Clemson University

TigerPrints

All Dissertations   **Dissertations**

8-2017

The Role of Implementation in Leisure, Experiential Education, and Recreation

Ryan J. Gagnon
Clemson University, rjgagno@g.clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/all_dissertations

Recommended Citation

THE ROLE OF IMPLEMENTATION IN LEISURE, EXPERIENTIAL EDUCATION,
AND RECREATION

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Parks, Recreation and Tourism Management

by
Ryan J. Gagnon
August 2017

Accepted by:
Dr. Barry A. Garst, Committee Chair
Dr. Denise Anderson
Dr. Robert Bixler
Dr. Tracy Maineri
Dr. Dewayne Moore

ABSTRACT

Within Leisure, Experiential Education, and Recreation (LER) research, there is a shift to evidence-based practices (EBP) in program design and assessment, reflecting the transition to EBPs within the prevention, health, and social sciences. However, one area that still lacks conceptual development and application within LER relates to how programs are implemented. This study contributes both to LER and the broader social sciences by examining the multiple dimensions of implementation within residential summer camp. First, this study reviews the macro level factors that contribute to or inhibit implementation quality. Second, this study examines the factors that promote implementation quality at the facilitator level through the production of a conceptual framework and corresponding characteristic, trait, and behavior measurement recommendations. Third, this study introduces Situational Judgement Testing as a method to predict implementation quality through a content analysis of Subject Matter Experts' responses to scenarios reflecting the LER programming context. Fourth, this study evaluates a Multi-Level Model to explore relationships among facilitator traits, behaviors, and characteristics with implementation quality, related sub-domains of implementation quality, and parent perceptions of program outcomes. The study findings: (a) demonstrate partial support for contemporary implementation research findings in an underexamined context, LER, (b) provide researchers and practitioners with actionable guidelines for future implementation investigations, and (c) establish a platform for research regarding the necessity of implementation assessment within the LER sciences.

DEDICATION

To Mom and Rich, for showing me how to grind. To Mead High School, for telling me I couldn't do it. To Spokane Falls Community College, for opening the door. To Matt C., for showing me the way. To Jonathan S., for second chances. To Matt B., for your guidance. To Barry G., for your patience. To Lisa G., for your heart.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

Table of Contents (Continued)

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: PROJECT BACKGROUND

Within the context of leisure, experiential education, and recreation (LER), well-designed and implemented programs and events lead to positive economic (Oh, Richardson, & Lacher, 2016), social (Arai & Pedlar, 2003), and emotional development for the communities and constituents they are intended to serve (Kleiber, Walker, & Mannell, 2011). Correspondingly there has been a shift from the "benefits-based" movement within LER to one that is more evidence-based (Kaczynski & Henderson, 2007) in the form of randomized control trials, quasi-experimental designs, and rigorous utilization of theory to design LER programs and services (Witt & Caldwell, 2010). Furthermore, there is a notable transition in LER assessment and research towards utilization of the best available evidence and methods to justify program selection, resource allocation, and program retention (Berk & McGivern-Moon, 2016; Browne, Garst, & Bialeschki, 2011; Bruening, Clark, & Mudrick, 2015; Crompton, 2016). While this orientation toward evidence-based decision making is reflected in research relating to the outcomes of these programs and events, a notable gap exists in terms of our understanding of implementation quality (Gagnon & Bumpus, 2016; Mainieri & Anderson, 2015a). More specifically, within LER there is a surprising lack of *implementation research*, the science of taking a program to effect, to include the identification of factors that may contribute to the design, delivery, and/or selection of a program (Eccles & Mittman, 2006; Sloboda, Dusenbury, & Petras, 2014). This lack of research is not only germane to LER, rather several exhaustive reviews of implementation research illustrate that implementation quality is frequently ignored or

1

undervalued when presenting the results of studies promoting the efficacy (or lack thereof) of programs (Berkel, Mauricio, Schoenfelder, & Sandler, 2011; Century & Cassata, 2016; Dane & Schnieder, 1998; Durlak & DuPre, 2008). There are many potential consequences of the lack of implementation-science focus within LER studies, the primary consequences being higher rates of type I errors (stating that a program was effective when it is not properly implemented) and/or type II errors (inferring that a program was ineffective when it was not properly implemented). Furthermore, the assessment of program implementation is critical for evaluating the internal and external validity of programs (Berkel et al., 2011; Durlak & DuPre, 2008). Implementation assessment can highlight reasons for programmatic success or failure, how a program design may be improved, and how a program may be replicated in settings outside of the program designer's scope (Dane & Schneider, 1998; Durlak & DuPre, 2008). Despite ample evidence that the assessment of a program's implementation is a critical component of the program delivery and improvement process (Fagan, Hanson, Hawkins, & Arthur, 2008; Little, Sussman, Sun, & Rohrbach, 2013), implementation assessments still rarely take place (Sloboda et al., 2014).

The study of implementation in the context of LER is inherently multidimensional. More specifically, a program can be influenced by multiple factors, including those at the organizational level (e.g., resources and support), the community level (e.g., community support and trust of a program), the program level (e.g., program design and fit), and at the facilitator level (e.g., enthusiasm and training) (Gagnon, Franz, Garst, & Bumpus, 2015). Additionally, within each of these levels there are numerous

factors that have been demonstrated to help or harm, implementation quality (Berkel et al., 2011; Durlak & DuPre, 2008). More simply, the study of factors that influence implementation quality is both complex and daunting (Durlak, 2010). Thus, to continue the preliminary investigations of implementation quality within LER contexts (e.g., Mainieri & Anderson, 2015a; Morgan, Sibthorp, & Browne, 2016), this project examined implementation through the lens of the residential summer camp program facilitator, who ultimately bears the responsibility to ensure that a program is implemented as designed (Wanless, Rimm-Kaufman, Abry, Larsen, Patton, 2015b). Specifically, *the purpose of this dissertation was to explore how a facilitator's traits, characteristics, skills, and behaviors relate to implementation quality through a three-paper dissertation format.* The sections below provide a brief description of the three papers (i.e., chapters) including their research purposes, anticipated outcomes, and targeted journals for later publication.

Chapter Two: A Conceptual Framework Exploring Facilitators Characteristics and Influences on Program Implementation: Measurement Recommendations and Challenges

The field of implementation research is growing practically, conceptually, and technically (Hansen, 2014). More specifically, there is an increase in evidence of the importance and positive effect of high quality implementation on program outcomes (Durlak & DuPre, 2008). Despite this growth in evidence, there remains a gap between our understanding of how programs are implemented outside of the program designer's control (Berkel et al., 2011) and which factors contribute most to implementation quality.

Specifically, there are challenges in operationalization and measurement of the factors that contribute to implementation quality. Given the relatively low level of LER-focused implementation research (Maineri & Anderson, 2015a; Tucker & Rheingold, 2010) and the applied nature of LER, this chapter provides a conceptual framework to guide future LER implementation research at the facilitator level. To develop this framework, the chapter: (1) explores the macro-level factors that influence implementation quality, (2) examines implementation domains influenced at the facilitator level, (3) investigates micro-level facilitator characteristics, traits, and actions that influence implementation quality, and (4) provides recommendations for measurement and operationalization's of these factors, components, and traits based upon prior research.

This chapter discusses macro- and micro-level factors that can influence implementation quality. First, the review of macro level processes provides the reader with the broader context when considering an implementation assessment as part program design, delivery, and/or evaluation. In this vein, the targeted journal is the *Journal of Youth Development.* Second, the review of micro-level factors, conceptual framework, and measurement recommendations that influence implementation quality at the facilitator level provides the reader with clear guidelines to explore LER program and service implementation quality, thus the target journal is *Leisure Sciences*.

Chapter Three: Developing an Assessment of Facilitator Influences on Program Implementation

Much of LER research is orientated towards understanding human growth and development resulting from attending a program or series of programs, but not

necessarily "how" these programs achieve their desired outcomes. This chapter enhances understanding of this "how" within LER by investigating the implementation process and its corresponding assessment through an exploration of one level thought to contribute to program implementation: the program facilitator responsible for delivering the program. To this end, the chapter: (1) explores the elements theorized to contribute to facilitator implementation quality at a general level, (2) reviews how these elements have been previously assessed, (3) introduces situational judgement testing (SJT) as a new method for assessing implementation quality, and (4) describes how this method was developed within an LER context. Thus, the parallel goals of this chapter are: (A) to introduce SJTs to the field of LER and (B) to develop SJTs that capture the facilitator influence on implementation quality. These SJTs were develop by integrating prior implementation research from the broader social sciences with the results of semi-structured interviews conducted with subject matter experts responsible for youth-centered program design and delivery. While discussed in more detail later in this dissertation, to guide the reader, *SJTs* present a respondent "with a brief scenario and then ask him or her to select the best choice or indicate what he or she would do" in the given scenario (Barrett, Doverspike, & Young, 2010, p. 447). This chapter utilizes data from experienced experiential educators from the residential summer camp and challenge course industry and has implications for these professions, thus the target journal for this chapter is the *Journal of Experiential Education*.

Chapter Four: A Multi-Level Model Exploring the Relationship among Perceived Implementation Behaviors, Attitudes, and Outcomes

As explored in chapter two, there are multiple challenges within implementation research including poor construct validity, inconsistent or outdated measurement techniques, and "siloing" of highly correlated constructs (Mowbray, Holter, Teague, & Bybee, 2003). Furthermore, the factors that influence facilitator implementation quality are often inconsistent between studies. For instance, facilitator experience has been found to have a null (Pas, Waasdrop, & Bradshaw, 2015), negative (Rohrbach, Graham, & Hansen, 1993), or positive effect (Dusenbury, Branningan, Falco, & Hansen, 2005) on implementation quality, and is measured differently between studies. The contradictory (and perhaps unsurprising) findings between studies only further illustrate the opportunities within implementation research and the need for robust investigations utilizing consistent operationalization of terminology to understand the "magic" that occurs during program implementation.

Beyond issues of construct validity, inconsistent effects are also partly due to outdated statistical techniques, where data are treated at a single level. More specifically, the analytic techniques and approaches to understanding implementation should better reflect "the organizational levels at which the data are collected. Traditional analytic methods require that all data be aggregated or otherwise configured to a single specified 'unit of analysis,' the program, in most fidelity [implementation] research to date" (Mowbray et al., 2003, p. 334). In this chapter, a multi-level model is utilized to explore relationships among the facilitator, residential camp outcomes, and implementation

quality as this approach affords "the best means to represent the data structures that typically arise when multiple fidelity indicators are used to characterize the manner in which providers deliver an intervention to recipients" (Zvoch, 2012, p. 549).

At a simpler level, the goal of this chapter is to better understand implementation science in the context of LER; as such, this chapter: (1) examines the conceptual foundations of implementation assessment, (2) describes how implementation is evaluated at the facilitator level, (3) introduces a new method for the assessment of implementation, and (4) shares results of an experiment exploring the usage of this alternative method in comparison and/or in combination with more standard measures of implementation quality. This chapter utilizes a latent multi-level model to explore factors that contribute to implementation quality at the facilitator level within a common youth experience, residential summer camp. As such the target journal for this chapter is the *Journal of Youth and Adolescence*.

Summary

The purpose of this dissertation is to explore how a facilitator's traits, characteristics, skills, and behaviors relate to implementation quality through a three-paper dissertation format. In the chapters that follow a conceptual framework of the macro- and micro-level factors that influence implementation is provided (Chapter 2); a novel method for the prediction of implementation quality is introduced (Chapter 3); a multi-level model exploring the relationships among the facilitator, program outcomes, and implementation is shared (Chapter 4); and a summarization of results, future directions, and implications of this dissertation are provided (Chapter 5).

CHAPTER TWO: A CONCEPTUAL FRAMEWORK EXPLORING FACILITATORS' CHARACTERISTICS AND INFLUENCES ON PROGRAM IMPLEMENTATION: MEASUREMENT RECOMMENDATIONS AND CHALLENGES

Implementation research is the science of taking a program to effect, that is identifying the factors that may contribute to the design, delivery, and/or outcomes of a program (Eccles & Mittman, 2006; Sloboda et al., 2014). Research in this area has grown exponentially since the 1990's with the emergence of a journal specifically dedicated to the study of implementation in 2006 (Eccles & Mittman, 2006). Moreover, the systematic reviews of implementation research provided by Dane and Schneider (1998), Durlak and DuPre (2008), and Berkel et al. (2011) in both scope and applicability to the broader social science support this uptick in the importance of implementation research to the social sciences. As indicated by Sloboda et al. (2014) in Figure 1, implementation research continues to grow rapidly. However, little investigation into the role of implementation within the context of leisure, experiential, and recreation (LER) program settings has been conducted (Mainieri & Anderson, 2015a). This opportunity for research in this area is highlighted by the work of Gillis, Gass, and Russell (2008) and Tucker and Rheingold (2010), who demonstrated the potential benefits of implementation science to LER and correspondingly, the surprising lack of implementation-focused research given the applied nature of LER.
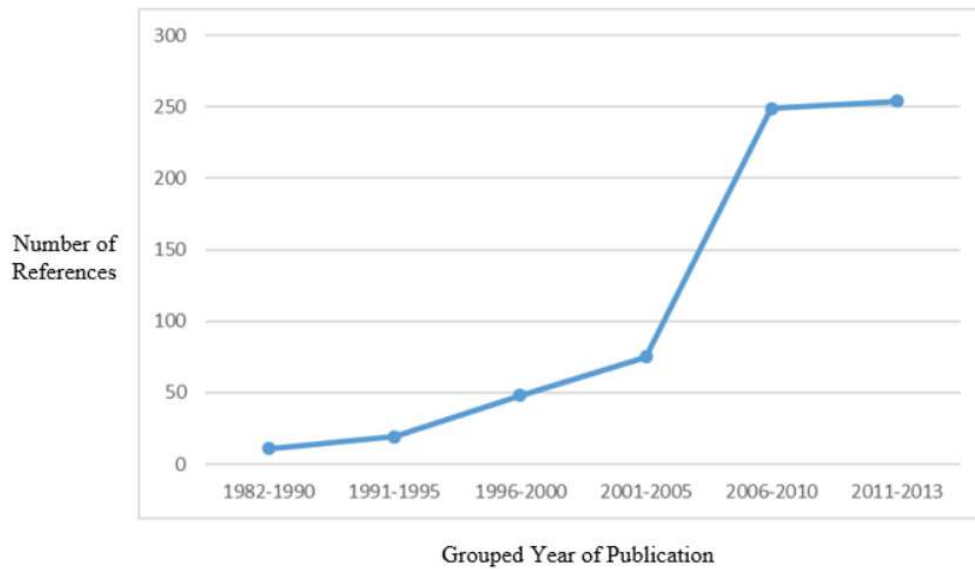
Figure 1. Number of references in Pubmed with term "implementation fidelity" in the title or abstract by grouped year of publication, adapted from Sloboda et al. (2014).

Responding to the necessity of implementation research within program assessments, Durlak (2015) shared that without it we cannot understand "if a program has been put to an adequate test. It may fail not because the intervention lacks value, but because the intervention was not implemented at a sufficiently high enough level to produce its effects" (p. 1124). In other words, without implementation quality research, we risk committing a Type II error when we assume that a lack of program effect was due to an ineffective design. Correspondingly without the support of implementation assessment guiding statements of program outcomes and efficacy, the risk of Type I error also increases (e.g., stating a program is effective when there isn't evidence that the program was delivered as designed). Expounding on this challenge and potential for Type II error within LER, Morgan et al., (2016) shared that as a field we must move beyond basic analyses of outcome achievement to a more in depth look on both why and how a

program achieved the desired outcomes. Beyond evidence of why a program may have failed, implementation assessment also helps to uncover why programs may succeed. Without the information and data relating to implementation quality it is not possible for an implementation evaluator to understand how the program did (and did not) influence the desired outcomes (Williford, Sanger-Wolcott, Vick-Whittaker, & Locasale-Crouch, 2015). In other words, an understanding that a program was delivered as designed is useful information for program stakeholders, but the additional information regarding why programs were not delivered is likely more useful for program designers and those charged with delivering future iterations of the program (Hill, Maucione, Hood, 2007).

While an evidence-based-practice like assessing for implementation quality, is important for understanding program success and failure, there is another less pronounced rationale for implementing programs as designed. As a broad body of research suggests, the presence of an implementation assessment component often leads to better program outcomes (Durlak & DuPre, 2008). Thus, it is important that programs are implemented as designed to ensure the the best possible program outcomes (Lillehoj, Griffin, & Spoth, 2004). In other words, as researchers and practitioners committed to preventing problems and improving the circumstances of those they serve, programmers must implement programs as designed. This implementation-focused process will likely result in greater improvement in outcomes for program participants. Finally, at a practical level, both researchers and practitioners have also begun to recognize that in a competitive field implementation assessment is now assumed to be part of an evidence-based proposal, Wanless and Domitrovich (2015a) noted, "many funding agencies that

endorse the use of evidence-based interventions…now require researchers to include measures in research proposals to monitor intervention implementation" (p. 1037).

Despite this growth in evidence, there remains a gap between our understanding of how programs are implemented outside of the program designer's control (Berkel et al., 2011) and which factors contribute most to implementation quality. Specifically, there are challenges in operationalization and measurement of the factors that contribute to implementation quality. Given the relatively low level of LER-focused implementation research (Maineri & Anderson, 2015a; Tucker & Rheingold, 2010) and the applied nature of LER, this paper aims to provide a conceptual framework to guide LER implementation research. To develop this framework, the sections below: (1) explore the macro level factors that influence implementation quality, (2) examine implementation domains influenced at the facilitator level, (3) investigate micro level facilitator characteristics, traits, and actions that influence implementation quality, and (4) provide recommendations for measurement and operationalization of these factors, components, and traits based upon prior research.

Macro Level Factors Influencing Implementation Quality

In a review of implementation research related to community Extension programs, Gagnon et al. (2015b) highlighted how implementation quality is influenced by four primary factors: (1) organizational characteristics, (2) community characteristics, (3) program characteristics, and the focus of this conceptual framework, (4) facilitator characteristics. As illustrated in Figure 2, Gagnon et al. (2015b) described how these factors intersect to influence later programmatic outcomes.

Organizational Characteristics. The characteristics of the organization and

administrative leadership responsible for funding, selecting, designing, and/or delivering

a program influence implementation quality. For instance, the motivation and/or rationale

of an organization for choosing a program developed by a third party or developing one

themselves is often based on the availability of resources (e.g., human, property,

financial) in combination with their "buy-in" to provide a program (Dane & Schneider,

1998). This combination of resource availability and organizational support (e.g., belief

at an organizational level a program will make a meaningful difference within a

community or group) directly influences implementation quality. For example, in an

examination of an in-school violence prevention program, Elliott and Mihalic (2004)

proposed that despite a high availability of resources and this resource level being clearly

communicated to organizational leaders responsible for implementing the program, a lack

of organizational buy-in and support for the program ultimately led to poor program

attendance for both program staff and participants. This finding highlighted how at the

organizational level, the level of resources and buy-in to provide a program may be inter-

dependent, indicating that the presence of only one of these elements does not ensure

implementation quality.

Beyond the intersection of resources and motivation to provide a program, the

organizational enthusiasm to select and/or design and evaluate a program often varies. An

inconsistent level of organizational enthusiasm for a program may lead to challenges to
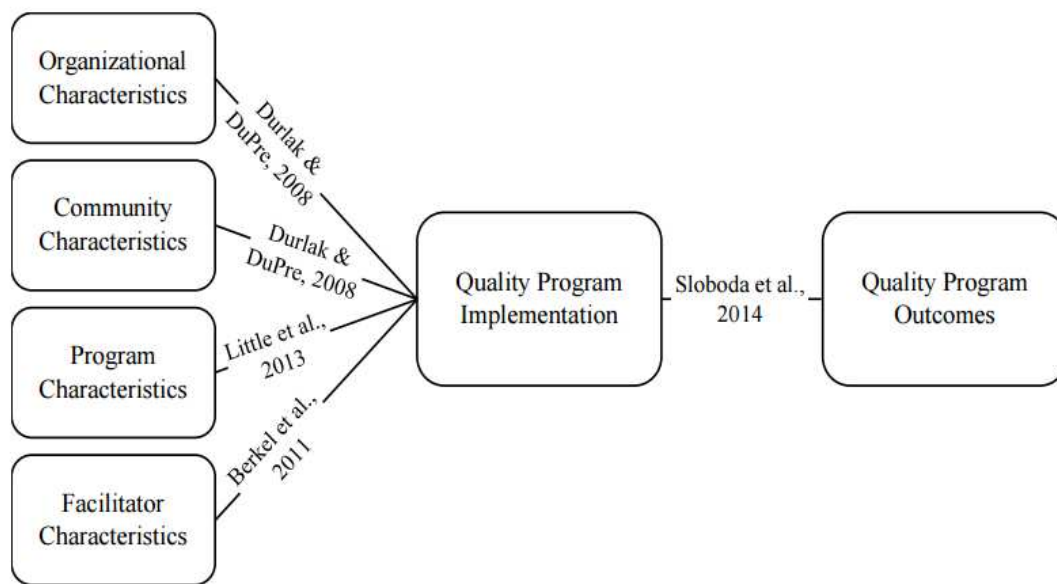
Figure 2. Conceptual model of the factors contributing to program implementation. Adapted from Gagnon et al. (2015b).

implementation quality. For example, the rationale for selecting/designing a program may be due to compliance requirements with a funding or accrediting organization (Kam, Greenberg, & Walls, 2003). In other words, an organization may be "bought-in" to program ABC due to their prior experience, fit, and/or familiarity with it, but the organization must deliver program XYZ to fulfill external demands and to sustain access to available resources and/or support. This "requirement versus preference" perspective can cause and organization negatively deliver a program. For example, in an examination of evidence-based programs being implemented at a statewide level, Fixsen, Blasé, Metz, and Dyke (2013) highlighted that top-down/external approaches (e.g., a state-level organization requiring a sub-organization to provide a specific program to maintain funding) often resulted in organizations implementing required programs poorly, due to a lack of internal motivation and an organizational perception of the program being a poor fit within their constituency.

In addition to an organization's resources, motivation, and program rationale, the organizational culture towards assessment also influences implementation quality. More specifically, as many organizations have an extrinsic pressure to sustain funding and demonstrate success in a program, there may be a corresponding growth in intrinsic organizational motivation to deliver a program as designed. For example, in a review of factors that contribute to poor program implementation, Fixsen, Blasé, Naoom, and Wallace (2009) suggested that funding agencies who compel organizations responsible for delivering programs should both demonstrate the value of a program and provide evidence that a program was implemented as designed, as this extrinsic pressure enhanced program outcomes and implementation quality. In summary, organizational characteristics including resources (i.e., financial, property, and personnel), level of support and enthusiasm for a program, and internal cultural view of program assessment, may influence program outcomes and implementation quality.

Community Characteristics. Paralleling organizational characteristics, the characteristics of the community being served also play a role in achieving desired program outcomes and implementation quality. At the community level, "implementation research can provide an understanding of the organizational and human capacities and motivation necessary to successfully adopt, implement, and sustain programs" (Mihalic & Irwin, 2003, p. 310). In a review examining factors that influence programmatic success at the community level, Lefebvre and Flora (1988) highlighted five factors: (1) the *communities' belief, support, and/or trust* that a program will make a meaningful difference within their community, (2) the *reach* of a program, (3) the amount of human

and financial *resources* available to the community to engage in the development and marketing of a program, (4) the communities *responsiveness* and support for a program, and (5) *programmatic saturation*, the level of similar programs historically and currently offered within the community. Each of these factors can have a unique influence on implementation quality.

The first community characteristic is the level of *community support, belief, and/or trust* that a program can achieve a desired effect can vary based upon a community's perceived need for a program, experience(s) with the organization responsible for providing the program, and the prevalence of past or ongoing programs within the community (Carroll et al., 2007; Elder et al., 2007, Wandersman et al., 2008). In a review of implementation research, Durlak and DuPre (2008) highlighted the importance of community-based decision making and support to achieve quality implementation, specifically that programs are generally sustained for longer and implemented better when shared decision making occurs.

Within implementation research, at both the organizational and community level, support and/or belief that a program will achieve desired outcomes is a key component to achieving implementation quality. Indeed, as suggested within the definition of *implementation fidelity* (i.e., the degree to which a program is delivered as designed) fidelity is influenced at multiple levels (Dusenbury, Branningan, Falco, & Hansen, 2003). When a program is not marketed to the designated population as recommended by program developers, a compromise to implementation quality can occur. For example, Fagan et al., (2008), highlighted how community level factors (e.g., lack of receptiveness

to the program as marketed) negatively influenced program reach, participant recruitment, and correspondingly implementation quality due to small numbers of participants (i.e., not enough participants to run programs as designed). Beyond a program achieving desired participant numbers, the second community characteristic, *program reach* also reflects the "extent to which participants being served by the program are representative of the target population" (Berkel et al., 2011, p. 24). Specifically, implementation quality can suffer if a program is being delivered to a population outside of a designer(s) intention and/or theory of change, as the initial design may not reflect this cultural difference.

