, J. (2004). Building quality, scale, and effectiveness in after-school programs: Summary report of the TASC evaluation. Washington, DC: Policy Studies Associates, Inc. Prepared for the After School Corporation. Retrieved from imap.www.tascorp.org/sites/default/files/building_quality_school_effectiveness_afterschool.pdf.

Roussos, S.T. & Fawcett, S.T. (2000). A review of collaborative partnerships as a strategy for improving community health. *Annual Review of Public Health, 21*, 369-402.

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys.* New York: John Wiley and Sons.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, *277*(5328), 918-924.

Scales, P. C., Benson, P. L., & Mannes, M. (2006). The contribution to adolescent well-being made by nonfamily adults: An examination of developmental assets as contexts and processes. *Journal of Community Psychology*, *34*(4), 401-413.

Schorr, L.B. (2012). Broader evidence for bigger impact. *Stamford Social Innovation Review,* Fall 2012, 49-55.

Seldon, W., Jolin, M., & Schmitz, P. (2012, February 6). Needle-moving community collaboratives: A promising approach to addressing America's biggest challenges. New York, NY: Bridgespan, Inc. Retrieved from: http://www.bridgespan.org/Publications-and-Tools/Revitalizing-Communities/Community-Collaboratives/Needle-Moving-Community-Collaborative-s-A-Promisin.aspx#.VPBcALPF_9t.

Silverstein, M., Lamberto, J., DePeau, K., & Grossman, D. C. (2008). "You get what you get": Unexpected findings about low-income parents' negative experiences with community resources. *Pediatrics*, *122*(6), e1141-e1148.

Small, S., & Supple, A. (2001). Communities as systems: Is a community more than the sum of its parts? In A. Booth and A.C. Crouter (Eds.), *Does it take a village? Community effects on children, adolescents, and families,* (pp. 161-174). New York: Psychology Press.

Smyth, K.F., & Schorr, L. B. (2009). A lot to lose: A call to rethink what constitutes "evidence" in finding social interventions that work. *Malcolm Wiener Center for Social Policy Working Paper Series.*

Spencer, M. B. (2006). Phenomenology and ecological systems theory: Development of diverse groups. In W. Damon and R. Lerner (Eds.), (6th Edition) *Handbook of Child Psychology,* vol. 1, (Chap. 15, Theory Volume) (pp. 829-893). New York: Wiley Publishers.

Spencer, M. B. (2013). Cultural cognition and social cognition as identity correlates of Black children's personal-social development. *Beginnings: The Art and Science of Planning Psychotherapy*, 215.

Spencer, M. B., & Swanson, D. P. (2013). Opportunities and challenges to the development of healthy children and youth living in diverse communities. *Development and Psychopathology*, *25*(4, 2), 1551-1566.

Spoth, R., & Greenberg, M. (2011). Impact challenges in community science-with-practice: Lessons from PROSPER on transformative practitioner-scientist partnerships and prevention infrastructure development. *American Journal of Community Psychology*, *48*(1-2), 106-119.

Spoth, R., Greenberg, M., Bierman, K., & Redmond, C. (2004). PROSPER community-university partnership model for public education systems: Capacity-building for evidence-based, competence-building prevention. *Prevention Science, 5*(1), 31-39.

Swanson, C. B. (2009). Cities in crisis: Closing the graduation gap. Educational and economic conditions in America's largest cities. *Editorial Projects in Education*.

Syed, M., Azmitia, M., & Cooper, C.R. (2011). Identity and academic success among underrepresented ethnic minorities: An interdisciplinary review and integration. *Journal of Social Issues, 67*(3), 442-468.

Theokas, C., Almerigi, J. B., Lerner, R. M., Dowling, E. M., Benson, P. L., Scales, P. C., & von Eye, A. (2005). Conceptualizing and modeling individual and ecological asset components of thriving in early adolescence. *Journal of Early Adolescence*, *25*(1), 113-143.

Trochim, W.M.K. (2006). *The regression-discontinuity design.* Retrieved from http://www.socialresearchmethods.net/kb/quasird.php.

United States Department of Education. (2013, August 9). *Programs: Promise Neighborhoods.* Retrieved from: http://www2.ed.gov/programs/promiseneighborhoods/index.html.

United States Department of Education. (2014). *Institute of education sciences: What works clearinghouse.* Retrieved from: http://ies.ed.gov/ncee/wwc/.

Urban, J. B., Hargraves, M., & Trochim, W. M. (2014). Evolutionary Evaluation: Implications for evaluators, researchers, practitioners, funders and the evidence-based program mandate. *Evaluation and program planning*, *45*, 127-139.

Urban, J. B., Osgood, N. D., & Mabry, P. L. (2011). Developmental systems science: Exploring the application of systems science methods to developmental science questions. *Research in Human Development*, *8*(1), 1-25.

Velicer, W.F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41,* 321-327.

Velicer, W. F., & Jackson, D. N. (1990). Component Analysis Versus Common Factor-Analysis – Some Further Observations. *Multivariate Behavioral Research, 25(1),* 97-114.

Verhelst, N. (2008). An efficient MCMC algorithm to sample binary matrices with mixed marginal. *Psychometrika, 73,* 705-728.

Vlahov, D., Freudenberg, N., Proietti, F., Ompad, D., Quinn, A., Nandi, V., & Galea, S. (2007). Urban as a determinant of health. *Journal of Urban Health*,*84*(1), 16-26.

Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, S., et al. (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.

Warne, R. T., & Larsen, R. (2014). Evaluating a proposed modification of the Guttman rule for determining the number of factors in an exploratory factor analysis. *Psychological Test and Assessment Modeling*, *56*(1), 104-123.

Wilson, W.J. (1987). *The truly disadvantaged: The inner city, the underclass, and public policy.* Chicago: University of Chicago Press.

Wolff, T. (2001). A practitioners guide to successful coalitions. *American Journal of Community Psychology, 29*(2), 173-191.

Wright, B.D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement, 14*(2), 97-116.

Zaff, J.F. (2011). A cease and desist order for school reform: It is time for educational transformation. *Applied Developmental Science, 15*, 1-7.

Zaff, J.F., Donlan, A., Pufall Jones, E., Lin, E.S. (in press). Supportive developmental systems for children and youth: A theoretical framework for comprehensive community initiatives. *Journal of Applied Developmental Psychology.*

Zaff, J. F. & Smerdon, B. (2009). Putting children front and center: Building coordinated social policy for America's children. *Applied Developmental Science, 13*(3), 105-118.

Zakocs, R.C. & Edwards, E.M. (2006). What explains community coalition effectiveness?: A review of the literature. *American Journal of Preventive Medicine, 30*(4), 351-361.

Zakocs, R. C., & Guckenburg, S. (2007). What coalition factors foster community capacity? Lessons learned from the Fighting Back Initiative. *Health Education & Behavior*, *34*(2), 354-375.