In addition to program reach, recruitment, and marketing, the third community level characteristic is the *level of resources* to select, develop, and sustain a program also influence implementation quality. In a review of the factors that influence implementation quality at a community level, in a review of the community level factors that influence implementation quality, Goodman (2000) shared that in order for high quality implementation to be achieved there is a high burden on the part of the community to develop financial and political capital. If a community does not have the resources necessary to deliver a program as designed, nor the training or ability to evaluate a programs implementation or outcomes, it is unlikely that a program will be sustained or achieve the desired effects (Preskill & Boyle, 2008). Additionally, in order for evaluation to remain a continuous focus of those charged with implementing programs, "sustainable evaluation practice also requires the development of systems, processes, policies, and plans that help embed evaluation work into the way the

17

organization accomplishes its mission and strategic goals" (Preskill & Boyle, 2008, p. 444). Thus, the ongoing assessment of a program's implementation quality is a core component of the evaluation process, and the characteristics of the community providing it will influence a program's implementation, outcomes, and the thoroughness of the assessment taking place. Beyond the evaluation resources needed at a community level for effective program implementation, the context (and resource quality therein) in which a program is provided also can have meaningful influence on its implementation quality. For example, Williford et al. (2015) found in an assessment of a problem-behavior-prevention program, that when the program took place in a higher resourced context, in terms of facility quality and a lower teacher-student ratio, greater levels of both dosage and quality of delivery were achieved.

The fourth characteristic of the community being served relates to *participant responsiveness* to the selected program, defined as the degree to which "participants react to or engage in a program" including "participants' level of interest; perceptions about the relevance and usefulness of a program; and their level of engagement" (James Bell Associates, 2009, p. 2). The level of responsiveness and fit of a program at the community level can influence a program's implementation quality (Elder et al., 2007). As highlighted by Carroll et al. (2007) in a conceptual framework exploring the community-level factors influencing implementation quality, there is a direct relationship between participant responsiveness and implementation quality, where lower levels of responsiveness can lead to poorer quality implementation. Furthermore, in an assessment of a school-based substance abuse prevention program, Ennett et al. (2011), proposed that

participant responsiveness can influence a facilitator's quality of delivery, thus compromising implementation, but also could diffuse into other facilitator level elements including the facilitator levels of competence, confidence, and adherence. This study also indicated that a program's fit within a community, measured by the level of participant responsiveness, can also lead to deviations from the program design.

Finally, the fifth community level characteristic that influences implementation quality is the presence of similar past and ongoing programs. The number of these related and unrelated programs is referred to as the level of *program saturation*, and can also influence the marketing and promotion of similar programs. In other words, a community's past and current direct experience with similar programs (e.g., participating within similar programs or knowing others who have) or indirect experience (e.g., familiarity with similar programs due to marketing and promotion) may influence the community's receptiveness to an intervention, its implementation quality, and corresponding success (Lefebvre & Flora, 1988). To address this potential saturation at the community level, an assessment team must look to the community to determine their knowledge of current and past offerings within the desired program service area. This saturation data helps to mitigate contamination and confounds that may influence how a program is implemented in terms of recruitment, participant responsiveness, program differentiation, and program outcomes. Expounding on this potential limitation, Zief, Henke, Knab, and Zaveri (2011) described the how program saturation could be a challenge to implementation quality, specifically if a community was already saturated with similar programming, this information could be utilized for later conclusions

regarding compromises to implementation quality or lower than anticipated outcome levels, due to confounds outside of the program implementer's control. In summary, at the community level, implementation quality can be influenced by multiple factors including: a community's support and engagement with the program at both a broad and participant level, by the resources available to a community to provide and assess a program, and the level of past and current programs offered within a community that may have a similar audience or intended outcome.

Program Characteristics. The characteristics of a program itself may also influence its implementation quality. For instance, if a program is too complex, implementation quality may suffer. As highlighted in a review of implementation research, Durlak and DuPre (2008) indicated that as complex programs are delivered repeatedly within a community, the program quality may decrease. This decrease in program quality and outcomes is often due to the restrictions and complexity embedded within a program's design and the ability of an organization, community, and/or facilitator to deliver a program in a "real-world" setting without the support of the program-development team. For example, in a study of a complex teen pregnancy prevention program, Lesesne et al. (2008) highlighted ten distinct processes necessary to ensure high-quality implementation:

> (1) Needs & Resource Assessment; (2) Goal & Objective Setting; (3) Identification of Best Practices; (4) Assessing Fit; (5) Assessing Capacity and Readiness; (6) Program Planning; (7) Program Implementation & Process

Evaluation; (8) Outcome Evaluation; (9) Continuous Quality Improvement; and

(10) Program Sustainability. (p. 384)

While these processes are reflective of an evidence-based approach to program selection

(Wandersman, Imm, Chinman, & Kaftarian, 2000), they also highlight the level of rigor

necessary at a program level, to achieve high implementation quality, and this rigor

potentially increasing exponentially with growth in a program's complexity. To address

challenges related to complexity and delivery of a program in the real-world, the presence

of a "program-champion" to mitigate stakeholder concerns and challenges is a strategy

that has demonstrated a positive influence on implementation quality. For instance, when

a program requires extensive training to ensure facilitators can fully implement the

program as designed, this strain on organizational resources may cause training to be

condensed or compromised; a program champion can help to highlight the need to

maintain the organizations fidelity to designer recommended training levels and the

benefits of doing so (Mihalic & Irwin, 2003). In a study examining factors that promoted

or inhibited implementation quality within an in-school-substance abuse prevention

program, Ennett et al. (2011) highlighted how schools with the most knowledgeable,

supportive, and well trained program facilitators (i.e., program champions) tended to

have better overall implementation and program outcomes.

In addition to the challenges to implementation quality arising from program

complexity, another difficulty relates to the interdependency between a program and the

characteristics of the community and organization it is provided to. For instance, a

program's complexity and corresponding available resources impact implementation

quality (Mihalic, Irwin, Elliott, Fagan, & Hansen, 2004). This relationship between program complexity and resources is highlighted in a study of a youth violence prevention program, where implementation quality (defined as the number of core program components provided) was influenced by training quality and staffing levels (Mihalic & Irwin, 2003). More specifically, the primary cause of poor program implementation was resource related, in that the initial program design called for higher levels of trained staff to implement the program, but due to resource issues at the community and/or organization level, desired staffing levels (e.g., number of trained staff delivering the program) were not achieved, thus compromising implementation quality.

An additional program level characteristic associated with implementation quality relates to "cultural mismatch" (Castro, Barrera, & Martinez, 2004). This mismatch can range from a program being designed for English speakers and delivered to Spanish speakers to a program designed for those with higher SES than the group being served (Castro et al., 2004). In the instance where programs will likely encounter a cultural mismatch, program designers should provide methods to alter a program so the program better suits the current context (Carroll et al., 2007). The need for a cultural match when a program is provided, highlights the potential incompatibility high-fidelity program models may experience when programs are diffused outside of the researcher's control. Specifically, the primary goal of implementation research is to ensure that a program is delivered as designed and understand why they are not (Elder et al., 2007). Well-designed programs should include aspects of adaptability to ensure a later cultural match, thus anticipating or preventing these challenges, rather than reacting to them. In summary,

program characteristics such as the level of complexity required for design and delivery, the presence of a program champion, the levels of resources available (i.e., personnel, financial, property), and the cultural fit of a program all can influence implementation quality.

Implementation Quality Domains Influenced at the Facilitator Level

As noted earlier, an increasing body of evidence indicates that when programs are implemented as designed they tend to have better outcome levels than those that do not, suggesting that an implementation-focused culture may positively influence program performance (Durlak & DuPre, 2008). Correspondingly, there is ample evidence that suggests implementation quality is influenced at multiple levels (i.e., organizational, community, and program); however, as noted by Wanless et al. (2015b), ultimately it is the responsibility of the facilitator to ensure that a program is delivered as designed. At the facilitator level, there are six domains of implementation quality that a facilitator has partial or total control over: (1) adherence, (2) dosage, (3) quality of delivery, (4) participant responsiveness, (5) program differentiation, and (6) adaptation (Hansen, 2014). While these domains have been identified as contributing to implementation quality, there remains ambiguity in our understanding of how these domains collectively influence program outcomes (Berkel et al., 2011) and how these domains influence each other. For example, there remains a prevailing assumption that all six domains are positively related to each other (Ennett et al., 2011). However, this assumption, while logical, does not always bear out when applied to "real-world" examinations of implementation quality. For instance, Williford et al. (2015) highlighted how domains of

23

implementation were dependent on some domains, but not others, "implementation components were modestly or not significantly associated with one another. Dosage was positively linked with quality, and quality was positively linked with generalized practice. However, dosage was not significantly related to generalized practice" (p. 1061). This finding demonstrates a partial lack of congruence with conceptual implementation theory, prior implementation research, measurement issues, and evidence (e.g., Durlak and DuPre, 2008; Schoenwald et al., 2011). Furthermore, the findings of Williford et al. (2015) indicate: (1) it is critical that when implementation assessment is conducted that all potential inputs and outputs are captured, as (2) it remains unclear how implementation domains interact and/or cause and/or relate to each other.

Revisiting this manuscript's purpose, the development of a conceptual framework of the facilitator characteristics that influence implementation quality, the sections below (1) describe the six domains of implementation influenced at the facilitator level, (2) provide examples of facilitator characteristics and traits that influence these domains, (3) where applicable, share differentiation in operationalization between studies of implementation, and (4) supply recommended measurements of constructs.

Domain One: Adherence. Within the research investigating facilitator-level contributions to implementation quality there is a degree of ambiguity regarding the operationalization of adherence. For example, Fagan et al. (2008) defines *adherence* as the "degree to which implementers taught the required program objectives or fulfilled the program's core components" (p. 242). Embedded within this definition are two constructs: (1) the degree to which implementers (i.e., facilitators) taught the required

objectives and (2) whether facilitators fulfilled the program's core components. Fagan et al. (2008) operationalized objectives as dichotomous (i.e., provided or not) and components as either specific program elements or as meeting a program goal, benchmark, or milestone. While in this study Fagan et al. (2008) suggested a dual definition of adherence, other studies suggest adherence is better operationalized as a dichotomous variable. For example, in a study of facilitator burnout and its influence on implementation quality, the measurement of adherence was dichotomous, where program raters simply counted whether a specific program component was observed or not (Wehby, Maggin, Moore-Partin, & Robertson, 2012). This operationalization reflects the first part of the definition of Fagan et al. (2008), and seems to capture the more focused description of adherence provided in the systematic reviews provided by Durlak and DuPre (2008), "the extent to which the innovation corresponds to the originally intended program" (p. 329) and of Berkel et al. (2011) "whether prescribed program components were delivered as instructed in program protocol" (p. 24). However, in both systematic reviews, adherence was used synonymously with fidelity. This definitional incongruence is further highlighted with the term, *implementation fidelity*, the degree to which a program is delivered as designed (Dusenbury et al., 2003) and *program integrity*, the degree to which a program is delivered as planned (Dane & Schneider, 1998). Within these four definitions, the common operationalization indicates that adherence refers primarily to the degree of a program component that was delivered; thus, facilitator *adherence* is the degree to which a facilitator(s) follows the program schedule, manual, and/or curriculum guide.

Domain Two: Dosage. In contrast with adherence, there seems to be slightly less confusion regarding the operationalization and definition of dosage at the facilitator level. Specifically, *dosage* is defined as how much of the program was delivered as designed (Durlak & DuPre, 2008). However, in some studies of implementation quality, dosage is measured with more specificity. For example, Borrelli et al. (2005) operationalized dosage as a combination of measures including each program's session length, total amount of session time across all program sessions, and percentage of content delivered. Similarly, in an investigation of a socio-emotional development program, Caldwell et al. (2008) operationalized dosage as the number of persons in attendance of the program [measured on a 1 (none present) to 5 (almost all present) scale] and the percentage of content covered [measured on a 1 (0%) to 5 (100%) scale].

Some evidence suggests that there is a definitional crossover between adherence and dosage, indicating that some researchers use the terms interchangeably. For instance, Durlak (2015), defined dosage as the percentage of a program or intervention delivered. Similarly, Mainieri and Anderson (2015a) measured fidelity (referring to amount of program content delivered) using a checklist [measured on 0 (none) to 1 (all material present) scale] measuring implementation quality in terms of percentage delivered. In both examples, there seems to be a lack of precision on where dosage and adherence are distinct constructs, likely reflecting confusion in the extant implementation literature. In both Durlak (2015) and Mainieri and Anderson (2015a), the authors utilize *the degree of a program that was delivered* as a measure of implementation quality [paralleling the definition of adherence provided by Durlak and DuPre, (2008)]. However, despite the

unintended definitional crossover, Dusenbury et al. (2005), address this challenge by defining *dosage* as, "providing sufficient exposure to the program" (p. 308) and *adherence,* "as following program methods and completing its delivery as outlined in a manual or curriculum guide" (p. 308). In other words, the operationalization of dosage provided by Durlak (2015) could be reframed to also reflect the percentage of the *core* program or intervention delivered, mirroring the definition provided by Mainieri and Anderson (2015a), where dosage (referring to fidelity) is defined as the percentage of core components delivered. In this regard, a more holistic conceptualization of *dosage* reflects the degree, percentage, or amount of the core original program that is actually delivered by the facilitator(s).

Domain Three: Facilitator Quality of Delivery. In addition to adherence and dosage, a facilitator's quality of delivery can have a meaningful influence on implementation quality and corresponding program outcomes (Spoth, Guyll, Lillehoj, Redmond, & Greenberg, 2007). At the facilitator level *quality of delivery* refers to the facilitators enthusiasm, motivation, buy-in, and engagement with the program and its participants (Ennett et al., 2011). However, as with the definition of adherence provided earlier (i.e., Fagan et al., 2008) this definition suggests two sub-dimensions: (1) the facilitator's attitude while delivering the program and (2) the facilitator's attitude towards the program (not necessarily during program delivery). While there is a logical relationship between a facilitator "buying-in" to a program (i.e., demonstrating a positive attitude and belief regarding the program's value) and thus delivering it well (i.e., with quality), it is important to recognize that these constructs are often treated as independent

27

variables within implementation research. For instance, Gagnon and Bumpus (2016) assessed a pre-college leadership development program (i.e., StepOne) and examined both *facilitator buy-in* to the program (e.g., a sample facilitator item was "StepOne will make a meaningful difference") in addition to *facilitator quality of delivery* (e.g., a sample facilitator item was, "I will deliver StepOne well"); Gagnon and Bumpus found that facilitator buy-in did have a meaningful effect on program implementation, but quality of delivery did not. These results indicate that if the measures were combined, the effect of buy-in may have been suppressed, highlighting the importance of definitional congruence, and the separation of buy-in and quality of delivery.

Regardless of the partial definitional confusion within some implementation research regarding quality of delivery, a body of evidence suggests quality of delivery has a positive effect on program implementation quality. For example, in a study of a character development program for youth-at-risk, Malloy et al. (2015) found that a facilitator's *quality of delivery* was positively associated with the number of program sessions provided per week and the use of additional program specific materials when facilitating outside of the program curriculum. A similar interpretation of this more unidimensional definition of facilitator quality of delivery is provided by the definition of Durlak and DuPre (2008), where "quality refers to how well different program components have been conducted…are the main program elements delivered clearly and correctly" (p. 329). At a more detailed level, Dusenbury et al. (2003) operationalized quality of delivery as "ratings of provider effectiveness which assess the extent to which a provider approaches a theoretical ideal in terms of delivering program content" (p.

244). In other words, at the facilitator level, *quality of delivery* refers to the amount of enthusiasm, skill, and/or competency to which the facilitator delivers the program.

Domain Four: Participant Responsiveness. In the context of implementation science, *participant responsiveness* refers to the level of a participant's motivation, engagement, and interest in a program (Dane & Schneider, 1998). Within the broader implementation research there are several potential sub-dimensions embedded within the operationalization of participant responsiveness, including: (1) the participant's level of satisfaction with the program (Hansen, 1996), (2) the participant's level of program-specific knowledge (Rohrbach et al., 1993), (3) the participant's level of program (re)attendance (Cantu, Hill, & Becker, 2010), and (4) the participant's level of interest and engagement with the program (Carroll et al., 2007; Rohrbach et al., 2010). Additionally, there is conflation between attendance and dosage in some of the extant implementation literature. In addressing this challenge in a review of implementation research, Berkel et al. (2011) indicated that dosage refers to the number of sessions that were offered, but participant attendance is a measurement of program sessions attended.

The direct effect of participant responsiveness on implementation quality is unclear. Some researchers contend that participant responsiveness is an outcome of a facilitator maintaining implementation quality (see Berkel et al. 2011, Figure 1), where a facilitator behavior (e.g., a program adaptation) causes participant responsiveness and in turn participant responsiveness causes and/or mediates program outcomes. However, the primary challenge relating to participant responsiveness may not be one of directionality (e.g., facilitator quality of delivery *causing* participant responsiveness) rather a more

foundational one. In the broader implementation literature, participant responsiveness is treated as a sub-dimension of implementation quality, *not an outcome*. For example, in Durlak and DuPre (2008) this treatment of participant responsiveness as a sub-dimension of implementation quality is highlighted: "…aspects of implementation (e.g., quality, adaptation, reach, program differentiation, and *participant responsiveness*)" (p. 342). Similarly, Carrol et al. (2007) uses this integrated phrasing when operationalizing implementation quality "…these are: adherence to an intervention; exposure or dose; quality of delivery; *participant responsiveness*; and program differentiation" (p. 2). Likewise, in Cantu et al. (2010) participant responsiveness is treated as component of "…implementation quality (adherence, *participant engagement* or facilitation quality) …" (p. 20). In a special issue of *Prevention Science* investigating implementation, Durlak (2015) provided a commentary indicating that participant responsiveness is a sub-dimension of implementation quality, sharing "the components of implementation that were studied included dosage, fidelity, quality of delivery, and *participant responsiveness*…" (p. 1123). This blended conceptualization suggests that participant responsiveness within implementation research is a subdimension of implementation quality rather than a consequence (i.e., outcome) of overall implementation quality or another implementation sub-dimension (i.e., domain).

The causal structure between implementation quality and participant responsiveness is unclear, where a program that is implemented well may actually cause participant responsiveness to increase, suggesting a recursive relationship between participant responsiveness and other domains (Carroll et al., 2007). An example of this

recursive relationship is highlighted in Figure 3, where higher levels of participant responsiveness may cause higher levels of facilitator adherence thereby causing higher levels of participant responsiveness. Essentially, there appears to be some conflict within implementation research regarding participant responsiveness as an outcome of implementation quality, but a broader body of literature that supports participant responsiveness as a sub-dimension of implementation quality. Additionally, the implementation science literature suggests that participant responsiveness is malleable by program facilitators (Wanless et al. 2015b); indicating that at the facilitator level, *participant responsiveness* can be defined as the degree to which program participants are engaged and/or involved in the program tasks and/or responsibilities due to the facilitator.

Domain Five: Program Differentiation. The program differentiation domain also reflects a degree of definitional and conceptual ambiguity at the facilitator level. Specifically, there is a lack of clarity regarding the relationship of program differentiation with implementation quality; some implementation research suggests program differentiation is a sub-dimension of implementation quality (e.g., Berkel et al., 2011; Durlak, 2015) while other research suggests it is an external factor that can inhibit or promote implementation quality. This distinction is proposed in a review of implementation research by Century et al. (2010):

> …differentiation is not a dimension of fidelity per se but rather is an analytic
> process by which an evaluator determines the degree to which the critical
> components that distinguish one program from another are present or absent.

Thus, it is a process that one undergoes before, during, or after measurement of

implementation. (p. 208)

Here Century et al. (2010) suggest program differentiation can influence implementation

quality, it is not a subdimension of implementation quality. However, it is possible that

program differentiation can have an independent effect on implementation quality. For

example, program differentiation may influence program saturation, where participants

associate differing programs with the one being implemented (Lefebvre & Flore, 1988).

This suggests that program differentiation is a subdimension of implementation quality,

dependent on its conceptualization. Additionally, the reviews of Carrol et al. (2007) and

Century et al. (2010) both suggest program differentiation is conceptualized as outside of

the control of the program facilitator. However, further conceptualizations of program

differentiation indicate that program differentiation may act as a subdimension of

implementation quality rather than an outside influence upon it. For instance, Durlak and

DuPre (2008) noted in their systematic review of implementation research that program

differentiation refers to how a program's theory and application differs from other

programs where the facilitator's level of implementation quality reflects: (1) a program's

core components and (2) the degree this implementation quality is differentiated from

unrelated programs. This facilitator-centric approach is supported by Dusenbury et al.

(2003), who measured program differentiation as, "the degree to which elements which

would distinguish one type of program from another are present or absent" (p. 240). In

other words, at the facilitator level, the domain of program differentiation captures the

degree to which a facilitator emphasizes a program's core components in terms of: (1) the
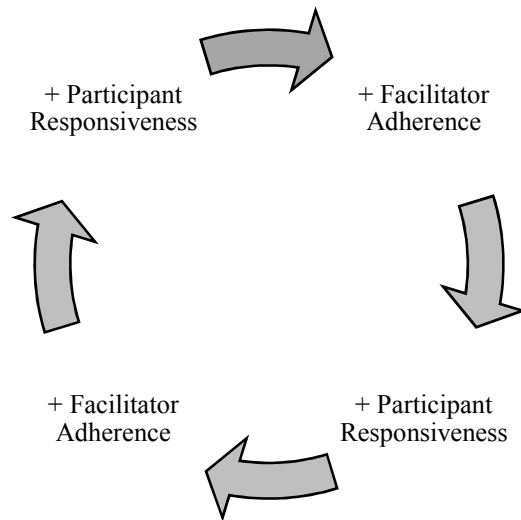
Figure 3. Example Recursive Relationship Between Implementation Domains

program plan, (2) the facilitator's level of ability to provide the program, (3) navigation

of challenges due to external factors (e.g., community or organization level issues),

and/or (4) mitigation of challenges due to variable levels of participant ability (Berkel et

al., 2011; Dane & Schneider, 1998). This facilitator-driven effort to maintain

implementation quality through various emphases of core program components reflects

how, at the facilitator level, program differentiation may offset deficits in the additional

domains of implementation quality (e.g., adherence, dosage, quality of delivery,

participant responsiveness, and adaptation).

Moreover, while the conceptualization of the role of program differentiation

provided by Century et al. (2010) is meaningfully different from those of Durlak and

DuPre (2008) and Dusenbury (2003) in terms of the relationship between program

differentiation and implementation quality, there remains a strong parallel between the

distinct conceptualizations, relating to participant experiences. For example, it is likely

program participants have experienced or been exposed to an unrelated program or

intervention prior to engaging in the program of interest (Lefebvre & Flore, 1988); thus, there is degree of potential "contamination" possible due to this prior experience. In discussing this contamination, Camire, Forneris, Trudel, and Bernard (2011), shared that one of the greater challenges with a facilitator (i.e., coach) implementing a program was enhancing a facilitators' understanding that program participants likely have prior experience within similar contexts. This facilitator recognition of prior participant experiences in related contexts may cause the facilitator to adapt, remove, ignore, or modify a program component, undermining its implementation quality. However, program differentiation can act as a counterbalance to other potentially compromised domains of implementation quality, where a higher emphasis in one area can address lower emphases in other areas (Fixsen et al., 2009).

In summary, program differentiation can be influenced by both facilitator and/or community level factors; thus, suggesting a multi-part definition, where *program differentiation* refers to program components and outcomes that can be: (1) directly attributed to the current program and (2) independent from prior/current participant exposures to unrelated programs. Further, this attribution and independence can be influenced by the program facilitator and/or external factors.

Domain Six: Adaptation. Within the implementation science literature there is an often-mentioned fidelity-adaptation debate (Hecht & Miller-Day, 2010) suggesting that the terms *fidelity* and *adaptation* are mutually exclusive within investigations of implementation quality. In other words, the opposite of fidelity (i.e., adherence) is adaptation, suggesting that the adherence definition provided earlier, *the degree to which*

34

a facilitator(s) follows the program schedule, manual, and/or curriculum guide, could also apply to the adaptation domain, as decreases in adherence represent changes to the program during implementation. While this opposition may be parsimonious for future implementation research, this definitional dual-operationalization (i.e., the opposite of one indicates the presence of the other), is not reflected in assessments of facilitator implementation, where information regarding adherence levels is commonly collected in parallel with adaptation data. More simply, current recommendations for implementation assessment suggest that adaptation and adherence should be measured concurrently due to ongoing conceptual development in both areas (Century et al., 2010; Mowbray et al., 2003).

Further, there is evidence adaptation is multi-dimensional, and correspondingly can have negative, positive, or neutral influences on implementation quality. More specifically, adaptations should be thought of as occurring along a continuum, where some adaptations can help a program's outcomes and others may harm it. In other words, adaptations can be, "*positive* (aligned with the goals and theory); *neutral* (neither aligned with nor deviated from the goals and theory); or *negative* (deviated from the goals and theory)" (Moore, Bumbarger, Rhoades-Cooper, 2013, p. 151). Thus, when collecting adaptation data, the type, reason, and frequency of adaptation should also be captured, which further supports the importance of definitional separation between adherence and adaptation. It is in this continuum-based context that, at the facilitator level, *adaptation* represents the frequency, degree, and style of change(s) made outside of the original program design during implementation.

Micro Level Factors Influencing Facilitator Implementation

Implementation researchers and evaluators must navigate the balance between collecting enough information to determine how/why/what occurred during a program, but not to the degree that it inhibits the programs delivery and outcomes (Meyers, Durlak, & Wandersman, 2012). As noted earlier there are many potential influences on implementation quality, such as characteristics of the organization, community, and program. Nonetheless, it is ultimately the responsibility of the facilitator providing the program to determine the degree to which a program is implemented as designed (Wanless et al., 2015b). As such, an understanding of the facilitator characteristics that best contribute to implementation quality is useful information for program stakeholders. As noted by Jelalian et al. (2014), an understanding of the facilitator characteristics that best contribute to implementation quality and program outcomes will aid those charged with the selection, development, and training of facilitators to best reflect those characteristics where possible.

While the use of facilitator characteristics to predict implementation quality is promising, the strength of relationship between implementation quality and facilitator characteristics remains unclear. Specifically, dependent on the implementation study, facilitator characteristics had a positive effect (e.g., Lillehoj et al., 2004), a negative effect (e.g., Hill et al., 2007), and in some cases no effect (e.g., Cantu et al., 2010) on implementation quality. This lack of clear relationships highlight how both facilitator characteristics and implementation quality may be dependent on the program, community, and organization. As such, to better understand the effects and measurement

of facilitator characteristics in the context of implementation science, the sections below: (1) describe eight facilitator traits/characteristics within various implementation studies, (2) illustrate how these micro-level factors, dependent on their measurement and context, may influence implementation quality differently, (3) provide recommendations for measurement in future investigations, and (4) present a conceptual framework (see Figure 4) of how the characteristics interact with the six domains of implementation quality presented earlier.

Facilitator Experience and Implementation Quality

Within the implementation literature, it is frequently proposed that more experienced facilitators will have higher quality implementation, due to the assumption that this experience will aid facilitators in their conceptual understanding of the program and later implementation (Berkel et al., 2011; Fixsen et al., 2009). However, the relationship between facilitator experience levels and implementation quality in previous research is indistinct due to the: (1) differing styles of measurement and (2) conflicting study results suggesting higher levels of experience can help, harm, or not influence implementation quality. This lack of clarity highlights the importance of exploring the influence of experience on implementation quality, as this information will assist program designers and evaluators in their understanding of whether a program was delivered as designed (Morgan et al., 2016).

As with many dimensions of implementation quality, there are differences in the conceptualization of facilitator experience in terms of measurement and later analysis within investigations of program implementation. At the facilitator level, experience is
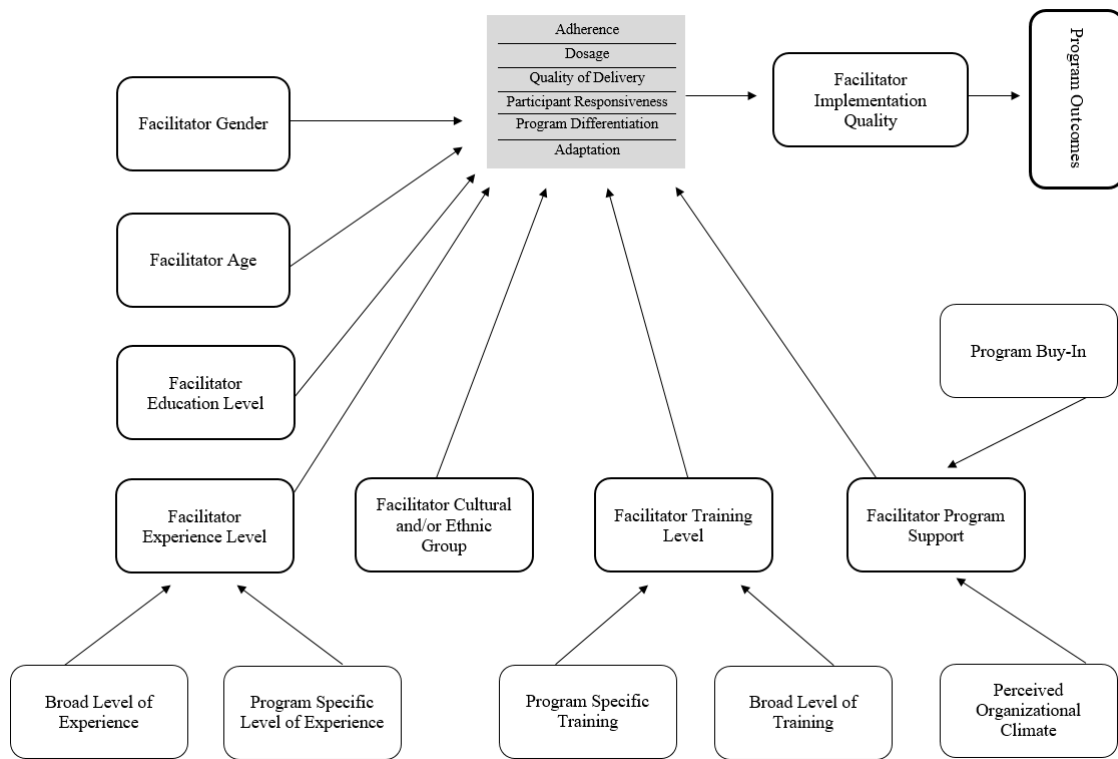
Figure 4. A Conceptual Model of Facilitator Characteristic Influences on Implementation Quality

frequently operationalized at one of two levels: (1) the facilitator's broad level of experience delivering programs, both related and unrelated to the program of interest and (2) the facilitator's level of experience delivering the specific program of interest (Dusenbury et al., 2003). When measured at a broader level, facilitator experience has been found to help, harm, or not influence implementation quality. For example, Desimone and Lee-Hill (2017) examined a middle school youth science program where teacher experience was operationalized as novice or non-novice, with novice teachers possessing two years or less of experience. Utilizing this approach, Desimone and Lee-Hill (2017) found no significant influence of experience on implementation quality. This null result parallels the findings of both Pas, Waasdrop, and Bradshaw (2015) and

Domitrovich et al. (2015). These results indicated that there was no statistically significant influence of experience (measured as total years teaching) on implementation quality. While these findings indicate facilitator experience, measured broadly as total years, has no influence on implementation quality, other studies examining this relationship have found differing results. For example, Lillehoj et al. (2004) examined the relationship between implementation quality and facilitator experience (reported as a broad experience category in number of years) within a youth substance abuse prevention program. The authors found that a facilitator's level of experience had a positive influence on both implementation quality and program outcomes. Similarly, in an examination of a program to increase socioemotional development (i.e., life skills) in youth, Dusenbury et al. (2005) found that program facilitators with higher levels of broad experience (measured in number of years) tended to be more adherent to the program design, to better meet program objectives, and to deliver greater levels of core program components. In this study, Dusenbury et al. (2005) also noted that broad facilitator experience had a positive influence on participant responsiveness (i.e., engagement), indicating that experience, measured at a broad level, can influence multiple domains of implementation quality (e.g., adherence and participant responsiveness).

While the relationship between a facilitator's broad experience and implementation quality is relatively uncertain, a similar unclear relationship is present when examining the relationship between program specific experience and implementation quality. For example, in a study of a drug-abuse prevention program, Pankratz et al. (2006) found no meaningful effect of either facilitators' broad experience

nor program specific experience (both measured in number of years) on implementation quality. Similarly, in a study of a cardiac health promotion program, Riley, Taylor, and Elliot (2001) also found no meaningful effect of broad facilitator experience (in years) nor program specific time (measured in average hours per day) on implementation quality. In partial contradiction to these findings, Dusenbury et al. (2005) did not find a meaningful influence of program specific experience on adherence (as a sub-dimension of implementation quality); however, Dusenbury et al. (2005) did find that facilitators with higher levels of broad and program specific experience made more positive adaptations to a program.

While there appears to be a disparity in investigation between the two measures of facilitator experience (i.e., broad or program specific), this is not a reflection of the lack of importance of either level of facilitator experience; rather, this disparity likely reflects that many assessments of implementation quality are conducted within the pilot phase of programs, where a research team exerts a high level of control on aspects of the program's design, implementation, and assessment (Ennett et al., 2011; Fixsen et al., 2013). More specifically, when a program is in the pilot or developmental phase, facilitators are less likely to have sufficient program specific experience. Thus, given the formative phase of implementation science the relationship between program specific experience and implementation quality should become clearer as programs designed with implementation assessment continue (e.g., Moore et al., 2013; Spoth et al., 2007).

Additionally, the measurement of experience appears to have conceptual challenges. For instance, in a pre-K problem behavior program, Williford et al. (2015)

measured experience in terms of total years working within a pre-K environment. While this approach reflects that of other studies examining the relationship between facilitator experience and implementation quality (e.g., measuring experience in years), it may not capture enough of the variable (experience) to explain a relationship (or lack thereof) between it and implementation quality. In other words, facilitator experience may be "under-measured" when it is quantified only as a facilitator amount of "time" (i.e., hours, days, years) delivering programs and/or working within the program context. In addressing this shortcoming, Weekley and Ployhart (2005) indicated that many measures of work experience are prohibitively unidimensional, failing to recognize the potential variables that also constitute work experience, and differences embedded within work experiences (time in current position vs. total time with current organization vs. total time in differing organizations). These differences may inflate or inhibit the potential role of experience on program outcomes (Weekley & Ployhart, 2005).

In addition to facilitator experience likely being a more multi-dimensional construct than previously measured, there are also challenges with how experience is measured in later relationship testing. For example, Desimone and Lee-Hill (2017), noted "exploration of a continuous teacher experience control variable yielded null results, as we might expect, because a continuous variable assumes a strict linear relationship, which contradicts previous research, which shows substantial gains after the first 2 to 3 years" (p. 8). This linear approach is also reflected in LER studies of implementation (e.g., Gagnon & Bumpus, 2016) examining the relationship between experience and implementation quality (e.g., camp counselors, coaches, after-school program facilitators,

41

challenge-course facilitators). While common to the implementation literature, this continuous linear approach to the measurement of experience presumes that a program facilitator will change at an identical rate in performance as their experience level also increases. This is a likely reason for the parsing of experience from a continuous level of measurement (e.g., number of years) to a categorical one (e.g., novice vs. expert) where facilitator experience likely reflects a more inconsistent influence (e.g., facilitator experience becomes increasingly valuable over time). Given the challenges with both the operationalization and measurement of experience, future studies exploring the relationship between implementation quality and facilitator experience should utilize multiple measures of experience (also see Table 1), including: (1) a facilitator's broad level of experience providing programs both similar and dissimilar in nature, (2) a facilitator's experience delivering the program of interest, (3) if applicable, the amount of time a facilitator spent in the program as a participant, (4) and the frequency and/or number of programs the facilitator has delivered.   Finally, it is possible that the measure of experience may not be discrete enough to capture variance across facilitators; specifically, many studies utilize a year as a measure of experience, but more discrete measurement of experience as time (e.g., hours, days, weeks) in addition to number of sessions a facilitator provided may aid future evaluators to explain the relationship (or lack thereof) between facilitator experience level and implementation quality.

Facilitator Training and Implementation Quality

Of the facilitator level factors thought to contribute to implementation quality, training likely has the most robust body of research (Durlak, 2015; Fagan et al., 2008;

Ringwalt et al., 2003; Meyers et al., 2012). High quality program facilitator training has been positively linked to high quality program outcomes (Wehby et al., 2012); however, there remains a degree of ambiguity about the features and dimensions of facilitator training that influence implementation quality and further the degree to which these features and dimensions influence corresponding program outcomes (Dusenbury et al., 2003). Furthermore, the content (including the relevance of implementing the program as designed) covered as part of facilitator training may also influence implementation quality. For example, Wandersman et al. (2008) indicated facilitator training emphasizing the importance of implementation, as part of overall programmatic facilitator training, may positively influence implementation quality and program outcomes. However, Hill et al. (2007) indicated that this may be too simplistic, specifically "improving implementation is likely not simply a matter of stressing the importance of fidelity during training" (p. 31), rather it is more likely the identification of reasons for program adaptation will better mitigate future compromises to implementation quality.

Training shares similar dual-dimensionality with the facilitator experience characteristic, that is, a facilitator's level of training can also be measured as two-dimensional, consisting of (1) specialized training associated to the program of interest and, where applicable, (2) broader level training with related and unrelated programs. Additionally, despite evidence suggesting that training may influence implementation quality, "research examining the specific features of training that promote effectiveness has been very limited" (Dusenbury et al., 2003, p. 249). This lack of examination of the specific features of training (e.g., role playing, lecture-based, facilitator observation) has

led to training being measured as a characteristic that a facilitator does or does not possess, thus suppressing the likely variability of trainings influence on implementation quality. Within a study of a student literacy program, Zvoch (2012) noted, while training is often provided at differing levels for facilitators, dependent on organizational and community characteristics, it is frequently treated as a dichotomous variable. This dichotomous approach is reflected in many studies of implementation quality. For example, in an implementation assessment of a middle-school science-achievement program, where teachers did (or did not) get the program specific training as part of a randomized group assignment, Desimone and Lee-Hill (2017) noted, that despite random group assignment, facilitator training (or lack thereof) did not have meaningful influence on the desired program outcome (i.e., science achievement). However, training (measured dichotomously) did have a positive influence on implementation quality. Surprisingly, the researchers also found that program specific training had a negative influence on teacher content knowledge (i.e., competency).

The challenge in understanding the relationship between implementation quality and facilitator training is that in some studies training has no effect, a positive effect, or a negative effect on implementation quality. Moreover, the reasons for the relationship (or lack thereof) between implementation quality and training may be due to a combination of measurement error and a training effect. For example, in further exploration of why training did not have a meaningful influence on program outcomes in their study, Desimone and Lee-Hill (2017) indicated that their measure of training did not adequately reflect potential outside sources of training to the non-program-trained comparison group

44

and indicated that training is likely a more multi-dimensional construct. As a dichotomous measure, training did not have a significant influence on implementation quality. This result partly parallels the findings of Rohrbach et al. (1993), who in an examination of a substance abuse prevention program, found no significant effect of staff training (comparing intensive versus limited training) on implementation quality or program outcomes.

However, the finding of Rohrbach (1993) partially contradicts other research indicating that staff training positively influences implementation quality. For example, in a youth substance abuse prevention program, Lillehoj et al. (2004) found that both program specific training and prior unrelated training had a positive effect on program implementation and the desired outcomes (e.g., more negative attitudes towards substance abuse). Supporting this finding, in a study of a municipal youth program, Morgan et al. (2016) noted that program-specific training was likely a direct cause of high-quality implementation in the program they assessed. However, somewhat in contradiction to the null findings of Rohrbach et al. (1993), and the positive findings of Lillehoj et al. (2004) and Morgan et al. (2016), in a study of pre-college leadership development program, Gagnon and Bumpus (2016) found significant negative relationships between general training levels and self-reported fidelity, program buy-in, and pro-fidelity beliefs. Their findings indicated that in the context of a pre-college preparation program, higher levels of general (i.e., unaffiliated with current) program training may have a negative influence on implementation quality; although, Gagnon and Bumpus (2016) did not find a meaningful relationship between program specific training

45

and implementation quality. In summary, the relationship between training and implementation quality is likely unclear due to measurement issues as well as the breadth, depth, and source of training.

The comprehensiveness of the training provided also can influence implementation quality. For example, in a study of a high-school drug abuse prevention program, Little et al. (2013), highlighted how program support (i.e., a favorable attitude toward the program achieving the desired outcomes) and training were interactive for program facilitators. Specifically, Little et al. (2013) explored the differences between a comprehensive training (e.g., a higher level of technical assistance, support and coaching, emphasizing the importance of fidelity to facilitators) to a standard training (e.g., no follow-up or coaching) and found that the more comprehensive training was associated with increased implementation quality. This more robust training and support, beyond the initial pilot program, has also been shown to be an effective model of achieving desired program outcomes (Stein et al., 2008). In a study of community youth development programs, Fagan et al. (2008) found that ongoing training and support (i.e., comprehensive training) that "includes program implementation monitoring tools and procedures…can help bridge the gap between the quality of program implementation typically achieved in prevention research studies…" and thus will enhance facilitator "adherence to program protocols and procedures, and, as a result, enhance their likelihood of realizing anticipated benefits to program participants" (p. 247). Similarly, Rohrbach et al. (2010) found that as part of a drug abuse prevention program, comprehensive training positively influenced implementation quality; however, Rohrbach

et al. (2010) also noted that this increase in implementation quality did not meaningfully influence program outcomes, indicating further research is needed in this area. For instance, in some studies, training influences implementation quality, but not outcomes (e.g., Desimone and Lee-Hill, 2017; Rohrbach et al., 2010) and in others, training influences both implementation quality and outcomes (e.g., Little et al., 2013).

Beyond the effect of training style and content on implementation quality, there are also potentially more nuanced factors that can influence how training quality itself is vulnerable to implementation quality issues. Specifically, many programs use an approach where an outside organization provides initial training and later follow-up (e.g., in-service) program training is provided by internal organizational members. Pas et al. (2015) identified that the potential for "slippage" (i.e., error) in this transition can not only influence the quality of training later program facilitators receive, but also influence later implementation quality. Similarly, in a review of clinical implementation research, Gearing et al. (2011) referred to this as "drift" where those responsible for training and/or implementing a program progressively began to modify the curriculum by condensing, adapting, or adding material to the original program. In this review, Gearing et al. (2011), noted this drift as a gradual process rather than more obvious deviations from a program curriculum, and correspondingly indicated the importance of regular program monitoring (e.g., implementation assessment and technical assistance) to mitigate and/or correct this training drift.

Furthermore, additional exploration of the unique effect of training on implementation quality compared to (or in combination with) additional facilitator

characteristics (e.g., gender, experience, education) may highlight the importance of training's contribution to implementation quality. For example, in a study of a pediatric weight control program, Jelalian et al. (2014) noted that program specific training led to quality implementation of the program; their findings also indicated that a lack of experience delivering similar programs did not influence programmatic implementation quality. Moreover, investigation of the relationship between training and implementation quality should shift beyond simply whether training was delivered to also include exploration of training efficacy (e.g., facilitator competency improvement). The study of Desimone and Lee-Hill (2017) is one of the first to also explore this improvement in competency as a component of training.

Interestingly, the findings of over 20 years of implementation research (i.e., Rohrbach et al., 1993 to Durlak, 2015) indicate that the effect of training on implementation quality remains unclear. This ambiguity only further highlights the importance of collecting training related data from facilitators to better understand the effect facilitator training may have on a specific program's implementation quality, benefiting both the program, as well as the broader implementation sciences. Given the relatively broad support for facilitator training having a positive influence on implementation quality (e.g., Durlak & DuPre, 2008; Fagan et al., 2008; Ringwalt et al., 2003), training remains a facilitator characteristic that warrants further exploration. Specifically, measurement of facilitator training (see also Table 1) should include: (1) the broader level of facilitator training prior to a program, (2) the specialized-program training that a facilitator receives, (3) the facilitator's level of competency prior to and

48

after training, (4) the role of the person(s) responsible for providing facilitator trainings organizational role and their experience providing said trainings, and (5) the number and duration of ongoing or later trainings. Within these measurements, there is also a potential dosage effect that could occur. For example, for instance, when a trainee attends 4 of 5 program specific trainings sessions or the trainee receives some training from an outside organization. Finally, as with facilitator experience, the training exposure (e.g., measured in units of time) could also influence the effect of training on implementation quality in later relationship testing.

Facilitator Ethnic or Cultural Group and Implementation Quality.

In the context of implementation science there has been not only a push to ensure that programs are implemented as designed, but also a more recent emphasis on ensuring that programs are culturally relevant to the groups they are intended to serve (Castro, Barrera, & Martinez, 2004). This push for cultural relevance is reflected in the intentional adaptation of programs to ensure they reflect the cultural needs of both the facilitator and the person(s) being served. However, there remains a conceptual misunderstanding of when an adaptation shifts from a compromise to implementation quality to a cultural adaptation. This tension is highlighted in a review of recommendations relating to cultural adaptation by Barrera, Berkel, and Gonzalez-Castro (2016):

> Adaptations convey the sense that changes are often made spontaneously during intervention sessions, perhaps in reaction to unanticipated conditions (e.g., disruptive classroom events) or perceived needs of participants (e.g., misunderstood session content). Cultural adaptations and local adaptations differ

on several dimensions. Cultural adaptations are developed prior to broad-scale implementation, are intended to reach populations (e.g., hypertensive African-Americans in Georgia's community health centers), and focus on cultural fit. Local adaptations are made just prior to or during intervention sessions, are directed at specific intervention sites (e.g., a community, clinic, or classroom), and could be done to improve cultural fit as well as a number of other idiosyncratic considerations including lack of time or physical resources. (p. 2)

More simply, "local" adaptations may represent genuine compromises to implementation quality, while "cultural" adaptations are positive in nature (e.g., reflecting a programs goals and theory, Moore et al., 2013) and are made at a more macro level. Thus, while a program being culturally relevant is a key component of ensuring implementation quality, these adaptations are generally outside of the facilitator's control. However, a cultural adaptation of a program could also include program designers ensuring an ethnic and/or racial match of the facilitator to the group(s) being served by the program (Lau, 2006).

Within the context of a facilitator's racial and/or cultural group, there are two dimensions that may influence implementation quality: (1) the facilitator's race and/or ethnicity and (2) the facilitator's race and/or ethnicity in relation to their participants (e.g., Chinese facilitator with Japanese participants). However, the research underpinning the influence of a facilitator's race and/or ethnicity is relatively limited in both measurement and breadth. Many studies of implementation quantify facilitator ethnicity as white or non-white (e.g., Pas et al., 2015), primarily in response to smaller sample sizes in non-white facilitator groups (Little et al., 2013). The limited research in this area

suggests the relationship between a facilitator's race and/or ethnicity and implementation quality is unclear. For example, in a study of the implementation quality of substance abuse prevention program, Rohrbach et al. (2010) measured facilitator ethnicity as white/nonwhite and found no meaningful effect of facilitator ethnicity on program implementation. Conversely, in a study of an intervention program orientated towards preventing problem behaviors, Williford et al. (2015) found that minority facilitators delivered the program with lower quality and at a lower rate. In the instance of Williford et al. (2015), the program did not have a cultural adaptation for non-white groups.

The interaction between a program being culturally adapted and the race and/or ethnicity of a facilitator may act as a confound to implementation quality. For example, in a study of youth programs being delivered beyond the control of program designers, Moore et al. (2013) found that one of the most common reasons programs were not delivered as designed was due to a facilitator's poor cultural fit (e.g., the program was designed for an English speaking group, where Spanish is the facilitator's primary language) with a program and/or the facilitator's perception that a program was not a cultural match with that of the participants (e.g., program materials and marketing reflected African Americans where the program was being delivered to Chinese Americans). Similarly, in a study examining attitudes towards evidence-based practices (EBPs), including delivering programs as designed, Aarons et al. (2010), found lower levels of support for EBPs in non-white respondents. In partial contradiction to the findings of Moore et al. (2013) and Aarons et al. (2010), in a study of a family development program, Cantu et al. (2010) found a positive relationship between

51

adherence and minority status match, indicating that implementation quality can improve when cultural adaptations are present within a program. Paralleling this finding, in a study of an HIV prevention program, Dolcini, Catania, Gandelman, & Ozer (2014) noted that facilitator race influenced implementation quality. More specifically, facilitators of the same race as the participants tended to deliver more of the program content (i.e., higher implementation level).

In summary, there are two primary challenges regarding the relationship between a facilitator's cultural group and implementation quality: (1) the research exploring the combination of cultural adaptations, facilitator racial and/or ethnic group, and implementation quality is limited by both measurement and number of studies, and (2) the relationship among a facilitator's racial and/or ethnic group, that of the program participants, and implementation quality is also unclear. The lack of research in this area suggests that measurement efforts (See also Table 1) should be further explored in both culturally-adapted and non-adapted program contexts including the facilitator's race and/or ethnic group (measured beyond dichotomous approaches, e.g., beyond white/non-white where possible), the facilitator's race and/or ethnic group in relation to program participants' race and/or ethnic group (also measured beyond dichotomous approaches, e.g., beyond white/non-white where possible).

Facilitator Education and Implementation Quality.

In the context of implementation research, a facilitator's education can have dual meaning. The U.S. Census operationalizes education as educational attainment, specifically the highest degree of education a person has completed (e.g., high school

diploma, bachelor's degree) (US Census, 2015). However, within implementation

science, education level is frequently operationalized as the possession of a degree

beyond a bachelor's degree (Little et al., 2013), and/or area of specialization,

certification, or major (Williford et al., 2015). In a study of a problem behavior

prevention program, Wehby et al. (2012) illustrated this point; specifically, the authors

operationalized facilitator education as possessing either a bachelors or master's degree

and additional program related certifications. However, in some studies of

implementation there is a greater range of educational operationalization. For instance, in

a substance abuse prevention program study, Johnson et al. (2010) operationalized

education as either a bachelor's degree, some post graduate work, or a master's degree.

This broader range is occasionally increased by also including educational area of

specialization. In an examination of a youth science-achievement program, Desimone and

Lee-Hill (2017) operationalized education level in terms of the degree in relation to the

content (i.e., major) focus, specifically "Science, Technology, Engineering, and Math

(STEM)", "Education," or "Other." Desimone and Lee-Hill (2017) also measured

competency as part of their assessment, through a standardized content knowledge

assessment reflecting the program topics (e.g., biological science, physical science, and

geology). Desimone and Lee-Hill (2017) reflected the greater range of education levels

also provided in Johnson et al. (2010), specifically, measuring education level from

"some-college" to "doctorate."

In exploring the influence of facilitator education level on implementation quality,

there is a lack of clarity regarding what effect education level has on implementation

quality. In some studies, the differing dimensions of education (i.e., major, emphasis area, or number of years) can have differing effects. For instance, Domitrovich, Gest, Gill, Jone, Sanford-DeRousie (2009) operationalized education as: (1) educational attainment including, high school, some post high school, associate's degree, and 4-year degree (or more), and (2) as a certification, including a child development or teaching certificate. In this study, Domitrovich et al. (2009) found that certification did not predict implementation quality, but education level positively predicted implementation quality. Other studies of implementation quality have indicated the opposite (i.e., certification influencing implementation quality); for instance, in an examination of a program (titled Banking Time) orientated towards preventing problem behaviors, Williford et al. (2015) found that when teachers possessed an early childhood education specialization they delivered the Banking Time program more often. However, Williford et al. (2015) found education level did not meaningfully influence implementation quality. While these studies found some effect of education on implementation quality, some research suggests that education level may have no effect on implementation quality. This point is illustrated by Little et al. (2013), who explored facilitator level characteristics and their influence on implementation quality and found no effect for advanced degree (i.e., above bachelor's degree) or teacher specialization (i.e., health focus) on implementation quality. Similarly, in a study of a classroom management program, Wanless et al. (2015b) operationalized education as Master's degree or not (i.e., bachelors) and found no meaningful effect of either level of education on implementation quality. Further illustrating this null effect, Rohrbach et al. (2010), operationalized education as the

facilitator's degree area of specialization, comparing physical education to health education, and found no effect on implementation quality.

Perhaps unsurprisingly, as much of implementation research is school-based, and thus most programs investigated are implemented by teachers who generally possess at a minimum a bachelor's degree, there is a frequent lack of range within education as a measured variable. Also, similar to other facilitator characteristics, the term "education" can have multiple meanings in the context of implementation science, and is used as a synonym for training and/or experience in some implementation investigations (Berkel et al., 2011). This aggregation of facilitator education may explain the effect (or lack thereof) of education on implementation quality. Additionally, these findings indicate that at a broader level the lack of range embedded within the measurement of education may be a cause of inconsistent findings. For example, within Domitrovich et al. (2009), the authors utilized a greater range of education levels and uncovered an effect of education on implementation quality. This positive result suggests that the usage of a dichotomous measure of education (e.g., Rohrbach et al., 2010) may suppress education's effect on implementation quality. In future studies of implementation quality, education should continue to be captured to determine in which circumstances and programs facilitator education can enhance implementation quality. However, where possible, a greater range of educational levels should be explored; specifically, facilitator education should be measured in terms of (1) terminal degree attainment, (2) number of years to obtain each degree, (3) where applicable, the time to obtain certification, (4-5) if current certification level requires prior certification level (i.e. prerequisite) and expired certifications, (6)

Grade Point Averages (GPA) and related performance measures from all terminal degrees, (7) progress (in time) towards non-conferred degree, and (8) unrelated and/or non-program-required certifications.

Facilitator Support and Implementation Quality.

Facilitator program support, facilitator buy-in, and perceived organizational climate can have a powerful effect on implementation quality (Baker, Kupersmidt, Voegler-Lee, Arnold, & Willoughby, 2010; Berkel et al., 2011; Durlak & DuPre, 2008). This facilitator support for a program influences a facilitator's quality of delivery (Fagan et al., 2008), how they promote a program to their peers (Ennett et al., 2011), and how receptive they are during program training (Moore et al., 2013). Within the context of facilitator support for a program there are multiple sub-dimensions including the facilitator's "buy-in" to a program and the facilitator's perception of organizational climate (Hoy & Feldman, 1987; James & James, 1989). These two sub-dimensions share a degree of crossover in how they are interpreted and operationalized in studies of implementation (Williford et al., 2015), but also reflect a degree of separation in level of unique effect on implementation quality (Baker et al., 2010). Specifically, a facilitator's program buy-in reflects their perceived "need for the innovation, believe the innovation will produce desired benefits, feel more confident in their ability to do what is expected (self-efficacy), and have the requisite skills" (Durlak & DuPre, 2008, p. 336). Conversely, perceived organizational support/climate reflects the facilitator's perception of organizational climate (e.g., leadership quality, openness to communication and

56

feedback, administrator support) and the degree to which organizational leadership is "bought-in" to the program (Dusenbury et al., 2003).

The relationship between facilitator buy-in and implementation quality is unclear. In some studies, there is a link between lower levels of facilitator buy-in and implementation quality. For instance, in a study of youth programs being delivered beyond the control of program designers, Moore et al. (2013) found that one of the most common reasons programs were not delivered as designed was a lack of facilitator support and belief that a program would achieve the desired effects. Further, within an implementation assessment of a substance abuse prevention program, Rohrbach et al. (1993) found that facilitators with higher levels of buy-in (i.e., enthusiasm toward towards the program) had higher quality implementation. Additionally, in a study of a community-based prevention program, Hill et al. (2007), found that the most common reason that facilitators did not implement programs as designed was due to their own disagreement with the program content in terms of their program's usefulness to the group being served, indicating that facilitator buy-in, when favorable towards the program, can positively influence implementation quality. Further evidence of the link between implementation quality and facilitator buy-in was provided by Morgan et al. (2016) who noted that the high degree of fidelity within their municipal youth program was likely due to high levels of facilitator buy-in.

However, some evidence suggests that the link between buy-in and implementation quality is less positive. For example, Malloy et al. (2015), measured facilitator buy-in to youth social and character programming as a dichotomous variable (1

= enthusiastic; 0 = cooperative) and found partial support for full implementation of the program and stronger support for implementation of supplemental program materials outside of designed curriculum. While Malloy et al. (2015) found only emerging (e.g., *p* ≤ .10) support for implementation quality and facilitator buy-in, additional studies have found a negative relationship. For instance, within a study of a drug abuse prevention program, Little et al. (2013), found that when facilitators went through a standard program training, their level of support (e.g., favorable beliefs in the program achieving desired outcomes) for the program negatively influenced implementation quality; more simply, higher levels of facilitator buy-in led to lower implementation quality. This contradiction is further illustrated by Williford et al. (2015), who hypothesized teachers (i.e., facilitators) with lower pro-program beliefs (e.g., that a child will benefit from the program) would have worse program implementation than their peers with higher program support. However, Williford et al. (2015) found that these lower scoring teachers actually implement more of the Banking time program. In other words, Williford et al. (2015) demonstrated that lower facilitator buy-in led to higher implementation quality.

This incongruence in the relationship between facilitator buy-in and implementation quality parallels the inconsistent evidence about the relationship between implementation quality and a facilitator's perception of their organization. As mentioned earlier, the characteristics of an organization can influence a program's implementation quality (Dane & Schneider, 1998). Correspondingly, a facilitator's organizational perception can influence implementation quality. At the facilitator level, this

organizational perception is a combination of four dimensions: (1) a facilitator's

perception of organizational *leadership* including the facilitator's level of trust and belief

in organization leaders goal orientation, (2) the facilitator's level of *work-related stress*,

(3) the facilitator's perception of job *autonomy and challenge* in their role, and (4) the

quality of relationships (e.g., warmth and friendliness) that facilitators have with their

peers (James & James, 1989).

Some evidence suggests that facilitators' views towards their organization can

have a meaningful positive effect on implementation quality (Baker et al., 2010). For

instance, in a review of the factors that influence implementation quality, Dusenbury et

al. (2003) shared that facilitator perceptions of organizational climate (e.g., leadership

quality, openness to communication and feedback, administrator support) were positively

related to implementation quality. However, there is evidence that suggests the opposite

effect. Specifically, in an examination of a youth-at-risk character development program,

Malloy et al. (2015), found a negative relationship between a facilitator's autonomy and

challenge in their role (termed participatory decision-making) and their quality of

delivery of the program, suggesting that the when facilitators perceived more autonomy

and challenge in their role (a subdimension of organizational climate) the more facilitator

quality of delivery suffered. Malloy et al. (2015) suggested when organizations provide

support for facilitator decision-making, the facilitator may actually deliver less of the

program due to their decreased level of perceived organizational oversight. While Malloy

et al. (2015) suggested their finding of a negative relationship between perceived

organizational climate and implementation quality was surprising, a study conducted by

Pas et al. (2015) indicated similar results; specifically, as facilitators increased in their perception of organizational climate and their comfort within the organization, they decreased in implementation quality. Furthermore, some evidence suggests a null relationship between facilitator perceptions of organizational climate and implementation quality. Illustrating this point, Rohrbach et al. (1993) examined facilitator perceptions of encouragement by organizational leadership to deliver the program as designed and found no influence of facilitator perceptions on implementation quality.

In aggregate, this research suggests that further exploration of the relationship among facilitator buy-in, organizational support, and implementation quality is warranted. Moreover, similar measurement and analyses problems to those mentioned earlier are also evident within the facilitator perceived organizational support and buy-in constructs, specifically, the dichotomization of continuous level variables (e.g., Malloy et al., 2015) and the use of non-traditional p-values (e.g., $p \le .10$ rather than $p \le .05$) to justify an effect (Baker et al., 2010). However, despite these limitations, there are psychometrically valid measures available for assessing both facilitator organizational climate and buy-in. An exemplar of these potential measures is highlighted in Baker et al. (2010), where the researchers examined the relationships between facilitator characteristics and implementation quality (referring to preschool teachers); specifically utilizing previously developed measures of facilitator buy-in (e.g., attitude toward the intervention, self-efficacy) and facilitator perceptions of organizational climate (e.g., perception of work environment and center director). Baker et al. (2010) indicated mixed support regarding relationships among implementation quality, buy-in, and organizational

support. However, the approach of Baker et al. (2010) demonstrated within the context of implementation assessment that the measure of these variables can inform researchers of factors influencing implementation quality. Further studies utilizing measures of facilitator buy-in have also demonstrated their predictive value towards implementation quality (e.g., Gagnon et al. 2015b; Gagnon and Bumpus, 2016). Table 1 provides recommendations for future measurement of facilitator buy-in and organization support.

Facilitator Age and Implementation Quality.

The relationship between a facilitator's age and implementation quality suggests that age does not have a measurable effect on implementation quality. For instance, Pas et al. (2015) found no link between a facilitator's age and their implementation quality. In this study, the researchers operationalized age as a dichotomous variable, 20-30 vs. 31 or older. Similarly, in a study of a substance abuse prevention program, Little et al. (2013) found no effect of facilitator age on implementation quality. Additionally, comparable results were found outside of school settings; for example, in a study of community-based program implementation, Hill et al. (2007) found that facilitator age had no effect on implementation quality. While these studies suggest that a facilitator's age does not have meaningful effect on implementation quality, other studies provide conflicting evidence. In a study exploring the effect of a socio-emotional development program Domitrovich et al. (2015) found that younger teachers tended to implement the program more frequently (i.e., dosage); however, there were no links between this enhanced implementation quality and program outcomes, indicating that the effect of age to increase implementation was not meaningful overall.

The challenge within implementation research is that many facilitator characteristics, such as age, are poorly measured to the point where unique effects of individual characteristics are likely suppressed (Century et al., 2010). For example, in the implementation studies of Pas et al. (2015) and Domitrovich et al. (2015), facilitator age was measured as "young versus old;" more specifically, facilitators were divided into a young age group (e.g., 20-30 years of age) and an older age group (31 or older). This grouping is problematic for a number of reasons, including the reduction of a continuous level variable (e.g., age in years) into a categorical one, and that with this style of analysis, a 29-year-old and 20-year-old facilitator are treated as equivalent. To the author's knowledge, there are no investigations exploring the interaction of facilitator age and program participant age and the effect of this interaction in terms of implementation quality. As highlighted in Table 1, future investigations of implementation quality should move beyond the dichotomous approaches reflecting in many studies, or provide a theoretical justification on why a grouping of facilitators (20-30 versus 31+) is prudent in the context of implementation. An examination of the potential combined effect similar findings to those experiments utilizing minority matching techniques (e.g., Cantu et al., 2010).

Table 1.

Recommended Operationalization's of Facilitator Traits and Characteristics

Facilitator Trait or Characteristic	Potential Measurements	Unit Measurement Recommendation
Facilitator Experience	(1) Broad level of experience providing programs both similar and dissimilar in nature	Hours, days, weeks, and/or years
	(2) Specific experience delivering the program of interest	Same as above
	(3) The amount of time a facilitator spent in the program as a participant	Same as above
	(4) The frequency and/or number of programs the facilitator has delivered	Discrete number of sessions and/or programs (e.g., 21 sessions)
Facilitator Training	(1) Broad level of training outside of program	Program dependent: time (e.g., hours, days, weeks) and/or dosage (e.g., total trainings)
	(2) Specialized (program specific) training	Same as above
	(3) Facilitator competency (pre- and post-training)	Program dependent: test score, ability assessment
	(4) Trainer status (e.g., internal or external employee) and trainer experience	Internal or external (0 no, 1 = yes), experience, program dependent: time (e.g., hours, days, weeks) and/or dosage (e.g., total trainings)
	(5) In-Service and follow-up training	Program dependent: time (e.g., hours, days, weeks) and/or dosage (e.g., total trainings)
Facilitator Gender	Facilitator Gender	Female, Male, Non-binary (where applicable)
	Facilitator Gender in Relation to Participant Gender	Same as above

Table 1 (Continued).

Recommended Operationalization's of Facilitator Traits and Characteristics

Facilitator Trait or Characteristic	Potential Measurements	Unit Measurement Recommendation
Facilitator Education	(1) Terminal degree	Terminal degree and number of years or semesters in pursuit of degree
	(2) Time to complete each degree	Number of years or semesters in pursuit of completed degrees
	(3) Current level and number of certification	Ascending level of certification (where comparable) and/number of certifications
	(4) Time to complete certification	Certification dependent: time (e.g., hours, days, weeks) and/or dosage (e.g., total trainings)
	(5) Prior certifications and/or prerequisite certification to current certification	Number of past prior certifications (including those required to obtain more advanced certification level
	(6) Educational performance	GPA, class standing, related course performance
	(7) Progress towards non-conferred degree	Number of years or semesters in active pursuit of non-completed degrees
	(8) Unrelated or non-program-required certification	Certifications unrelated to program of interest

Table 1 (Continued).

Recommended Operationalization's of Facilitator Traits and Characteristics

Facilitator Trait or Characteristic	Potential Measurements	Unit Measurement Recommendation
Facilitator Support	Attitude[1] or Buy-in[2] Towards Intervention	Where possible extend range beyond 1-5 and treat variables as latent rather than composite (e.g., adding and averaging)
	Job Satisfaction[1]	Same as above
	Job Commitment[1]	Same as above
	Self-Efficacy[1,3] or Perceived Preparedness[2]	Same as above
Organizational Perception	Perceived Organizational Health[4,5]	Same as above
	Perception of Work Environment[1,6]	Same as above
	Perception of Organizational Leadership[1]	Same as above
Facilitator Age	Facilitator Age in Years	Number of Years
	Facilitator Age in Years in Relation to Participant Age	Same as above
Facilitator Cultural Group	(1) Facilitator race and/or ethnic group	To at least reflect the 7 categories of US Census (2010): (1) White, (2) Black or African American, (3) American Indian or Alaska Native, (4) Asian, (5) Native Hawaiian or Pacific Islander, (6) Hispanic Latino, and (7) two or more races.
	(2) Facilitator race and/or ethnic group in relation to participants (e.g., minority match)	Same as above

Note: 1 indicates Baker et al. (2010); 2 indicates Gagnon and Bumpus (2016); 3 indicates Sherer et al. (1982); 4 indicates Hoy and Feldman (1987); 5 indicates Pas et al. (2015); 6 indicates Jorde-Bloom (1996).

Facilitator Gender and Implementation Quality.

The relationship between a facilitator's gender and implementation quality is relatively unclear, partially due to the conflation of gender (a social construct) and sex (a biological construct). More specifically, some evidence suggests that a facilitator's gender can have a meaningful influence on implementation quality or program outcomes. For example, Lillehoj et al. (2004) found that male facilitators had better participant outcomes than their female counterparts, but facilitator gender did not have a meaningful influence on implementation quality. Similarly, in a study exploring attitudes towards Evidence-Based-Practices (EBPs), Aarons et al. (2010) found that the respondents sex did have a meaningful influence on attitude towards EBPs scores, with women scoring higher (i.e., more likely to utilize EBPs including implementing programs as designed). Furthermore, evidence suggests that the interaction between facilitator gender and participant gender may have a meaningful influence on implementation quality. Illustrating this interaction, Dolcini et al. (2014) found that facilitator gender in relation to participant gender led to lower rates of implementation quality, more so for female facilitators. That is, female facilitators had lower rates of implementation quality regardless of participant gender, but more so if their participants were also female. This finding suggests that when exploring the influence of facilitator gender on implementation quality, analyses should also include the gender of program participants.

While a few studies have suggested that facilitator gender can have a meaningful effect on implementation quality, others have found null results (Baker et al., 2010). For example, in a study examining the implementation of a youth substance abuse prevention

program, Rohrbach et al. (2010) found no effect of facilitator gender on implementation

quality or program outcomes. Paralleling this finding, in a study of a drug abuse

prevention program, Little et al. (2013) found no meaningful influence of gender on

implementation quality. Finally, in a study of community-based program implementation,

Hill et al. (2007) found that gender had no effect on implementation quality. The

differing study results indicate that gender does influence implementation quality in some

program contexts, but not in others, and thus suggests that gender remains an important

characteristic to measure when conducting implementation research at the facilitator

level. Furthermore, the results of Dolcini et al. (2014) indicate that studies exploring

implementation quality should also examine a facilitator's gender in relation to that of

program participants.

Discussion and Future Directions

As synthesized in this review, multiple factors can influence implementation

quality. In fact, "we are left with the impression that...almost everything seems to matter"

(Mihalic & Irwin, 2003, p. 312). The sheer number of potential influences are a common

reason that implementation research is not conducted as the process is prohibitive and

unlikely to provide clarity (Durlak, 2010). While complex, the study of implementation

illustrates the intersection of macro- and micro-level factors, corresponding effects on

implementation quality, and program outcomes. As evidenced in figure 2, figure 4, and

other conceptual models of implementation (e.g., Berkel et al., 2011; Carroll et al., 2007),

a multitude of correlated factors influence implementation quality; in many ways, these

correlations parallel those of ecological systems models (e.g., Bronfenbrenner, 1994)

where causality between factors (e.g., community and organization) is often recursive rather than "input – output" driven. More simply, it seems that there is a level of complexity among the factors that influence implementation, and correspondingly the influence of many facilitator characteristics is dependent upon the levels of others. Furthermore, this chapter highlights how the study of implementation can never truly be "done" due to the nearly unlimited factors that could be studied under the umbrella of implementation science (Barrera et al., 2016).

However, despite the multitude of factors that influence implementation quality and the corresponding difficulty in conducting implementation research, a robust foundation of evidence suggests that when programs are implemented as designed, participants generally experience better outcomes (Durlak & DuPre, 2008). Thus, it is critical that efforts towards understanding the factors that influence implementation quality continue. In this vein, the goal of this manuscript was to untangle and critique some of the complexity embedded within implementation research. This was accomplished by narrowing to the facilitator level to determine which facilitator characteristics, traits, and/or beliefs inhibit or promote implementation quality. The results of this review (See Table 1) suggest at least 23 characteristics or traits (e.g., facilitator education, experience, gender, training) and seven beliefs (e.g., buy-in, perceived organizational health) contribute to implementation quality. However, between the studies shared in this review, there was a "consistent inconsistency" in terms of the influence (or lack thereof) of each characteristic, trait, and/or belief on implementation quality. More simply, the relationship among facilitator characteristics, traits, and beliefs

with implementation quality remains unclear. There are many potential reasons for the inconsistency in findings across studies, but there are likely three primary culprits: (1) a genuine lack of effect of a specific characteristic, trait, or belief on implementation quality, (2) measurement issues of facilitator characteristics, traits, beliefs, and implementation quality, and (3) a lack of careful theoretical explication of these constructs, so as to better inform measurement and interpretation. The first explanation is the simplest one; that is, in some contexts and programs, a facilitator's individual characteristics, traits, or beliefs may not matter. This lack of effect could be due to the regimented design of a program, the quality of training provided to facilitators, or other external influences.

However, while the first explanation would be ideal for several reasons (e.g., program stakeholders being driven towards parsimonious explanations for program outcomes or lack thereof), it is unlikely outside of the simplest programs this would be the case. As evidenced throughout this review, a more likely reason for this inconsistency across studies is the second explanation. The varying results suggest that there are measurement issues embedded within the measurement of implementation in terms of the dimensions that are measured (Durlak & DuPre, 2008) and those that are not. Further, the "compositing and averaging of implementation quality" likely has resulted in suppression of unique relationships within implementation. In other words, there may be an oversimplification of the variables that influence implementation quality and thus gaps in our understanding of relationships and effects. This oversimplification of variables (e.g., dichotomization of facilitator experience into novice and experienced) likely led to both

suppression and/or magnification of effects. In other words, many studies of implementation provided in this review may have committed Type I and Type II errors due to these analytic choices. Perhaps ironically, due to an effort for parsimony in an inherently complex research area (e.g., implementation science), the modification of continuous variables to categorical levels led to overly simplified explanations of facilitator effects on implementation (e.g., experience did not matter).

A further challenge highlighted in this review, and suggested in Figure 4, is that facilitator characteristics, traits, and beliefs are in effect "siloed" within many studies. More simply, training and experience are treated as unidimensional and non-recursive; however, logic would imply in many cases that facilitator characteristics would interact with each other. For instance, it would be reasonable to assume that as facilitator program specific experience increases so does their broad experience. This increase in correlation (e.g., towards multicollinearity) could likely suppress the unique effect of either variable. Moreover, examination of relationships between facilitator characteristics and beliefs (e.g., experience levels and buy-in) could highlight other facilitator level dimensions that promote or inhibit implementation quality. Indeed, this review identified many facilitator-level characteristics, traits, and beliefs that could promote or inhibit implementation quality, but "research on community-driven implementations is sparse" (Cantu et al., 2010, p. 27), suggesting there are likely more factors at the facilitator level that influence implementation quality (e.g., facilitator competence) and thus a great deal of work remains (Schoenwald et al., 2011).

Ultimately, it is the facilitator's responsibility to ensure that a program is delivered as designed (Wanless et al., 2015b) and thus future investigations of implementation quality should continue to explore the effects of facilitator characteristics, traits, and beliefs on implementation quality both independently and in the aggregate. It is likely there is a unique mix of facilitator characteristics, traits, and beliefs between programs and contexts, but the identification of these mixes will only improve program design and implementation quality and thus program outcomes. While there are many reasons to explore the effect of the facilitator on implementation quality (e.g., staff selection, training, retraining), ultimately the primary motivation should continue to be ensuring the very best outcomes for program participants.

CHAPTER THREE: DEVELOPING AN ASSESSMENT OF FACILITATOR INFLUENCES ON PROGRAM IMPLEMENTATION

In the past few decades, recreation and leisure sciences have shifted towards evidence-based practice in both program design and assessment. This evidence-based shift is in recognition of the shortcoming embedded within some past recreation and leisure research (Kaczynski & Henderson, 2007) due to an external poor regard of leisure research in the broader social sciences and an internal challenge relating to inconsistent development as a field methodologically "conflicting conceptual and methodological development" and "inconsistent terminology" (Henderson, 2016). This shift towards "conscientious, explicit, and judicious use of current best evidence in making decisions" is also partly due to increasingly robust funder expectations and a deeper recognition of the complex needs of those constituents served by recreation and leisure programs (Sackett, Roseberg, Gray, Haynes, & Richardson, 1996, p. 72). Furthermore, this change is echoed in the focus of program designers and evaluators utilizing more methodologically rigorous procedures to design, implement, and understand programmatic success (Witt & Caldwell, 2010), including randomized control trials and quasi-experimental designs. This growth in research quality has allowed the field of recreation and leisure sciences to catch up to many in the broader social sciences. One notable gap persists relating to our understanding of program implementation within leisure and recreation contexts (Mainieri & Anderson, 2015a; Morgan et al., 2016; Tucker & Rheingold, 2010). More specifically, much of leisure, experiential, and recreation (LER) research is orientated towards understanding human growth and
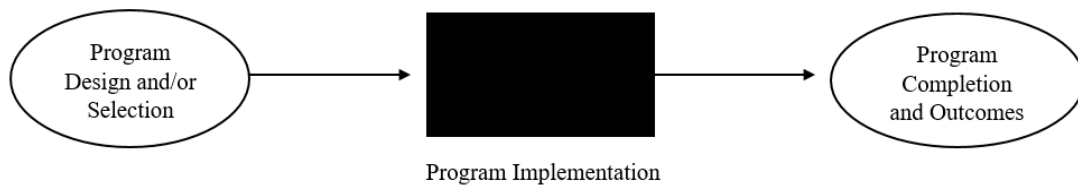
Figure 5. The black box of program implementation, adapted from Mainieri (2016).

development resulting from attending a particular program or series of programs, but not

necessarily "how" these programs achieve their desired outcomes, known more simply as

the *black box* (Ewert, 1983). As presented in Figure 5, this black box represents the gap

between the design and/or selection of a program and the outcomes thought to result from

the program (Mainieri, 2016). In this manuscript, the terms "program" and "intervention"

are utilized interchangeably, due in part to much of the extant literature referenced within

this study also utilizing these terms interchangeably.

This study enhances understanding of the black box within LER through an

investigation of the implementation process and its corresponding assessment by

exploring one of the elements thought to contribute to program implementation, the front-

line staff person responsible for delivering the program (hereafter referred to as the

facilitator). To this end, the sections below: (1) explore the elements theorized to

contribute to facilitator implementation quality at a general level, (2) review how these

elements have been previously assessed, (3) introduce situational judgement testing (SJT)

as a new method for assessing implementation quality, and (4) describe how this method

was applied to an LER context. Thus, the primary goals of this study are to introduce

SJTs to the field of LER and to develop SJTs that capture an understanding of the

facilitator influence on implementation quality. This will be accomplished by integrating

prior implementation research from the broader social sciences with the results from semi-structured interviews conducted with camp professionals responsible for program design and delivery. While discussed in more detail later in this manuscript, *SJTs* present respondents "with a brief scenario and then ask him or her to select the best choice or indicate what he or she would do" in the given scenario (Barrett et al., 2010, p. 447).

Factors Influencing Implementation

Program implementation is influenced by multiple aspects including organizational characteristics, characteristics of the community and participants being served by the program, characteristics of the program itself, and the skills and traits of the facilitators delivering a program (Gagnon, Franz, Garst, & Bumpus, 2015b). Figure 6 provides a visual representation of how these factors are theorized to collectively influence implementation quality and corresponding programmatic outcomes.

Organizational Characteristics. The characteristics of an organization providing a program can have a profound influence on the quality of a program's delivery, implementation, and outcomes. For instance, the motivation and/or rationale of an organization for choosing an external program (i.e., one developed by a third party) or developing one themselves is often based on the availability of resources (e.g., human, property, financial) as well as the organization's "buy-in" to providing programs due to the potential positive corresponding outcomes (Dane & Schneider, 1998). Additionally, the reasons programs are selected, designed, evaluated, and provided often vary, from the program being delivered to fulfill a compliance requirement with funding or accrediting organizations, to produce meaningful change for a given group, or somewhere in the
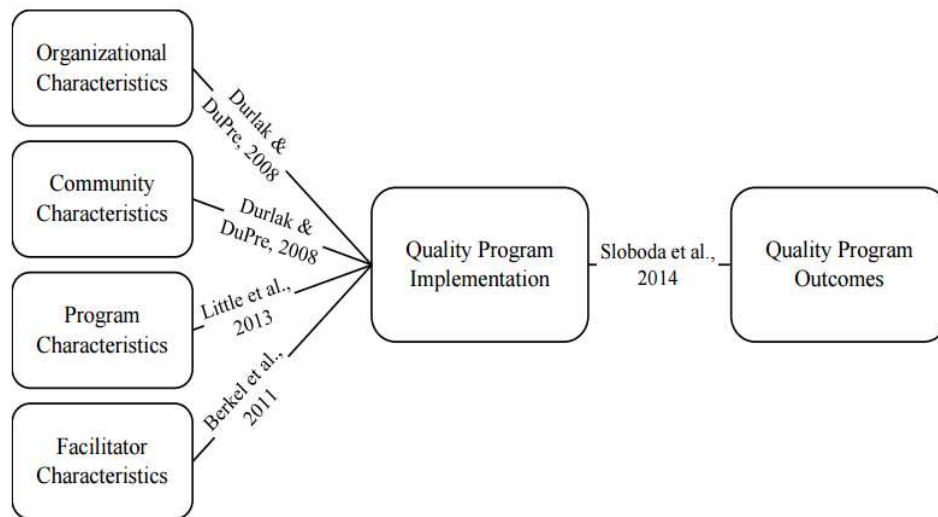
Figure 6. Conceptual model of the factors contributing to program implementation. Adapted from Gagnon et al. (2015b).

middle (Fredericksen & London, 2000; Kam et al., 2003). Correspondingly, the reasons

for conducting an evaluation of said program(s) may influence the quality and support

provided to a program by the organization charged with its design, delivery, and/or

assessment (Wandersman et al., 2008).

An organization's culture, orientation, and support for programmatic evaluation,

including implementation assessment, also may influence the quality of a program's

delivery and outcomes (Durlak & DuPre, 2008; Wandersman et al., 2008). For example,

in a case study examining organizational support for, and challenges to, the evaluation

process, Bechar and Mero-Jaffe (2014) highlighted how an organization's negative

perspective towards evaluation led to poor attendance of focus groups designated for

program assessment and potentially skewed results in later evaluation reports. Indeed, the

organization's lack of receptivity to evaluation and corresponding reporting led to the

evaluation team sharing that "staff members and students were reluctant to be

interviewed, and program staff was not prepared for, nor always informed about, our presence. In addition, our access to information was limited, and we were not updated about changes in activities" (p. 367). The results of this case study highlight the importance of positive organizational support for both an intervention and corresponding assessment, but also the corresponding challenges that may arise during the implementation process when this support is lacking.

Community Characteristics. Paralleling organizational characteristics, the characteristics of the community being served also play an important role in terms of implementation quality and the degree to which program outcomes are achieved. The support of a community for a program can vary widely based on stakeholders' perceived need for a program, their experience with similar programs and the organizational provider, the presence of a program champion (Elder et al., 2007), and the prevalence of current or recent programs within the community with differing and/or similar focus (Carroll et al., 2007; Wandersman et al., 2008). This program saturation, including the marketing and promotion of similar programs, can influence community receptiveness to an intervention, its implementation quality, and corresponding success (Lefebvre & Flore, 1988). Thus, when considering community-level factors that may influence program quality, it is important that an analysis of the potential program audience is also considered.

Another important characteristic of the community being served relates to *participant responsiveness* to a particular program, refers to the level of engagement, motivation, and interaction that participants demonstrate during a program defined as the

degree to which (James Bell Associates, 2009). The level of responsiveness and fit of a program at the community and at the individual program participant level may influence a program's implementation quality (Elder et al., 2007). In a conceptual framework exploring the factors influencing implementation quality, Carroll et al. (2007) indicated that participant responsiveness when low can negatively influence implementation quality as the facilitator may modify and/or adapt a program to enhance responsiveness levels. In an assessment of a school-based substance abuse prevention program, Ennett et al. (2011), proposed that participant responsiveness, can influence a facilitator's quality of delivery, thus compromising implementation, but also could influence other components thought to influence implementation including the facilitators charged with providing the program levels of competence, confidence, and adherence.

Beyond the buy-in of the community being served by a program and the responsiveness of participants, program evaluation capacity also can influence the degree to which a program is delivered as designed (Preskill & Boyle, 2008). If a community does not have the resources necessary to deliver a program as designed, nor the training or ability to evaluate its implementation or outcomes, it is unlikely that a program will be sustained nor achieve the desired effects (Preskill & Boyle, 2008). Thus, the ongoing assessment of a program's implementation quality is a core component of the evaluation process and the characteristics of the community providing it will influence both a program's quality, but also the quality of the assessment taking place.

Programmatic Characteristics. There is a high level of intersection and dependency both within and between the organizational and community levels identified

in Figure 6. This interdependency between levels is also reflected within the characteristics of the program, community, and organization, particularly when the program's complexity and corresponding resources for a program impact the level to which it is implemented (Mihalic et al., 2004). For example, if a program design requires the use of video, but the "real-world" setting of the program lacks electricity, this incongruence may compromise implementation quality, as not all of the program can be delivered as designed. Beyond the potential mismatch of design-based resources and the "real-world" setting, an additional challenge to the maintenance of high quality implementation relates to the program's complexity. Specifically, as program complexity increases, the level of efficacy in terms of program quality (to include implementation) generally decreases (Durlak & DuPre, 2008). This decrease in program quality and outcomes is often due to the restrictions and complexity embedded within a program's design and the ability of an organization, community, and/or facilitator to deliver a program in a "real-world" setting without the support of the program-development team (Carroll et al., 2007). To address the challenges related to complexity and delivery of a program in the real-world, the presence of a "program-champion" is also a key component to the achievement of program outcomes. For instance, when a program requires extensive training to ensure facilitators are able to fully implement the program as designed, this strain on organizational resources may cause training to be condensed or compromised. A program champion can help to highlight the need to maintain the organizations fidelity to designer recommended training levels. In a study examining factors that promoted or inhibited implementation quality within an in-school-substance

abuse prevention program, Ennett et al (2011) highlighted how schools with the most knowledgeable, supportive, and well trained program facilitators (i.e., program champions) tended to have better overall implementation and program outcomes.

An additional problem associated with programs being implemented as designed relates to a "cultural mismatch" (Castro et al., 2004). This mismatch can range from a program being designed for English speakers and delivered to Spanish speakers to a program designed for those with higher SES than the group being served (Castro et al., 2004). In the instance where programs will likely encounter a cultural mismatch, program designers should provide methods to alter a program so that it better fits the needs of the local population (Carroll et al., 2007). The need for a cultural match between a program and population, highlights the potential incompatibility of high-fidelity program models when the programs are diffused (provided) outside of the researcher's control. Well-designed programs should include aspects of adaptability to ensure a later cultural match, thus anticipating/preventing these challenges, rather than reacting to them (Elder et al., 2007).

Facilitator Characteristics. The facilitator of a program also influences the quality of program implementation (Berkel et al., 2011). Facilitator characteristics and traits, such as their experience level facilitating groups, quantity of program-specific and broader training, education level, age in relation to participant age, and gender, have all been demonstrated to have a relationship with the degree of implementation achieved within a program (Berkel et al., 2011; Caldwell et al., 2008; Cyr, 2008; Dusenbury et al., 2003; Little et al., 2013). Beyond these less malleable characteristics, a facilitator's

79

competence, the level of ability and skill to provide a program (Schoenwald et al., 2011) and *program buy-in*, the level of belief, support, and motivation a facilitator has when providing a program (Gagnon et al., 2015a), also have been found to influence implementation quality. Facilitators who report (or are observed to possess) higher levels of competence and/or program buy-in tend to have higher levels of implementation quality (Berkel et al., 2011; Dane & Schneider, 1998).

Domains of Implementation.

Beyond these characteristics, there are six distinct areas that the facilitator has control over and that influence the degree and quality of implementation present within a program: (1) adherence, (2) dosage, (3) quality of delivery, (4) participant responsiveness, (5) program differentiation, and (6) adaptation. At the facilitator level the "term *adherence* is often used synonymously with the concept of *fidelity* referring to the degree to which the program corresponds to the originally intended program (aka adherence, compliance, integrity, faithful replication)" (Durlak & DuPre, 2008, p. 329). Reflecting the operationalization and influence of adherence, *dosage* at the facilitator level refers to the total number of program sessions provided (Berkel et al., 2011). Essentially, a facilitator may skip or condense sessions of a program due to external issues (e.g., not enough time or resources), thus compromising the dosage of a program, and/or they may sequence the program differently than designed, compromising adherence. Dosage may also be influenced in combination at the organizational, community, and facilitator level (Bishop et al., 2013). where the number of program sessions (i.e., modules, components) is generally under the direct control of the

80

facilitator. However, the level of *program reach,* "the extent to which a program attracts its intended audience" (p. 1) may be a more organizational and/or community level factor as it is determined by factors outside of the facilitator's direct control, such as marketing, community support, and geographic scope (CDC, 2011).

The facilitator's quality of delivery also has an influence on program implementation quality and outcomes. As with many of the factors that inhibit or promote the level of implementation achieved in a programs delivery, *quality of delivery* is also a multi-dimensional concept, consisting of facilitator behaviors such as their enthusiasm, motivation, ownership, and buy-in when providing the program (Berkel et al., 2011; Spoth et al., 2007), facilitator competence and ability to provide the material reflecting their training and experience with the program specifically, and facilitator experience and training with similar programs (Ennett et al., 2011).

Of the six constructs, program differentiation is likely the factor most outside of the facilitator's control. This lack of control is highlighted in a review of 34 child-centered mental health programs conducted by Domitrovich and Greenberg (2000), who defined *program differentiation* as "attempts by the program evaluators to verify that only the experimental group received the intervention" (p. 195). Similarly, Berkel et al. (2011) defined program differentiation as the "distinctiveness of a program's theory and practices from other available programs" (p. 24). In both definitions, the authors examined implementation quality from the community and/or organizational level, areas generally beyond the control of the program facilitator.

An additional challenge embedded within the definition provided by Domitrovich and Greenberg relates to the idea that program participants have not received prior (or are not currently receiving) "treatment" with programs or services orientated towards similar goals. While random assignment as part of an initial study may be a mechanism to address this prior experience, random assignment is seldom sustainable or utilized beyond initial program implementation due to limited resources or capacity (Williams-White, Keonig, & Scahill, 2007). At the facilitator level, program differentiation may be affected in confluence with a well-designed program; specifically, a design that incorporates the potential of a program to have theoretical or practical similarities to others previously or currently provided within a community. For example, if while providing a team-building initiative a facilitator is informed by participants they "played that game last week" in an unrelated program, the facilitator should have (in a well-designed program) the training, competence, and a program design that reflects alternatives that still meet the theoretical and practical goals of the program of interest. In other words, the facilitator must have the competency (i.e., ability to recognize alternatives that meet the theoretical and practical goals of the program) and support (i.e., administrative approval to provide programmatic alternatives) to diverge from the initial program plan without compromising implementation quality, as they are still following the overarching and intended program plan. This divergence from the initial program plan does not compromise implementation quality, as these potential changes were embedded within the program design (Patton, 2008; Saunders, Evans, & Joshi, 2005).

While this modification due to issues with program differentiation may not represent a compromise to implementation quality, an adaption does. *Adaptation* at the facilitator level refers to the degree to which modifications are made to the program that are not captured in its design (Berkel et al., 2011). While adaptations may compromise program implementation quality, the relationship between programmatic adaptations and program outcomes is less clear. A deeper understanding of this relationship could highlight why adaptations positively, negatively, or do not influence outcomes and further provides another argument for the importance of conducting implementation assessments that include documentation of all programmatic adaptations (Bishop et al., 2013).). This lack of clarity is further highlighted by Moore et al. (2013), who in a study exploring a typology of programmatic adaptations shared, they "can occur within the context of high or low fidelity; that is, an adaptation may align with/enhance or misalign with/distract from the program's original design and/or theory of behavioral change" (p. 149). In summary, potential compromises to implementation quality and the rationale for why these compromises occurred should be documented through an assessment of implementation quality to not only capture *what* occurred, but also *why*. The information gained from this assessment can be utilized in future iterations of a program to correct errors and/or advances improvements to the design.

What is Implementation Assessment and Why Does It Matter?

Implementation assessment refers to determining the degree to which a program was delivered as designed, the reasons for omission, adaptation and/or deviation from a program's design, and in some circumstances the reasons for complete or partial

implementation (Duerden & Witt, 2012; Durlak, 1998). There are many considerations in understanding the motivation for conducting an implementation assessment. In the context of academic publishing, there is now a requirement in many of the "top" journals to include data on implementation (Durlak, 2015). Implementation assessment is also utilized to determine the potential causes, confounds, and compromises to program design (Durlak & DuPre, 2008), providing critical information when program providers and designers need to modify or redesign a program. As part of the program evaluation process, implementation assessment is particularly useful for two reasons. First, if a program is fully delivered as designed, but fails to achieve the desired outcomes, then there may be an issue with the program's design (Berkel et al., 2011). Second, if a program is not able to be delivered as designed due to external issues (e.g., not enough time, low participant engagement, poor facility conditions), then this may highlight a need to alter the program to better suit the needs of the community it is intended to serve (Dane & Schneider, 1998). Beyond the rationale of program improvement, there is evidence that the mere presence of an implementation assessment component as part of a larger program evaluation strategy leads to better programmatic effect. For example, in a meta-analysis of 542 studies conducted by Durlak and DuPre (2008), program evaluations that included an aspect of implementation assessment achieved positive programmatic effect sizes up to three times greater than those who did not conduct implementation assessments. This does not suggest that conducting an implementation assessment will simply enhance outcomes, rather the presence of implementation

assessment may reflect other elements within an organization that facilitate "better" programs.

Beyond the motivations for conducting implementation assessment, this investigation is also important for maintaining or establishing an evidence-based approach toward our understanding of our program's effect. For instance, if a researcher states that program XYZ was successful based on the participant's achievement of the desired outcomes, but does not assess for the program's implementation, then this statement may not be accurate. The researcher knows that *a* program was successful, but *they have insufficient evidence* indicating *their* program was successful (Sloboda et al., 2014). This potentially incorrect statement highlights that when program stakeholders are making statements about a program's efficacy, a corresponding investigation of the quality of program implementation is also necessary to fully support the accuracy of their statement (Fixsen et al., 2009). In summary, without the evidence an implementation assessment provides underpinning statements made to the veracity of a program, stakeholders may risk Type 1 Error (e.g., misstating that a program is effective when it is not properly implemented) and/or Type 2 Error (e.g., inferring that a program was ineffective when it was not properly implemented).

The Assessment of Implementation

The assessment of implementation typically takes one of three paths: (1) the use of indirect measures, including: self-reports, participant reports, participant interviews after a program, and daily logs/journals (Gresham, 1989; Mainieri & Anderson, 2015b), (2) the use of direct measures, such as program observations by trained observers

(Domitrovich & Greenberg, 2000), and (3) hybrid strategies where direct strategies like observations are combined with indirect strategies such as self-reports (Hansen, 2014). Each of the three assessment options have strengths and limitations. For example, persons providing self-reports may inflate their level of adherence to a programmatic design due to social desirability (e.g., the need to seem competent in their role), and conversely those being observed may adhere more strictly to a program than normal due to the presence of an observer (Gresham, 1989). Additionally, the funding needed for observations of program delivery (i.e., training and labor costs of the observers) may be excessive for organizations with limited resources (Fredericksen & London, 2000; Kam et al., 2003). As noted in Dane and Schneider (1998), when utilizing a direct strategy, a better predictive relationship between implementation quality and outcomes is present when the observer is independent/unaffiliated with the organization they are observing. As direct strategies may already be prohibitive for low-resourced organizations, the additional costs of an independent observer may introduce further challenges to limited resource pools.

Program implementation assessment methods still reflect a formative stage when compared to more traditional outcomes-based assessments in the broader social and prevention sciences (Hansen, 2014; Sloboda et al., 2014). Additionally, due to the exploratory/emerging nature of implementation measurement within the social sciences and potential limitations of some measures, alternative strategies have emerged including random brief observations of programs taking less than five minutes (Pettigrew & Gagnon, 2016), the use of video in lieu of direct observation (Pettigrew & Gagnon,

2016), the use of peer observation as self-report and/or journaling (Mainieri & Anderson, 2015a; Pettigrew & Gagnon, 2016), and self-reports of pro-implementation beliefs compared with/or in addition to self-reports of adherence (Gagnon, 2014; Gagnon & Bumpus, 2016). The preponderance of evidence suggests that given the relative newness of current implementation assessment approaches, implementation science is still ripe for exploration using a variety of alternative methods to both assess and predict implementation quality (Gresham, 1989; Hansen, 2014). One promising approach, informed by industrial organization psychology research, may be the use of situational judgement testing to predict implementation quality.

Situational Judgement Testing

Situational judgement tests (SJTs) historically are designed and implemented to measure and/or predict an employee's (or potential employee) performance within common work settings (McDaniel & Nguyen, 2001; Russell, 2010). At a basic level SJTs present respondents with a scenario they may encounter in their work and possible reactions/solutions to the scenario (McDaniel, Hartman, Whetzel, & Grubb, 2007). SJTs are generally utilized for personnel selection and promotion assessment in fields such as law enforcement, human resources, finance, and sales (Barrett et al., 2010; Russell, 2010; Schmit & Strange, 2010). While to the author's knowledge there are no studies examining the usage of SJTs within LER settings, in the broader field of organizational psychology there is a strong body of evidence suggesting SJTs are useful tools to predict future job performance as part of the hiring process (Chan & Schmitt, 2002). Indeed, when compared to more general assessments of personality and behavior (e.g., the five-

factor model of personality assessment) SJTs have been found to be less subject to faking (McDaniel et al., 2007) and better predictors of later job performance (Chan & Schmitt, 2002; Russell, 2010). The use of SJTS as predictors of performance is also advantageous over personality and behavior assessments, as "most job performance situations are complex, good judgment in these situations is likely to be a function of multiple, more narrowly defined traits and abilities" (Chan & Schmitt, 2002, p. 233) which are captured in the use of SJTs.

There are two general categories of SJTs: knowledge assessment and behavioral tendency assessment (McDaniel & Nguyen, 2001). In *knowledge-assessment-SJTs*, respondents are asked to "select the correct or best possible response or to rate the effectiveness of responses" (McDaniel et al., 2007, p. 64). For example, a respondent may be asked when teaching a group how to canoe the order in which they would present a lesson (e.g., communication, paddle strokes, personal-floatation-device usage). Conversely, in *behavioral-tendency-SJTs*, respondents are asked to "select the response that represents what the respondent would likely do or to rate the likelihood that they would perform an action" (McDaniel et al., 2007, p. 64). For instance, the respondent may be asked how they would react and manage if a participant in a canoe course refused to wear a personal-flotation-device.

In this study, behavioral-tendency SJTs are utilized, as the goal of this paper was to develop global measures of implementation behaviors. More specifically, the goal was to assess factors that may influence/predict implementation quality across LER programs, often with differing goals and outcomes. In this study, knowledge-assessment orientated

SJTs are too specific and not related to the constructs of interest of this study (i.e., adherence, dosage, quality of delivery, participant responsiveness, program differentiation, adaptation). Thus, this study explores the behaviors that may influence implementation quality, not the steps that make up the actual program process, as these steps vary from program to program. This more general approach is congruent with prior SJT research, specifically relating to implicit trait policy. At the basic level, *implicit trait policies* describe "inherent beliefs about causal relationships between personality traits and behavioral effectiveness" (Whetzel & McDaniel, 2009, p. 190). In this study, these beliefs are examined through an investigation of how a facilitator's judgement in a given scenario reflects pro-implementation attitudes. The goal was to utilize these responses to predict future "behavioral effectiveness" [i.e., behaviors (and reactions) that may influence the quality of program implementation]. Within behavior orientated SJTs there are several factors to consider ensuring that the SJTS reflect a reliable and valid measure of the constructs of interest (Whetzel & Reeder, 2016).

Validity of Situational Judgement Tests

Within the development of SJTs there are two primary types of construct validity to consider: (1) the validity within the scenario presented as relevant and applicable to the respondent and (2) the validity of the response choices available to the respondent (McDaniel et al., 2007). In both instances, *construct validity* refers to whether the SJT effectively measures what it is intended too (Drasgow, Nye, & Tay, 2010). To ensure construct validity within SJTs, one of two strategies are typically employed: (1) the use of subject-matter-experts (SMEs) or (2) a construct-orientated approach (Russell, 2010).

With the first approach, SMEs provide work samples or vignettes that reflect successful job performance. This information is then integrated into scenario(s) where the researcher/SJT developer(s) produces SJTs reflecting this information. In the second approach, researchers/SJT developers produce "draft" SJTs that reflect current research and areas identified to predict work performance which are then audited, reviewed, and/or modified by SMEs to ensure they reflect the language and approach of the organization and/or community of interest (Chan & Schmitt, 2002; Russell, 2010).

The utilization of SMEs in the development SJTs is not without limitation, as SMEs are often mid-to-senior level employees and therefore may be out of touch with the front-line or less experienced worker's experience or perspective (Russell, 2010). Additionally, some SMEs may be highly exceptional employees and thus may not reflect the "average" employee's view or experience (Schippmann, 2010). Finally, SMEs may only have a view that reflects "the right-way" to complete tasks and/or react to a scenario, and consequently lack the ability to consider alternative strategies to address a situation that may reach a similar desired outcome (McPhail & Stelly, 2010). This "right-way" is highlighted in the work of Weekly and Ployhart (2016), who in an investigation of why SJTs may fail to accurately predict performance, as highly qualified and/or experienced employees may not perform as expected on measures of interpersonal skills. This finding demonstrated how, in some cases, an SME may no longer possess the ability to identify all skills necessary to perform in their role, in spite of the ability to perform their job at a high-quality level.

Despite these limitations, SMEs can be used to develop and produce valid SJTs. For example, SMEs can be asked to describe their day-to-day work environment, common problems, and challenges they encounter to a researcher, who then integrates their responses with those from other SMEs to build scenarios that a prospective employee may encounter (McPhail & Stelly, 2010). However, this approach is typically utilized to produce knowledge- or task-based SJTs and thus may not generalize to other dissimilar work environments (Russell, 2010; Schippmann, 2010). More simply, the resulting SJTs may reflect a specific end goal (e.g., successful operation of a particular make/model of firetruck ladder), but may not provide adequate insight to the decision-making process of the person(s) completing the SJT. "For example, tasks directly related to putting out fires would serve as adequate operationalization of firefighters' job performance because these tasks constitute the core technical proficiency of the firefighter job" (Chan & Schmitt, 2002, p. 235). In other words, the use of firefighter technical skills may be helpful to understand their performance in terms of skill level, but this approach may not reflect the firefighter's ability to be a public/municipal employee or contributing member of a firefighter team.

Van Scotter, Motowidlo, and Cross (2000) refer to this difference as *task* versus *contextual* performance, where task performance (e.g., putting out a fire) is referred to as the "can do" performance, and more nuanced behaviors (e.g., motivation, interpersonal, psychological) reflects a "will do" performance. In the contextual scenario, there may be multiple "right" answers, when compared with the task scenario in which there may be only one right answer (Chan & Schmitt, 2002). In this study, where the goal is to utilize

SMEs to assist in the development of situations and responses that may promote and/or inhibit implementation quality, there could be multiple correct choices/judgements for the person(s) in the facilitator role to either sustain or compromise implementation quality (Patton, 2008). In summary, the use of SMEs to develop SJTs can result in quality measures of task performance, but SMEs may not be as critical in the development of SJTs reflecting contextual performance.

An alternative approach in the utilization of SMEs to develop SJTs is to utilize their expertise as a criterion validity check to previously developed scenarios. In this instance *criterion validity* refers to the degree to which a response choice is congruent with actual behavior that is likely to take place because of a given scenario (Sackett, 2010). This approach also allows for response choices that are likely to take place rather than choices that are clearly unrelated or incorrect. In this study, construct-orientated SJTs were developed utilizing definitions (i.e., constructs) from previous implementation research (Russell, 2010). More specifically, six scenarios were developed based on the work of Berkel et al. (2011), Hansen (2014), and Dusenbury et al. (2003; 2004), who suggested that at the facilitator level six factors (see Table 2 for definitions) may directly influence implementation quality: (1) adherence, (2) dosage, (3) quality of delivery, (4) participant responsiveness, (5) program differentiation, and (6) adaptation. In the sections below, the process utilized for the development and refinement of the six SJTs produced in this study is presented.

Method

As mentioned earlier the goal of this study was to develop, design, and implement SJTs that globally reflect factors that may influence implementation at a broader facilitator level, rather than program-specific factors within common youth-centered LER contexts. The development of behavioral-tendency SJTs reflects a nine-step process culminating in the analysis of the SJT results (See Figure 7). The development of the SJTs, including their refinement and future directions is described below.

Development of Situational Judgement Test Scenarios

As discussed earlier through a review of the implementation literature, six factors were identified that explicitly contribute to implementation quality at the facilitator level. Six scenarios were designed (See also Table 2) that reflect this literature, using an approach congruent with behavioral-tendency-SJTs (McDaniel & Nguyen, 2001), where the respondent is presented with a situation and choice(s) to assess how they would behave given specific information.

Subject Matter Expert Identification and Recruitment

A key component of well-designed SJTs is the validity underpinning their development (Whetzel & McDaniel, 2009). SJTs must be relevant to the audiences they are intended to serve, in the case of this study the intended audience includes SMEs and later program facilitators. As such, in this study, SMEs in the residential summer camp industry were utilized to further develop the SJTs presented within Table 2, including providing response options. Beyond the refinement of scenarios and development of response choices, SMEs were also utilized to resolve points of ambiguity within the

93
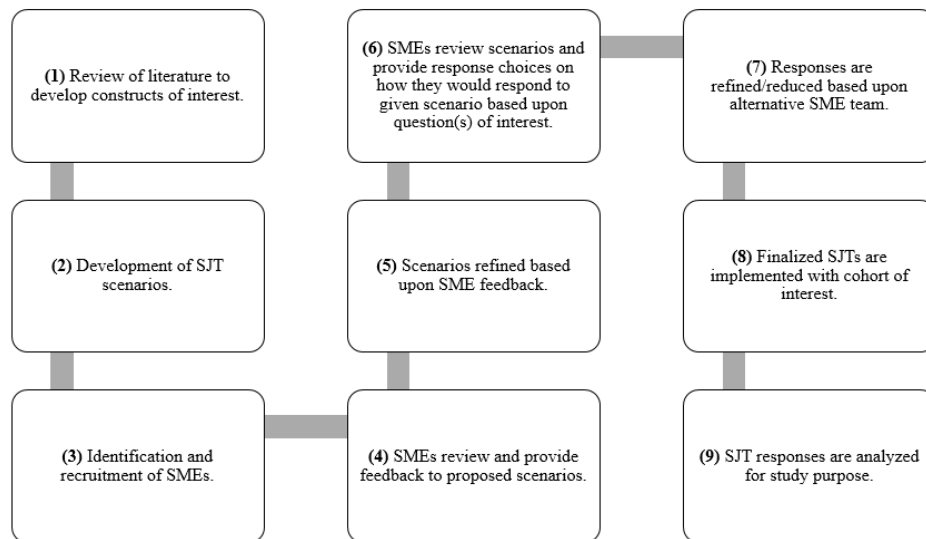
Figure 7. SJT development process.

scenarios, including response instructions prior to implementation with broader
audiences. This approach is congruent with recommendations for developing valid
measures in the reviews presented by McDaniel and Nguyen (2001) and Whetzel and
McDaniel (2009). To ensure that SMEs were appropriately qualified for inclusion within
this study, five screening criteria were utilized: (1) SMEs must be responsible for the
design of programs in a youth program setting, (2) SMEs must have at least three years of
experience as a director or supervisor in residential camp, (3) SMEs must have directly
delivered programs to youth (ages 8-17) within the last 3 years, (4) SMEs must be willing
to complete a follow-up interview, and (5) SMEs must be willing to review and complete
a 20-minute questionnaire. These screening criteria reflected similar studies addressing
the strengths and limitations of the use of SMEs and how their selection should proceed
within the development of SJTs (McPhail & Stelly, 2010; Russell, 2010).

Table 2.

Operational Definition & Situation (Pre-Subject Matter Expert Review)

Adherence	*The degree to which a facilitator(s) follow the program schedule, manual, and/or curriculum guide*
Adherence Situation	This morning your colleague called in sick and you were called in to deliver an 8-hour program that develops independence in 16-18 year olds. You recall practicing this program several months ago during staff training, but don't have any more resources than the program guide, which describes each step of the program in great detail. Please respond with what you believe will be the most effective strategy to achieve the program goals and maintain the program plan?
Dosage	*The degree, percentage, or amount of the core original program that is actually delivered by the facilitator(s)*
Dosage Situation	You and a co-facilitator are delivering a 4-hour program that promotes cooperation skills in youth-at-risk. After driving for 30 minutes to the site to deliver the program your co-facilitator realizes that he forgot the required program supplies at home. Driving back to get the supplies will cut the time you have to less than 3 hours. Please respond with what you believe will be the most effective strategy to achieve the program goals and maintain the program plan?
Quality of Delivery	*The amount of enthusiasm, skill, and/or competency to which the facilitator delivers the program*
Quality of Delivery Situation	You are delivering a 90-minute resilience program to a group of eight 13-15 year olds. Your co-facilitator is not adhering to the program plan and mentioned to you beforehand that he/she knew the teens would not be engaged. Her poor attitude towards the program is now causing the teens to disengage. Please respond with the strategy you believe would be the most effective to reengage your partner while maintaining program quality?
Participant Responsiveness	*The degree to which program participants are engaged and/or involved in the program tasks and/or responsibilities due to the facilitator*
Participant Responsiveness Situation	You are delivering a program aimed at developing self-efficacy to a group of 8-9 year olds that has only previously been delivered to 14-16 year olds. The 8-9 year olds are having trouble focusing on the activities and are not engaging in the program. Please respond with what you believe will be the most effective strategy to achieve the program goals and maintain the program plan?

Table 2. (Continued)

Operational Definition & Situation (Pre-Subject Matter Expert Review)

Program Differentiation	*Program components and outcomes that can be: (1) directly attributed to the current program and (2) as independent from prior/current participant exposures to unrelated programs; this attribution and independence can be influenced by the program facilitator and/or external factors.*
Program Differentiation Situation	You are delivering a five-hour program to develop self-regulative behaviors in a group of nine 12-14 year olds. About an hour into the program, several of the participants mention that they "played the same games" last week in their afterschool program you are not affiliated with. You know that it is important to stick with the program plan as evidenced in your own training. Please respond with what you believe will be the most effective strategy to achieve the program goals and maintain the program plan?
Adaptation	*The frequency, degree, and style of change(s) made outside of the original program design during implementation*
Adaptation Situation	You are providing the second half of a 2 day 16-hour program aimed at improving leadership quality in nine 11-13 year olds. As part of the program the youth developed a video to show what leadership meant to them. However, as soon as you moved to the video sharing component of the program, the electricity went out. This video was the central outcome of the program. Please respond with what you believe will be the most effective strategy to achieve the program goals and maintain the program plan?

Referenced from Berkel et al., 2011; Dane & Schneider, 1998; Durlak & DuPre, 2008; Dusenbury et al., 2003; 2005; Hansen, 2014

SMEs for this study were recruited from a "summer camp professionals" social media group (i.e., Facebook group). This method reflected a combined sampling approach (i.e., convenience, purposeful, and theoretical). More specifically, data were collected from a group that is readily accessible (i.e., convenience), this group is also reflective of the population of interest (i.e., purposeful), and the sample also reflects descriptive needs through the screening criteria (i.e., theoretical) detailed earlier (Bryant & Charmaz, 2007, p. 235). Due to the nature of the data collection a response rate was unavailable. Specifically, there was not a mechanism to determine the number of

potential respondents to the call for participation within the study to the actual number of respondents, due to the social media group interface.

Procedure

Prior to presentation of scenarios, SMES were asked to provide basic descriptive information regarding their gender, age, ethnic group, current job setting, years of professional experience facilitating groups, education level, and number of job related certifications. *Facilitation* was defined as the planning, guiding, and management of a group to achieve a goal, this could include, but was not limited to: training staff, providing experiential programs, teaching fitness classes, leading group trips, etc. SMEs who indicated they did not meet screening requirements through their responses to these questions were skipped to the end of the survey and dropped from later data collection/analyses. SMEs were then prompted to provide how they would respond to each of the six scenarios presented (see Table 2), in two short-answer formats: (1) *Please respond with what you believe will be the most effective strategy to achieve the program goals and maintain the program plan* and (2) *Please enter the strategy that you believe would be the most detrimental to achieving the program goals*. SMEs were given the option to describe what additional information or clarification they would like to see presented in the scenario. Through this approach SMEs will be less likely to provide "undesirable" responses due to the flexible nature of the questions; thus, responses should be considered valid as they are less prone to potential respondent social desirability and they allow for SMEs to provide feedback in areas the questionnaire that were unclear (Weekley & Ployhart, 2005; Whetzel & McDaniel, 2009).

Analyses

The generation of situation revisions and responses consisted of two phases: (1) content analysis of SME responses by three raters and (2) a member check with SMEs to establish trustworthiness of the data. To ensure that the codes produced by raters were valid in the context of this study and in potential future iterations utilizing the SJTs, the analysis was considered to be at one of three potential levels: (1) the *manifest* level (i.e., count level data, such as how many times a word appears in a response regardless of placement), (2) *latent* level (i.e., recognizing patterns within the data, such as phrases with similar meaning), and (3) *projective* level (i.e., projection/representation of coder's own paradigms and schema to the data, such as their own past experience with the phenomena of interest or view of respondent's needs or deficits) (Potter & Levine-Donnerstein, 1999). As the goal of this study was to produce alternative measures of situations germane to youth-centered LER settings that reflect global challenges a facilitator may face in their implementation of programs, the latent level of analysis was the most appropriate strategy, producing codes (i.e., situational responses) that represent what they are intended to measure (Curry & Nunnez-Smith, 2015) as directly produced by SMEs in their responses to the six presented situations; no hidden or inferred level meaning was projected by raters into response coding. Additionally, to ensure codes were reliable (i.e., trustworthy), an audit trail was utilized reflecting all stages of the content analysis process including raw data preparation, independent code creation, member checks, code refinement and ranking, inter-rater agreement, and development of the final SJT products (Thomas, 2006). This reliability process was utilized to produce

98

intersubjective convergence, which "gives readers the sense that the patterns in the latent content must be fairly robust and that if the readers themselves were to code the same content, they too would make the same judgments" (Potter & Levine-Donnerstein, 1999, p. 266).

Data were shared with raters after being exported from Qualtrics into a spreadsheet for *content analysis,* defined as "the subjective interpretation of the content of text data through the systematic classification process of coding and identifying themes or patterns" (Hsieh & Shannon, 2005, p. 1278). Each independent rater had broad experience designing, facilitating, and evaluating youth programs in addition to specific training in the development and purpose of SJTs. Raters utilized a *constant comparative approach* where they reviewed "data line by line and coding sections as concepts become apparent, then comparing the text segments with segments previously assigned the same code to decide whether they represent the same concept" (Curry & Nunez-Smith, 2015, p. 379). Codes in this phase of analysis represented potential responses to each of the six situations. After independent response codes were identified, the three raters met to determine which shared response codes were generated from the independent coding and to discuss areas of disagreement between raters. This style of analysis reflects a *general inductive approach*, where the goal is to examine the "raw" data to develop themes and codes through the rater(s) analyses of the data. (Thomas, 2006).

After potential scenario responses (i.e., codebooks) were generated that reflected the general interpretations of the three raters, the codes were further refined to convey the essence of each theme (Thomas, 2003, p. 238). More specifically, through constant

99

comparison, codes with a high degree of overlap were refined to reflect only one response choice. For example, in response to the quality of delivery scenario (See Table 2) where SMEs shared a strategy that would be the most detrimental to program goals, SME A shared "*continuing on the course, and allowing the other facilitator to continue to lead without a positive engagement in the program*" and SME B shared "*ignoring the co-facilitator's attitude and continuing on as planned despite the group being disengaged.*" The common response code that emerged from the rater's refinement was "*move forward without addressing co-facilitator.*" This approach also allowed for SME responses to be broken into multiple situational responses where necessary. For instance, in responding to the adherence situation with a strategy to achieve program goals and maintain the program plan, SME F shared:

> *Honestly assess your ability and attitude in delivering the program. If necessary, cancel, reschedule, or offer a program which you are more comfortable leading and which meets the same or mutually acceptable outcomes. A poorly presented program can by much more detrimental than the inconvenience of making an adjustment.*

Here raters indicated several potential situational choices were embedded within SME F's response, including: (1) canceling the program, (2) rescheduling the program, (3) delivering a more familiar (to the facilitator) program with identical outcomes or, (4) delivering a program where the participants received alternative/differing outcomes. Raters also identified that three of the coded responses (1, 2, 4) did not fit with the situational prompt: *respond with what you will believe will be the most effective strategy*

to achieve the program goals and maintain the program plan, rather these coded

responses seemed to better fit with the second prompt, *strategies that would be the most*

detrimental to achieving the program goals.

After the establishment of response codes in the form of a common codebook,

raters (including members of the author team) independently ranked codes generated

from each situation reflecting strategies most likely to achieve program goals and

maintain program plans to those that were least likely. Raters then shared their rankings,

and in cases where there was disagreement on the order of pro- to anti-program goal

achievement, they reached a consensus. To further ensure the trustworthiness of the data

(e.g., validity and reliability) a member check was conducted with the SMEs to determine

if the near-final SJTs reflected the intent of their responses (i.e., affirming their

perspective was represented within the SJTs). SMEs were solicited, through their earlier

provided email address, to determine if they had additional revisions/or additions to their

initial responses. However, perhaps due to the timing of the follow-up questionnaire

(mid-summer 2016), none of the initial respondents shared additional concerns or

feedback on the near-final SJTs. Of the 17 initial SME respondents, eight opened the

follow-up questionnaire as indicated by the Qualtrics software, but none added additional

feedback, potentially indicating no-further concerns or feedback. SME responses were

also examined for common dimensions across responses to better understand how

experienced facilitators may address the challenges presented within scenarios in the

form of a brief thematic content analysis.

Results

The 17 respondents did not all provide complete responses to demographic/descriptive questions; however, those who did indicated the sample in this study was primarily male ($n = 7$; female = 2), white ($n = 8$, 88.9%), highly educated (88.9% of sample reporting a bachelor's degree or greater), experienced in their role, with an average of 13.11 years of experience facilitating groups ($SD = 7.25$), well-trained as average respondents possessed 6.14 job-related certifications ($SD = 3.29$), and otherwise suitable for inclusion in the study. Respondents also provided a job title which included manager, coordinator, or director; this response indicating a supervisory level position.

A total of 68 potential SJT responses to the six scenarios were developed based on the content analysis. Codes were developed from the raw data by the raters reading SME response data line-by-line and producing codes (i.e., potential scenario responses) as concepts became apparent (Curry & Nunnez-Smith, 2015). Provided below are examples of scenario response choices reflecting the unique dimensions of each scenario, more specifically those that reflect the choice most likely "to achieve the program goals and maintain the program plan" and conversely the choice most likely to "be detrimental to achieving the program goals." In Appendix A, the final scenarios and response options are provided. Additionally, in many cases SMEs shared similar answers across situations, potentially indicating that despite the differing factors that can influence both program outcomes and implementation quality, there may be universal responses to the challenges that facilitators encounter when balancing the need to maintain implementation quality with factors that are often outside of their control. Beyond the responses produced, two

broader themes also emerged from the content analysis, reflecting how SME respondents may navigate challenges germane to the facilitation process.

Adherence.

In the analysis of the adherence scenario, SMEs shared an orientation towards thorough review and preparation in their understanding of program materials prior to program delivery to ensure the achievement of implementation quality and program goals. For example, SME C responded, "*Use any time before the program to review the program guide*." SME A also shared this view "*I would immediately sit down with the program guide and outline what my day was going to look like. I would create a game plan, and gather my materials*." These responses were then narrowed into the SJT response "revisit program guide and plan" reflecting this facilitator orientation. An attitude of creating one's own program plan also emerged in response to this scenario, as SME A shared "*if necessary start them on an activity that will allow me to plan and adjust the program with what I know, and what I am able to accomplish*." This response was then condensed into the SJT response "Deliver alternative program." Respondents also shared responses that would detract from the program goals and implementation quality as highlighted by SME B "*Not paying attention to the program guide and creating my own program on the fly,*" which was modified into the SJT response "not utilize program guide, create own." SMEs provided further responses that could negatively influence the program including SME G:

> *The most detrimental strategy would be to self-doubt or lower my expectations for the group because I'm not familiar with the material. The worst thing I could do*

would be to give up on the group and blame the fact that I haven't run this

program recently

which was condensed into the SJT responses "presenting program poorly, using

circumstances as an excuse" and "lowering expectations of group."

Dosage.

Responses from the dosage scenario reflected similar breadth and quality on how

to achieve program goals and to maintain implementation quality. SME respondents

indicated a focus towards program adherence. For example, SME B shared "*If this were a*

ground activity/low rope only program I would choose to ask the co-facilitator to go back

for the equipment while the group and I start the day with no-prop icebreakers and field

initiatives" which was similar in nature to SME C's response "*If we decided that the*

supplies were integral for some components of the program, I would suggest that one of

us starts with an hour of low prop/no prop, while the other drives back for supplies."

Both responses reflected an acknowledgement that the program supplies were necessary

for high quality implementation leading to the SJT response of "proceed with non-

equipment-required activities, during which have co-facilitator pick-up equipment" and

"deliver program without supplies." SMEs also provided behaviors that could negatively

influence program outcomes in this scenario such as, "*Both facilitators turning around to*

get the supplies. Worse than losing an hour of the program is the message conveyed to

participants regarding your professionalism and lack of preparation" shared SME D.

This response was similar to SME H, who shared "*I think having both facilitators drive*

back to get supplies would be the worst idea." These strategies led to the SJT response "Leave participants at program site to pick up equipment."

Quality of Delivery.

In the quality of delivery scenario SMEs shared strategies that could both maintain the program plan and achieve program outcomes. The first strategy produced was "take a group break and visit with co-facilitator in private," which was generated from the responses of SME B, "*Call for a quick water/snack break for the group, pull facilitator aside and regroup,*" SME D, "*take a break and reorganize with input from your co-facilitator,*" and SME G, "*try to step up with a game or initiative that involves the participants working together so that I could catch a quick word with my co-facilitator.*" SMEs also shared strategies that could harm the program including SME B, "*Ignoring the co-facilitator's attitude and continuing on as planned*" and SME D, "*Continuing to move ahead without acknowledging the disengagement*" which led to the SJT response of "Move forward without addressing program facilitator."

Participant Responsiveness.

In the participant responsiveness scenario SMEs also shared a similar strategy that could harm program outcomes and implementation quality as shared by SME F, "*Continuing the program for 8-9 years old with the design and context set for 14-16 years old*" and SME B, "*Continuing with the same program as planned*" which generated the SJT response of "Continue without acknowledging issues." To address the poor program fit in the participant responsiveness scenario, SMEs provided a strategy that reflected adaptation with the goal of also maintaining the program plan as shared by SME

J, "*Adjust the activities to decrease abstract concepts and make debrief a more direct and concrete*" and SME G, "*tailored some of the program elements to the different group,*" which in both cases led to the SJT response of "Modify/Adjust program plan for age-appropriate behaviors."

Program Differentiation.

In the program differentiation scenario, SMEs were asked to provide strategies that would help/harm program outcomes and implementation when participants indicated they had already experienced similar programming. SMEs shared that rather than changing the plan and reflecting a greater breadth of activities they would rather reflect greater depth as shared by SME C, "*Engage these students in special roles during these activities. Chat with the group about how others with experience in an area can support the learning of a group and group goals*" and SME D, "*brainstorm with the participants how the games might be modified, have more in-depth debriefing, let some participants act as facilitators*" which led to the SJT response of "Utilize prior participant experience as part of facilitation, frame as positive." Conversely, SMEs shared that forcing participants to continue at the same level could harm program goals and implementation. For example, SME A stated, "*Telling the kids that's too bad because this is what we are doing*" which led to the SJT response of "Communicate to participants they "have" to do it anyway."

Adaptation.

The SME responses to the adaptation scenario also reflected a depth versus breadth orientation. Specifically, SME C shared "*support participants to share their*

video content in a live speech - reintegrating that they are using their flexibility as a leader, by adapting to what the situation dictates" which led to the SJT response of "Apply situation (flexibility) to leadership and discuss with kids." SMEs also indicated ending or canceling the program earlier would be a strategy that would be detrimental to the program goals and implementation. Specifically, SME D detailed "*Spending too much time trying to fix the electrical problem, leaving participants disengaged. Ending the program early or on a down note*" which led to the SJT responses of "end program early" and "have participants not share information that is on video."

Response Instructions.

While SME's did not provide feedback regarding potential modification of scenarios, they did share that the response instructions were unclear; specifically, that the scenarios were meant to provide program facilitators with scenarios that may not reflect their specific camp or program setting (e.g., the later respondents camp or program setting may not utilize video for their programs, thus limiting the relevance of the scenario). As such the response instructions were modified from: "*In this portion you will be presented with scenarios commonly encountered during the facilitation of programs. Read the scenario and then circle the action you would most likely take*" to also include an additional statement "*In this portion you will be presented with scenarios commonly encountered during the facilitation of programs. **The scenario details may not be an exact fit to your own programs, but we'd still like to know how you would respond in the given scenario.** Read the scenario and then circle the action you would most likely take*."

Common Dimensions.

Beyond the responses to each of the six scenarios, the raters also noted two distinct themes across SME responses. In the first theme, an orientation towards program adaptation and/or deviation emerged. In response to the best strategy to achieve program goals and maintain the program plan, SME D shared in the dosage scenario "*A good presenter should always be able to improvise seamlessly without negatively affecting program outcome… Otherwise, embrace the ability to "think on your feet" and restructure what you can.*" Similarly, in the participant responsiveness scenario, SME H shared "*I would have to adjust on the fly by adding elements to the games making them easier to complete. Or, I could adjust the games/activities to ones I thought would better suit this age group.*" It is unclear if these responses represent an adaptation, deviation, or some combination of both; where an adaptation reflects a modification to a program plan while maintaining a focus towards a program's goals, and conversely a deviation reflects a deliberate modification/omission of a program plan or component due to a facilitators lack of motivation or buy-in to the program material or participants. SME B shared a response reflecting adaptation to the participant responsiveness scenario, "*change the planned activities to either be less intense or change to new activities which will develop the same goals, while re-engaging the 8-9 year olds*" specifically, that despite the introduction of new activities, they maintained an orientation towards the same goals. Conversely, responding to the quality of delivery scenario, SME C shared,

> *…pull facilitator aside and regroup. Ask what activity they want to lead that*
> *would get the group back on track. When group returns I would have a quick*

reengaging talk and lead the next activity to hopefully boost engagement, then

support the other facilitator to lead their ideal activity with the group. At that

point I would reassess both the group engagement and that of the facilitator and

make my next choice.

In this response, SME C provided an example of a deviation, specifically having the non-engaged facilitator lead an "ideal activity" in the hopes of raising the facilitator's engagement level. While this modification was implemented indirectly to achieve program goals, it represents a deviation from the program plan, specifically the modification or omission of a program component due to the facilitator's motivational issue.

The second theme that emerged throughout the six scenarios was the SMEs' reliance on their own experience and training to achieve program goals and to maintain the program plan. For example, in response to the participant responsiveness scenario, SME E reflected on their own training as a facilitator, "*Wrong activities for the age group. ages and stages are critical to program success. sounds like the goals and plan are off*" and further elaborated how an experienced facilitator would be able to navigate the challenge "…*a good facilitator with great skills and experience working with groups of all ages would be the best answer in that situation.*" Similarly, in response to the dosage scenario, SME J shared:

It depends on what program elements were agreed to. Essential safety equipment

is worth the loss of an hour. Activity props are not. An experienced facilitator

relies on their knowledge of the process and group dynamics more than specific

activities

Here in both examples the SMEs indicated that experience is in fact a solution to many challenges facing program facilitators rather than activity/program design. More simply, when "something" goes wrong a more experienced facilitator may be more adept at managing the issue.

Discussion

The purpose of this study was to produce an alternative method to predict implementation quality; specifically, SJTs that reflect the six dimensions that may influence a facilitator's implementation quality and corresponding programmatic outcomes. Specifically, the primary goal of the study was to produce valid measures reflecting facilitator behaviors and/or reactions in each provided scenario based on the training and broad experience of 17 subject matter experts. The data provided by these SMEs and later content analyses by three trained raters familiar with both implementation science and the facilitation experience resulted in 68 responses across the six scenarios, reflecting behaviors that would promote and/or inhibit implementation quality and corresponding programmatic outcomes.

The results in all six scenarios were congruent with prior research on how facilitators may manage challenges they encounter that are normative to the facilitation experience. For example, in a study investigating the effect of quality of delivery on program implementation Little et al. (2013) found that for teachers to maintain program support, and thus quality of delivery, high quality initial and ongoing train is necessary.

This finding is similar to a SJT response to the quality of delivery scenario "revisit & refocus on program goals with co-facilitator." Additional congruence with past research was highlighted in SME responses to the participant responsiveness scenario. In a review of evidence-based prevention programs, Moore et al. (2013) shared,

> when adaptations to the target population were made, the program was delivered to a sample for whom the program was not initially designed and in which it was evaluated (e.g., with students who are younger or older). Collectively, these results indicate that changes to procedures, dosage, and content were the most commonly reported adaptations. (p. 154)

This finding is congruent with several of the SJT responses that emerged from this study including "modify activities to better fit group" and "modify/adjust program plan for age-appropriate behaviors."

Beyond the parallels with past research highlighted within the SJT response choices, similarities also emerged with prior research across the six scenarios on how SMEs may behave/react when faced with challenges to implementation and/or program quality. More specifically, a positive orientation towards pro-adaptive behaviors and a heavy reliance on their own experience and training as facilitators emerged. These responses are congruent with that of Gagnon and Bumpus (2016) who noted there was not a significant relationship between facilitator reported fidelity and facilitator reported experience level.; nor did Gagnon and Bumpus find a meaningful relationship between pro-fidelity beliefs (i.e., believing that it is important to deliver a program as designed) and experience. This congruence with past research may indicate that when facilitators

possess high levels of experience the maintenance of implementation quality is less important to achieving outcomes than their own ability and skill level. Indeed, SME D shared, "*remember that the outcome is not in completing the prescribed program, but in the change and growth you want to see in the participants.*"

Limitations

While the findings of this study are promising, there were limitations. The lack of ability to capture a response rate due to the nature of the social media group where SMEs were solicited for participation is cause for concern, specifically if there were meaningful differences between the respondents who chose to participate and those who did not. The timing of the member-check follow-up questionnaire was problematic as no respondents provided a follow-up response. While respondents indicated in the first questionnaire their ability to complete the follow-up, no feedback on the finalized SJTs was provided from SMEs. This could suggest that the respondents had no need for alteration with the proposed measures. However, an alternative explanation may be that as the member-check of SJTs took place in mid-summer, respondents employed within the OST industry simply were too busy to complete the follow-up measure. Future iterations of this study (e.g., refinement and/or expansion of SJTs) should take place in OST off- seasons.

An additional limitation related to how SMEs responded to questions. Specifically, in some instances SMEs would provide both positive (e.g., achieving program goals and maintaining program plan) and negative (e.g., detrimental strategies to achievement of program outcomes) responses within the response box termed "effective strategies to achieve program goals and to maintain the program plan." While this

112

differentiation was apparent in many instances to raters, the lack of response to the member check on the part of SME respondents did not allow for this differentiation to be further clarified.

Future Directions

One of the primary challenges relating to the usage of SJTs is that these measures may have low internal consistency leading to questions about their reliability (Chan & Schmitt, 2002). This is likely due to the high level of specificity often driving the design, implementation, and purpose of many SJTs; more specifically, SJTs are generally developed for use within one setting (e.g., a specific organization) and thus may reflect a culture that is specific to that setting. The advantage of the approach undertaken in this study is that the SJTs were developed to reflect a more global OST setting and situations that commonly occur across OST programs rather than a specific context. As such, future investigations utilizing the SJTs produced from this study should examine their reliability across OST settings and sites to determine if the responses provided are consistent with this purpose.

The finding that for many SMEs prior experience and training and thus program modification and/or adaptation seemed to exceed the importance of fidelity to a program plan, highlights a challenge embedded within our understanding of implementation research and related facilitator training. While a wide breadth of literature suggests facilitators with high levels of experience may have lower levels of implementation quality and poorer corresponding program outcomes, there may be a population embedded within highly experienced facilitators who are unaware of this shortcoming.

Developing trainings for more experienced facilitators that highlight this potentially negative relationship may help mitigate this "I know better and/or be flexible" culture. Conversely, in future research, the factors that cause experienced facilitators to maintain program quality should also be examined, to determine how to enhance both buy-in and maintenance of high implementation quality. Furthermore, facilitator trainings that highlight the relationship between high quality implementation and program outcomes in addition to the other reasons for maintaining implementation (e.g., validity of programmatic outcomes and plan) should be explored to determine how to better enhance high-experience facilitators' implementation quality and motivation.

While a strong body of evidence suggests that SJTs are both reliable and valid predictors of job performance, to the author's knowledge SJTs have not been utilized in OST settings to predict program outcomes or implementation quality. The SJTs produced from this study should be tested within an OST setting to determine their efficacy in predicting program outcomes and implementation quality in comparison and in combination with other measures of facilitator traits and characteristics. The findings of these future studies could highlight the best possible combination of strategies to predict later staff performance and thus provide higher quality OST programs.

CHAPTER FOUR: A MULTI-LEVEL MODEL EXPLORING THE RELATIONSHIP BETWEEN PERCEIVED IMPLEMENTATION BEHAVIORS, ATTITUDES, AND OUTCOMES

Within the context of leisure, experiential education, and recreation (hereafter referred to as LER) there is a growing body of evidence suggesting that well-designed and implemented programs and events lead to positive economic (Oh et al., 2016), social (Arai & Pedlar, 2003), and emotional development for the communities and constituents they are intended to serve (Kleiber, Walker, & Mannell, 2011). Correspondingly, there has been a shift in leisure assessment and research towards utilization of the best available evidence to justify program and event selection, funding, and sustainability (Berk & McGivern-Moon, 2016; Browne et al., 2011; Bruening et al., 2015; Crompton, 2016). Further, there is a shift in LER programming due to increasing challenges as funders pressure organizations to demonstrate how their programs achieve stated effects (Mainieri & Anderson, 2015a). While this orientation toward evidence-based decision making is reflected in research relating to the outcomes of these programs and events, a notable gap exists in terms of our understanding of their implementation quality (Gagnon & Bumpus, 2016; Mainieri & Anderson, 2015a). Indeed, several comprehensive reviews of the prevalence and influence of implementation quality on programmatic success highlight that the core concept of implementation quality is frequently ignored or undervalued when presenting the results of studies promoting the efficacy (or lack thereof) of programs (Berkel et al., 2011; Borrelli et al., 2005; Dane & Schnieder, 1998; Durlak & DuPre, 2008; Moncher & Prinz, 1991). This lack of evaluation-focused rigor

within many studies risks the likelihood of misstating that a program is effective when it is not properly implemented (e.g., Type 1 Error) and/or inferring that a program was ineffective when it was not properly implemented (e.g., Type 2 Error).

While implementation research is somewhat limited as compared to outcomes-focused evaluations, it is widely acknowledged that the assessment of program implementation is essential for evaluating the internal and external validity of programs (Berkel et al., 2011; Durlak & DuPre, 2008). Implementation research and assessment can highlight reasons for programmatic success or failure, how a program design may be improved, and how a program may be replicated in settings outside of the program designer's scope (Durlak & DuPre, 2008). Despite ample evidence that implementation assessment is a critical component of the program delivery and improvement process (Fagan et al., 2008; Frantz, Stemmler, Hahlweg, Pluck, & Heinrichs, 2015; Little et al., 2013), implementation assessments still rarely take place (Sloboda et al., 2014) and are often not included as part of the overall program evaluation process and/or report (McGrew, Bond, Dietzen, & Slayers, 1994; Pawson & Tilley, 1997). Reasons for the lack of implementation investigations may include a lack of awareness about implementation assessment and its importance (Dusenbury et al., 2003; Dane & Schneider, 1998), a lack of need for external program replication (Sloboda et al., 2014), and a lack of resources to support and/or conduct implementation assessments (Lillehoj et al., 2004). To better understand the role of implementation in the context of LER, this manuscript: (1) examines the conceptual foundations of implementation assessment, (2) describes how implementation is evaluated at the facilitator level, (3) introduces a new method for the

116

assessment of implementation, and (4) shares results of an experiment exploring the usage of this alternative method in comparison and/or in combination with more standard measures of implementation quality.

Evaluating Implementation Quality

Program implementation quality is typically influenced at four distinct, but related levels: (1) the characteristics of the organization providing and supporting the program, (2) the characteristics of the community and participants the program is intended to serve (Durlak & DuPre, 2008), (3) the characteristics of the program itself (Little et al., 2013), and (4) the characteristics of the facilitator(s) providing the program (Berkel et al., 2011; Wandersman et al., 2008). This study focuses on the fourth level, that of the program facilitator. While implementation quality may be influenced at multiple levels, it is ultimately the responsibility of the program facilitator (i.e., front-line staff member responsible for delivering the program) to deliver a program as intended by the program designers (Wanless et al., 2015b). In this facilitator level context, six primary components (Hansen, 2014) should be considered as contributing to implementation quality: adherence, dosage, quality of delivery, engagement, program differentiation, and adaptation (See Table 3, Berkel et al., 2011; Dane & Schneider, 1998; Durlak & DuPre, 2008; Dusenbury et al., 2003; Hansen, 2014). However, despite evidence that these six components (at high levels) may have both combined and independent positive effects on implementation quality, they are often not examined together to determine or predict total implementation quality (Durlak & DuPre, 2008). For example, in a review of 39 studies specifically investigating the role and influence of implementation quality, Hansen

117

Table 3.

Implementation Evaluation Component Definition List.

Component	Definition
Adherence	The degree to which a facilitator(s) follow the program schedule, manual, and/or curriculum guide.
Dosage	The degree, percentage, or amount of the original program that is actually delivered by the facilitator(s)
Quality of Delivery	The amount of enthusiasm, skill, and/or competency to which the facilitator delivers the program
Participant Responsiveness (i.e., Engagement)	The degree to which program participants are engaged and/or involved in the program tasks and/or responsibilities
Program Differentiation	The degree of program outcomes that can be directly attributed to the current program versus past programs and/or experiences
Adaptation	The degree to which a program's components are modified to suit the needs of participants or due to other factors

Referenced from Berkel et al., 2011; Dane & Schneider, 1998; Durlak & DuPre, 2008; Dusenbury et al., 2003; Hansen, 2014

(2014), highlighted that only one study, conducted by Ennett et al. (2011), evaluated all

six dimensions. In this study of a substance abuse prevention program for 7th-8th graders,

Ennett et al. (2011) emphasized that all six dimensions may need to be investigated to

have a more thorough understanding of a program's implementation quality, but the

authors also demonstrated that some dimensions (e.g., adherence and dosage) may be the

best predictors of implementation quality. Supporting the findings of Durlak and DuPre

(2008), the authors stressed that the six components of implementation quality should not

be considered independently, rather that they should be considered interdependent. For

example, an assessment of a facilitator's adherence to a program design without also

monitoring for corresponding adaptations may not fully capture a program's

implementation quality and the factors (or lack thereof) that contributed to the program

being delivered as designed. This assessment of facilitator implementation quality and the factors that may contribute or detract from it typically involve an indirect, direct, or hybrid approach (Gresham, 1989; 2009; Dusenbury et al., 2003).

The indirect strategy of implementation assessment is generally conducted through a self-report of the facilitator delivering the program (Gresham, 1989; 2009). For example, a facilitator may utilize a checklist/questionnaire indicating to what degree he/she followed the curriculum and fully delivered the program (i.e., adherence and dosage), their level of enthusiasm while delivering the program (i.e., quality of delivery), participant excitement toward the program (i.e., engagement or responsiveness), participant reported involvement with a similar program with comparable goals (i.e., program differentiation), and modification of the program due technical issues with equipment or low participant engagement (i.e. adaptation) (Ennett et al., 2011).

Conversely, the direct assessment strategy generally consists of a trained rater (e.g., person(s) trained in both the program and assessment) observing a facilitator delivering the program and utilizing a checklist format to determine: (1) the degree to which the program was delivered (i.e., adherence and dosage), (2) the provider's observed level of enthusiasm and engagement while delivering the program (i.e., quality of delivery), (3) participant levels of attentiveness and focus to the program and program provider (i.e., participant engagement), and (4) the providers' (i.e., the person(s) actually delivering the program) modification, removal, or addition of program components (i.e., program adaptation and/or deviation) and the reasons for this programmatic change (Lillehoj et al., 2004; Gresham, 2009; Hansen, 2014). As opposed to less intrusive self-

119

report (i.e., indirect) measures which typically require a low level of training on program providers part to complete (Floyd, Phaneuf, & Wilczynski, 2005), within the direct assessment strategy the implementation rater generally needs training in the program similar in quality to that of the facilitator, and enough experience within the program (and as a facilitator more broadly) to detect nuanced behaviors such as facilitator competency (i.e., quality of delivery), adaptation, deviation, and participant responsiveness (Dane & Schneider, 1998; Gresham, 1989). In programs that are implemented per a highly-structured manual or curriculum, a checklist following a schedule is often utilized to account for dosage and adherence (Eames et al., 2007; Gresham, 1989). However, in less manualized programs a checklist may be only partly reliable to ensure that all six components of a program were delivered as designed (Ennett et al., 2011; McGrew et al., 1994). This limitation is due to the lack of ability for a checklist to capture the nuanced behaviors on the part of the facilitator and participants, it may not capture deviations or adaptations to the program and the underlying rationale for modifications, nor may it possess evidence of validity in its measurement properties across raters and/or facilitators (Kogan, Holmboe, Hauer, 2009).

A hybrid assessment combines the indirect assessment of the facilitator self-reported measures of adherence, dosage, quality of delivery, participant engagement, and adaptation with the direct account of a rater (i.e., observer) also reporting on adherence, dosage, quality of delivery, and adaptation (Hansen, 2014). The goal of the hybrid assessment strategy is to tell a more holistic story of the implementation quality of a program, as some evidence suggests that indirect assessments tend to over-report

implementation quality and direct assessments tend to under-report implementation quality (Durlak & DuPre, 2008; Dusenbury et al., 2003; 2005). As such, a hybrid strategy may better capture a balanced perspective from these two sources of bias in describing program implementation quality (Borrelli et al., 2005). For example, in a study of a preventative intervention aimed at reducing risk behaviors in early adolescents (e.g., 10-14 years old), Lillehoj et al. (2004) found that facilitators tended to be more generous in their self-reported implementation quality ratings than observers (i.e., raters); however, implementation quality scores between observers and facilitators in this study were positively correlated.

The indirect, direct, and hybrid methods of implementation assessment may be prone to a few limitations. In the indirect method, the primary challenge to the validity of this approach is the potential for social desirability of the person completing the self-report (Lillehoj et al., 2004). For instance, a facilitator may over-report their own adherence to a program design or over-report participant engagement levels to be viewed as competent in their role as the program facilitator. In a study comparing self- and peer-reports of work behaviors and competencies, Fox, Spector, Goh, and Bruursema (2007) found that there were meaningful differences in scores, with self-reporters rating themselves higher in work behaviors and competencies than peer reports. An additional challenge embedded within indirect assessment relates to the tracking of participant responsiveness, as this implementation quality component is typically assessed retrospectively. Due to the practical challenge tracking participant responsiveness during program delivery, as this tracking may be both disruptive to the program and unfeasible

to accurately measure utilizing an indirect approach (Dufrene, Noell, Gilbertson, & Duhon, 2005). This retrospective approach may lead to the facilitator missing specific instances where the participants were low and/or high in engagement.

In the direct strategy, there is a degree of reactivity (i.e., social desirability) also potentially present (Domitrovich & Greenberg, 2000). For example, the presence of an observer (i.e., program rater) may cause the facilitator to deliver the program differently (i.e., reacting) and correspondingly compromise the validity of the assessment results as recorded by the observer. For example, in a study of a pediatric weight control program, Jelalian et al. (2014) suggested that the presence of observers may have increased the facilitator's level of effort in providing the program. In other words, the observer presence may cause the facilitator to deliver the program more accurately/ with fidelity (i.e., reacting) and thus does not provide an accurate assessment of facilitator behaviors and implementation quality when the observer is not present. An additional challenge embedded within the direct approach to implementation assessment may be the skill of raters to accurately measure a facilitator's ability to deliver and implement a program. For instance, in a study of principals' observations of a teacher's ability to produce outcomes in students (i.e., math and reading achievement), Jacob and Lefgren (2008) found that principals were accurate reporters in cases where the teacher's ability was very high or very low. In circumstances where a teacher's ability was more moderate, principal reporting was deemed less accurate than teacher self-report. Expounding on this limitation, Malloy et al. (2015) noted, "observational data are usually limited to short, intermittent blocks of time and may not capture teachers' true performance when they are

not being observed or videotaped" (p. 1092). An additional limitation of the direct

strategy relates to the levels of additional resources necessary such as labor-costs, staff

attrition, training refreshment (i.e., retraining), training materials, and logistical support

(Dolcini et al., 2014; Dusenbury et al., 2003; Fagan et al., 2008; Hansen, 2014). This

strain on resources is also compounded by the recommendations of Dane and Schneider

(1998) and Spoth et al. (2007), who proposed that in the utilization of observers as a

strategy for the direct assessment of implementation quality, these observers must/should

be independent of the program, that is, not directly involved in its design, delivery, or

evaluation. While such observer independence may be possible when a research team is

first implementing a program, the sustainability of this approach is questionable.

Additionally, in many direct implementation assessments, the observer often completes

the same, if not a greater amount of, training than the facilitator they are observing.

Specifically, an observer must have training in not only the program itself, but also in

how to conduct an implementation assessment of said program. Given the additional

resources necessary for direct implementation assessment, this approach may be

prohibitive for organizations with limited resources and funding (Durlak, 2015; Wanless

et al., 2015b). Finally, hybrid implementation assessment may have similar limitations to

the indirect and direct methods, including those relating to social desirability, reactivity,

and the resource heavy nature required for observers. Also, as highlighted in Ennett et al.

(2011), the hybrid method frequently produces contradictory results. For instance, the

results of the indirect implementation assessment (as part of the hybrid approach) may

indicate low participant responsiveness, but the direct method (also as part of the hybrid

approach) may indicate the opposite. This discrepancy may introduce a degree of

subjectivity when analyzing the results, and correspondingly compromise the usefulness

of the hybrid assessment (Borelli et al., 2005).

The purpose of direct, indirect, and/or hybrid implementation assessment is to

capture the degree of implementation quality present in a program session(s). However,

despite their usage and support, the methods of indirect, direct, and hybrid

implementation assessment are vulnerable to several limitations. Correspondingly,

several emerging alternative strategies have been associated with both the prediction of

implementation quality and corresponding programmatic outcomes, including the usage

of facilitator characteristics and traits to predict implementation quality, the use of

facilitator self-reported "pro-implementation beliefs," and the use of short vignettes to

assess a job candidate's likelihood of delivering a program. As such, this study will

investigate alternative strategies that may predict and/or influence programmatic

implementation quality and later program outcome achievement.

The Relationship of Facilitator Characteristics with Implementation Quality

Beyond the six implementation components listed in Table 3, it is also important

to consider facilitator characteristics and traits as these also may influence

implementation quality (Gagnon & Bumpus, 2016). The facilitator has a great deal of

influence over a program's implementation quality (Wanless et al., 2015), and

correspondingly, the outcomes that may or may not result from the program. Factors such

as the facilitator's level of experience delivering and facilitating programs, their

programmatic buy-in, their belief in the importance of delivering programs as designed

(i.e., pro-implementation attitudes/beliefs), and program related training all have been shown to have a strong influence on implementation quality (Gagnon & Bumpus, 2016; Perepletchikova, Treat, & Kazdin, 2007; Sloboda et al., 2014). Less malleable traits such as a facilitator's gender, race, body type, and age in relation to participants have also been found to influence program implementation quality (Dolcini et al., 2014; Gagnon & Bumpus, 2016; Lillehoj et al., 2004). For instance, in a study of an HIV prevention program, Dolcini et al. (2014) noted that facilitator race influenced implementation quality. More specifically, facilitators of the same race as the participants being provided the program tended to deliver more of the program (i.e., higher implementation level). Somewhat conversely, the authors found that facilitator gender in relation to participant gender led to lower rates of implementation quality, more so for female facilitators. In fact, female facilitators had lower rates of implementation quality regardless of participant gender, but more so if their participants were also female. Consistent with this finding, Lillehoj et al. (2004), in a study of a substance abuse prevention program, found that male facilitators tended to also have better implementation quality and programmatic outcomes regardless of participant gender. However, in both the case of the studies of Dolcini et al. (2014) and Lillehoj et al. (2004) the reasons for this gendered effect were unclear.

Facilitator experience also has been shown to have an influence on implementation quality and corresponding program outcomes. This characteristic is often defined at two levels: (1) a facilitator's broad level of experience delivering programs, both related and unrelated to the program of interest and (2) a facilitator's level of

125

experience delivering the actual program of interest (Dusenbury et al., 2003). Both types of experience of been found to influence implementation quality (Nobel et al., 2006; Zollo & Gottschalg, 2004), but the relationship between experience and implementation quality is unclear. In some studies, a facilitator's higher experience level with a specific program has been shown to have a negative influence on implementation quality (Macmillian, 1998) while in others a facilitator's experience level has been found to have a positive or null effect (Dusenbury et al., 2005; Stein et al., 2008). Lillehoj et al. (2004) found that a facilitator's specific experience delivering similar programs positively predicted outcomes, teacher buy-in (i.e., program support) and implementation quality in a youth-orientated substance abuse program, suggesting that it may be "beneficial when teachers see the value of implementing the prevention-related program and have the necessary background knowledge to do so" (p. 254). However, in an assessment of a similar program conducted by Rohrbach et al. (1993), the authors found that facilitator experience level was a negative predictor of implementation quality. These contradictions within the research exploring the relationship between implementation quality and facilitator experience indicate that more investigation exploring this relationship (or lack thereof) is warranted (Little et al. 2013). An additional challenge involving facilitator experience relates to their inconsistent and potentially more nuanced measurement. Prior research indicates that many measures of work experience are prohibitively unidimensional, failing to recognize the potential variables that also constitute work experience, differences embedded within work experiences, and the differing impact each "type" of experience may have on workplace performance (e.g., time in current position

vs. total time with current organization vs. total time in differing organizations) and these differences may inflate or inhibit the potential role of experience on program outcomes (Weekley & Ployhart, 2005). This finding highlights how experience as a variable can be misconstrued within LER research. For example, if three facilitators providing a rock climbing program to youth-at-risk were asked to provide their experience in years as a climbing instructor as part of a program evaluation and they shared a mean level of 3.1 years (SD = 1.4), and corresponding program evaluation findings indicated instructor experience level did not meaningfully predict outcomes in rock climbing skill or socio-emotional development in youth participants, does this indicate that experience did not matter? What if one of the instructors also had 15 years of therapeutic backpacking skill experience with a similar population and her peers did not? As these examples suggest, the experience variable as a construct is likely more multi-dimensional than some prior research suggests.

In contrast to the ambiguity regarding the relationship and/or influence of facilitator experience and implementation quality, a strong body of evidence suggests high quality intentional training of program facilitators has a positive and sustained effect on implementation quality (Berkel et al., 2011; Frantz et al., 2015). For example, Dolcini et al. (2014) found that when program facilitators received both initial and follow-up training following the program designer guidelines, they had higher levels of implementation quality. The researchers also noted that facilitators receiving poorer quality training tended to have lower levels of implementation quality. Paralleling the facilitator experience construct in terms of breadth versus specificity, training as a

predictor of implementation quality also consists of multiple levels: (1) the broad level of experience/training a facilitator may have in related and/or unrelated programs, (2) the program-specific training a facilitator may receive, and (3) the degree to which implementation quality is emphasized as part of a facilitator's program specific training (Cyr, 2008; Dufrene et al., 2005).

Another characteristic found to influence implementation quality is the buy-in of the program facilitator. More specifically, *facilitator buy-in* is the level of support and enthusiasm a facilitator has to deliver the program, their beliefs in the goals and outcomes of the program, and their level of agreement that the program will deliver the outcomes it designed to achieve (Dusenbury et al., 2003; Johnson, Mellard, Fuchs, & McKnight, 2006). The relationship between facilitator buy-in and implementation quality is somewhat unclear. In some studies, facilitator buy-in has been strongly associated with implementation quality (Dariotis, Bumbarger, Duncan, & Greenberg, 2008; Stein et al., 2008), but in others the direct influence of program buy-in is less meaningful. For example, in a study of teacher implementation quality of a classroom management strategy, Wehby et al. (2012) found that teacher engagement (i.e., buy-in to the program) did not have a direct influence on implementation quality, but it did have a moderational effect on the strength of relationship between program support available to teachers and corresponding implementation quality. These findings suggest that when conducting implementation assessment, it is important to consider the potential complexity of factors that may inhibit or promote implementation quality. The review of implementation research conducted by Dusenbury et al. (2003) also suggests that buy-in may also provide

a unique influence to the quality of delivery and participant responsiveness components described in Table 3.

Alternative Methods of Implementation Assessment

      Implementation Quality Prediction. An alternative method to the indirect, direct, and hybrid strategies of implementation assessment may be the use of variables such as pro-implementation beliefs to predict implementation quality and relationships with program outcomes. *Pro-implementation beliefs* are measured as the degree to which a facilitator places value on delivering a program as designed and with intention to the programmatic outcomes (Gagnon & Bumpus, 2016; Gagnon, Garst, & Stone, 2015a; Rohrbach et al., 1993). This method was utilized in a study of facilitators delivering a leadership program to incoming college students (i.e., transitioning high school seniors to first-year college students). The findings suggested that the pro-implementation belief variable was a promising predictor of implementation quality and corresponding programmatic outcomes (Gagnon & Bumpus, 2016). Pro-implementation beliefs in this study were measured as a sub-dimension of the Facilitator Characteristics and Program Contributions scale (FCPC). The FCPC was designed to be a global measure of implementation quality, specifically to measure pro-implementation beliefs, perceived competence, and perceived preparedness. The FCPC is also designed to be used interchangeably, capturing broader elements of implementation quality rather than program-specific factors. Further replications of the FCPC in an at-risk youth program and a college recreational leadership development program yielded similar results (Gagnon et al., 2015a), indicating that as facilitators scored higher in pro-implementation

beliefs they had higher levels of program outcomes than their less "pro-implementation" focused facilitators. These findings support past research suggesting that facilitator buy-in is a key component in the achievement of quality program outcomes and programmatic implementation (Larsen & Samdal, 2007).

Situational Judgement Tests. Beyond the approach outlined in the FCPC, a promising method for the prediction of implementation quality may be the use of situational judgement tests (SJT). In a situational judgment test, a respondent is presented with probable scenarios they will encounter as part of their work (Weekley & Ployhart, 2005). Applicants are then asked to assess each choice and respond with how likely they may be to choose that response in a given scenario (Whetzel & McDaniel, 2009). The goal underpinning the use of SJTs is to better understand a respondent's judgement when presented with a scenario they will likely encounter in their work environment (Russell, 2010). In alternative contexts, SJTs have been shown to positively predict both immediate and sustained job performance due to their validity and simplicity (Whetzel & McDaniel, 2009). More specifically, SJTs have demonstrated incremental validity above other cognitive approaches (McDaniel et al., 2007), and have less vulnerability to faking responses (i.e., social desirability) than self-reported measures of behavior such as those measuring personality type (Peeters & Lievens, 2005).

While SJTs have not yet been utilized to predict implementation quality, this does not suggest that their usage would be markedly different in implementation, analysis, or validity as compared to the other contexts in which they have been utilized. Paralleling prior investigations utilizing SJTs to predict future performance (e.g., leadership,

profitability, productivity, ethics), in the context of implementation quality the SJTs would provide an examination of how a facilitator responds to a likely scenario they could encounter as part of the program implementation process. The facilitator's response (i.e., judgement), captured as an option to a presented SJT scenario, would then be utilized to predict future performance (e.g., linking pro-implementation attitudes to programmatic outcomes). The SJT approach is founded on significant empirical and theoretical evidence that suggests when programs are delivered with high implementation quality, they tend to have better outcomes than those either lacking an implementation assessment and/or those with poor quality implementation (Berkel et al., 2011; Dane & Schneider, 1998; Durlak & DuPre, 2008; Dusenbury et al., 2003; Sloboda et al., 2014). Correspondingly, if a facilitator is presented with a scenario reflecting a "pro-implementation" response to a scenario and they choose said option, they should achieve better outcome levels than their peers who select options reflecting a lower "pro-implementation" quality option, due to the significant body of evidence suggesting SJTs are "better" predictors of performance.

Purpose and Contribution of Study

The goal of this study was to examine the effectiveness of SJTs, the FCPC, and/or facilitator characteristics as an alternative method to predict implementation quality and corresponding parent perceptions of program outcomes in a common youth LER setting, residential summer camp. This will be accomplished through an investigation of relationships among (a) facilitator SJT responses, (b) facilitator FCPC responses, (c) facilitator characteristics, and (d) parent perceptions of program outcomes by testing

131

eight hypotheses grounded in current implementation research. A summary of these hypotheses is available in Table 4. First, to the author's knowledge this is one of the first attempts to examine and/or predict implementation quality in the context of parent perceptions of residential summer camp outcomes (Mainieri & Anderson, 2015a). Research in differing contexts suggests that as facilitator pro-implementation behaviors positively increase, then corresponding programmatic outcomes should also increase. Thus, the first hypothesis is H_1, program facilitators with higher pro-implementation beliefs and behaviors will achieve better parent perceptions of programmatic outcomes. Second, the relationships among the FCPC, SJTs, and facilitator characteristics intended to capture or predict implementation quality has not been fully explored nor tested in a residential summer camp or learning context (Gagnon & Bumpus, 2016; Gagnon et al., 2015b); although research suggests that facilitators with higher levels of buy-in and perceived competence (i.e. preparedness) tended to have higher quality implementation and/or program outcome levels. Therefore, our second and third hypotheses are: H_2, facilitators with higher levels of buy-in will have higher levels of program outcomes and H_3, facilitators with higher levels of perceived competence will have better parent perceptions of program outcomes.

The third contribution of this study relates to facilitator characteristics and implementation quality. Specifically, the influence of facilitator characteristics on implementation quality and corresponding program success is unclear; some research suggests that female facilitators tend to have poorer implementation quality and program outcomes than their male peers regardless of participant gender (Dolcini et al., 2014),

132

where other research suggests females were more likely to implement programs as designed (Aarons et al., 2010), and finally in some studies, findings indicated gender had no effect on implementation quality (Rohrbach et al., 2010). This potential (or lack thereof) gendered influence on program implementation quality leads to a two-part hypothesis: male facilitators will, H_{4a}, produce better parent perceptions of program outcomes than female facilitators and H_{4b}, gender will not meaningfully influence perceived implementation quality regardless of participant gender. Additionally, despite some emerging evidence suggesting that gender may play a role in the achievement of program implementation and program outcomes, there is evidence that gender does not play a meaningful role in the SJT responses (Whetzel & McDaniel, 2009), leading to the fifth hypotheses: H_5, facilitator gender will not have a meaningful influence on SJT response choice.

The fourth contribution of this study explores the relationship of facilitator experience, components of implementation, and program outcomes by exploring two hypotheses. First, some implementation research suggests (e.g., Desimone & Lee-Hill, 2017; Domitrovich et al., 2015; Pas et al., 2015) facilitator experience has no effect on components of implementation quality or program outcomes as a dichotomous variable (e.g., novice or non-novice) and as such the sixth multi-part hypotheses is: facilitator experience as a dichotomous variable will have no effect on H_{6a} facilitator buy-in, H_{6b} perceived preparedness, H_{6c} pro-fidelity beliefs, H_{6d} SJT response choices, and H_{6e} parent perceptions of program outcomes. Second, some evidence (Lillehoj et al., 2004) suggests that the reason for the null effect of facilitator experience is due to the dichotomous

measurement of experience, indicating that a continuous level of measurement will have a positive effect on the same domains (Dusenbury et al., 2005). As such the seventh multi-part hypothesis guiding this study is: facilitator experience as a continuous variable will have a positive effect on: H_{7a} facilitator buy-in, H_{7b} perceived preparedness, H_{7c} pro-fidelity beliefs, H_{7d} SJT response choices, and H_{7e} parent perceptions of program outcomes.

The fifth contribution of this study explores the predictive qualities of SJTs and the FCPC to perceived implementation quality and program outcomes. Research into the usage of SJTs, within organizational contexts suggests that SJTs perform better than traditional survey measures in prediction of later staff performance (Whetzel & McDaniel, 2009), leading to the eighth hypothesis: H_8, SJT response choices will better predict parent perceptions of program outcomes than FCPC measures. This hypothesis is based in part on the work of Weekley and Ployhart (2005), who proposed that SJTs capture "more" in terms of incremental validity than personality and unidimensional measures (e.g., the FCPC), indicating that the combination of characteristics (e.g., facilitator experience and FCPC responses) may be less predictive of implementation quality as compared to SJTs. Following is a description of the setting in which this study took place, the methods utilized to examine the eight hypotheses, the results of hypothesis testing, and implications and future directions uncovered by this study.

Setting

The setting for this study consisted of two university affiliated organizations responsible for six summer OST programs. Sessions at the six OST program sites were

134

co-educational, multi-night, residential experiences, lasting between 5-7 days, and targeted outcomes germane to the summer camp experience (see Table 5). The first organization hosted an eight-week (eight, one week independent sessions) summer experience where each week was broken into independent "pre-college" style programs where youth participants, aged 11-17 years, attend college-style courses taught by university professors, to facilitate successful transition from their primary to secondary education. When not attending these pre-college courses, youth participants stayed in on-campus residence halls and participated in programs provided through on-campus facilitators. Beyond the technical skills learned in classroom sessions, this summer learning focused camp also targets the development of socio-emotional growth in three areas for program participants: responsibility, self-regulation, and exploration (operational definitions available in Table 5). The second university affiliated organization operates five residential summer camps also taking place over eight weeks (eight, one week independent sessions). The five sites support a common mission of enhancing three socio-emotional skills in camp participants: self-regulation, attitude, and responsibility (operational definitions available in Table 5). To achieve the development of these skills, sites provide a diverse array of activities and programs including shooting sports, marine sciences, and wilderness exploration, in which activities are intentionally designed and implemented to encourage skill development.

Samples

Upon institutional review board (IRB) approval of the study, data were collected from two cohorts: (1) program facilitators at one of the six residential summer camps

135

Table 4.

Summary of Hypotheses.

Hypothesis Number	Hypothesis
H₁	Program facilitators with higher pro-implementation beliefs and behaviors will achieve better parent perceptions of programmatic outcomes.
H₂	Facilitators with higher levels of buy-in will have higher levels of parent perceptions of program outcomes.
H₃	Facilitators with higher levels of perceived competence will have better parent perceptions of program outcomes.
H₄ₐ	Male facilitators will produce better parent perceptions of program outcomes than female facilitators.
H₄ᵦ	Facilitator gender will not meaningfully perceived implementation quality regardless of participant gender.
H₅	Facilitator gender will not have a meaningful influence on SJT response choice.
H₆ₐ	Facilitator experience as a dichotomous variable will have no effect on facilitator buy-in.
H₆ᵦ	Facilitator experience as a dichotomous variable will have no effect on perceived preparedness.
H₆ᵨ	Facilitator experience as a dichotomous variable will have no effect on pro-fidelity beliefs.
H₆ₔ	Facilitator experience as a dichotomous variable will have no effect on SJT response choices.
H₆ₑ	Facilitator experience as a dichotomous variable will have no effect parent perceptions of program outcomes.
H₇ₐ	Facilitator experience as a continuous variable will have a positive effect on facilitator buy-in.
H₇ᵦ	Facilitator experience as a continuous variable will have a positive effect on perceived preparedness.
H₇ᵨ	Facilitator experience as a continuous variable will have a positive effect on pro-fidelity beliefs.
H₇ₔ	Facilitator experience as a continuous variable will have a positive effect on SJT response choices.
H₇ₑ	Facilitator experience as a continuous variable will have a positive effect on parent perceptions of program outcomes.
H₈	SJT response choices will better predict parent perceptions of program outcomes than FCPC measures

(hereafter referred to as camps) and (2) from the parents of youth participants (ages 8-17 years) attending one of the six residential camps.

Facilitators. Facilitators in this study were temporary full-time (summer season) employees (i.e., approximately 40-60 hours worked per week, less than 120-day employment duration), typically current undergraduate students aged 18-23, who were hired two months prior to the eight-week session, beginning approximately June 1. Facilitators were employed by a university in the Southeastern United States through one of two on-campus organizations. Prior to beginning the summer season, facilitators participated in a 3-day "general" training that addresses basic youth and site management techniques, small group facilitation skills, and training in safety policies. Specifically, as part of their training facilitators were instructed in their organization's intended outcomes (e.g., a youth's socio-emotional development) and how to develop and achieve said outcomes through role playing, scenarios, and intentional reflection.

Parents. The second sample in this study consisted of the parents of youth attending one of the six camp sites. Based on a pilot study conducted with both organizations in the summer of 2015, these parents are typically white, highly educated, and have middle-to-high level incomes (Garst & Gagnon, 2016a). In this study parents reported on the observed growth of their child one week after the completion of their child's camp experience.

Sample Recruitment and Data Collection Procedures

  Facilitators. As part of the hiring process and parallel to a larger study assessing the influence of training quality on self-reported competency and comfort in their role as a program facilitator, facilitator participation in the current study began prior to their arrival for training. Facilitators were recruited by their organizational trainer for participation within the study (response rates provided below) and informed of the potential risks and benefits of their participation. Facilitators completed three questionnaires: prior to arrival for training (Time 1), after completion of training (Time 2), and at the end of the Summer Session (Time 3). The first questionnaire (see Appendix B) was completed electronically through a Qualtrics link prior to the facilitator's first day of training and provided through the administrative teams of the organizations providing training (May 1st - May 15th, 2016). This questionnaire consisted of demographic information, identifier information for tracking purposes across measurement occasion, information describing facilitator characteristics (e.g., their experience level, training level, program buy-in, and pro-implementation beliefs), and the FCPC Scale (described below). Additionally, the pre-training measure contained questions utilized for internal assessment of training quality by corresponding organizational leadership teams, outside of the focus of this study.

  The second facilitator questionnaire was administered in a paper format on-site immediately following completion of the four-day training. It also contained an identifier question (facilitator last name) to link responses to the first questionnaire, the SJTs, the FCPC, and internal measures of interest to assess training quality outside the scope of this

Table 5.

Camp Skill Development Operational Definitions

Outcome	Definition
Responsibility	The ability to start and accomplish tasks without external motivation [1]
Self-Regulation	The ability to adapt through stress and adversity [2]
Attitude	The ability to monitor and regulate behaviors[3]
Cooperation	The ability to work with others towards a shared goal[4]

Note: (1) Henderson, Bialeschki, & James 2007; (2) Ungar, 2012; (3) Glover et al., 2013; (4) Rhodes, Spencer, Keller, Liang, & Noam, 2006

study. Finally, the third facilitator questionnaire was administered via Qualtrics through the organizational administration teams at the completion of the summer camp season (approximately August 1st). It also contained an identifier question (facilitator last name) to link pre- and post-training responses, the FCPC, SJT responses, and internal measures of organizational interest.

Parents. Parent respondents were sent an email (through organizational administrative teams) one week following the completion of their child's camp experience with the questionnaire as a Qualtrics link (see Appendix C) embedded in the message. A week later parents received a reminder email to complete the survey if they had not already done so. Incentives (provided by the organizations) to participate in the study were offered in the form of entry in a drawing to receive one of three $100.00 gift-cards. The questionnaire (adapted from Garst & Gagnon, 2016a) contained 141 questions relating both parent and child demographics, satisfaction with the camp experience, prior camp experiences, parenting style, and outcomes parents observed in their child due to participation in a residential 5-7-day camp with one of the sites.

Measures

Parental Report of Outcomes. The *parent perceptions of developmental outcomes* (PPDO) construct(s) were operationalized as parental perceptions of their child's growth in four socio-emotional skills (See Table 5) resulting from their child's recent camp experience. This four-factor sub-scale was based upon an examination of the influence of camp director intention on programmatic outcomes conducted by Garst and Gagnon (2016a). In their study, Garst and Gagnon utilized this child-skill-development-assessment to determine if camp director practices (i.e., their intention levels towards program outcomes) influenced outcomes as reported by parents of campers. The authors utilized confirmatory factor analysis (CFA) and related statistics to determine the psychometric reliability and validity of the scale in a sample of 2,952 parents of residential campers. The results of the CFA suggested that the five factor PPDO demonstrated good convergent validity in terms of factor loadings (λ= .545 - .874), average variance extracted (AVE) scores (AVE = .548 - .672), and reliability (α = .825 - .894). Evidence of discriminant validity of the PPDO was indicated by relatively low between factor correlations (r = .446 - .689) and square root pf AVE[1] scores (.740 - .820).

While promising, the PPDO had a few limitations acknowledged by the researchers in its design and later assessment. The authors utilized a pre-post retrospective design to collect data; specifically, parents were first prompted with the statement: please compare your child's behaviors BEFORE camp with your child's

---

[1] Authors of the current study conducted square root AVE scores based upon statistics provided within the Garst and Gagnon (2016a) study, as they were not available within the manuscript.

behaviors AFTER camp, and consider how he/she may be behaving differently (or not) because of attending overnight camp this year. Parents were then presented with a behavior, for example "my child participates in a discussion" and asked to rate their child on that behavior both before and after the camp experience on a five-point Likert scale (1, strongly disagree to 5, strongly agree), with higher scores indicating higher levels of skill development. Pre-post retrospective approach designs are generally utilized due to problems associated with measurement standard invariance, incomplete data sets (e.g., respondents only completing one of two measures), and response shift bias (Howard & Dailey, 1979; Sibthorp, Paisley, Gookin, & Ward, 2007). However, this approach limits the validity of statements of growth and/or change shared by respondents due to their desire to show a learning effect (Lamb, 2005). Specifically, by presenting a respondent with an option to demonstrate growth, the researcher may unintentionally encourage a respondent to show more change than they would have if they did not have pre-program levels of skill readily available. In the current study, items were adapted from the pre-post retrospective design to a cross-sectional approach where items were prompted with: "As a result of camp my child now...participates in a discussion." Additionally, due to the shift in items from pre-post retrospective to a cross-sectional approach the initial 1-5 Likert scale was modified to a six-point Likert scale to extend the potential variability in responses: (1) Not at All to (6) A Great Deal More. The 1 to 6 range was also selected as it was the longest range possible within the Qualtrics software to remain "optimized for mobile" where all six score options would be viewable on one mobile device screen.

An additional challenge embedded within the Garst and Gagnon (2016a) study related to their management of missing data. Specifically, they utilized a technique known as listwise deletion, where if a respondent missed any items, the respondent is removed from the data set. The challenge with this technique is that a respondent who completed 99 of 100 questions and a respondent who completed 1 of 100 questions are treated as equivalent and removed from the study. While this approach is potentially normative within past research into OST and LER settings, it does not reflect contemporary methodological approaches regarding missing data management, specifically "ignoring this step is poor science, and results reported without attention to missing data can misinform our scientific understanding and misguide policy and practice" (Schlomer, Bauman, & Card, 2010, p. 8). In the current study, missing data were examined for systematic issues (e.g., patterns of missingness) and managed through two techniques: (1) full-information-maximum-likelihood (FIML) technique for initial psychometric testing and (2) an expectation maximization algorithm (EM) imputation technique for later multi-level modeling (both described in detail in the analyses section below). Garst and Gagnon's study was also a pilot in terms of the development of the PPDO items, and this was reflected in the number of items within the initial measure (35-items) compared to the final measure (consisting of 21-items). In the current study, the PPDO was adapted from the final 21-item five-factor measure as provided by Garst and Gagnon (2016a; 2016b) to reflect four factors and 19-items (see Table 5 for operational definitions). Specifically, as neither organization in the current study was focused on communication (i.e., the ability to articulate thoughts and feelings in a meaningful way)

as a targeted outcome of interest, the communication factor and corresponding items were not included in the current study.

Situational Judgement Tests. A purpose of this study was to determine the efficacy of SJTs as predictors of implementation quality. While the SJTs of interest for this study (See Appendix A) have not been tested as predictors of implementation quality, related components, or outcomes, there is a large body of evidence that suggests when SJTs are developed with both prior research and the use of SMEs, as they were within this study, they will have positive predictive qualities for future performance (Weekley & Ployhart, 2006). The six SJTs (see Table 2 for scenarios and operational definitions) utilized within this study were designed to reflect pro- and anti-implementation choices, where lower scores reflect a more pro-implementation attitude.

The Facilitator Characteristics and Program Contributions Scale (FCPC). A goal of this study was to determine the efficacy of SJTs (in comparison with and in addition to other measures of implementation quality), for predicting programmatic outcomes. In this study, the FCPC was utilized as it has been validated in similar LER contexts. The FCPC was designed to capture three factors that contribute to implementation quality and program outcomes: (1) *Program Buy-In*, the facilitator's level of belief and support that a program will achieve the desired outcomes (six-items), (2) *Pro-Fidelity Beliefs*, the level of facilitator support for delivering programs as designed (four-items), and (3) *Perceived Preparedness*, the facilitator perception they have sufficient training and experience to deliver a program well (seven-items) (Gagnon

& Bumpus, 2016). The 17 items were measured on a 1 (strongly disagree) to 7 (strongly agree) Likert scale, where higher scores indicated more positive responses to items.

The FCPC was designed to address a limitation embedded within many measures of implementation quality, that they are designed for a specific program and thus are not generalizable. Specifically, the FCPC measures were designed to be easily adapted to different programs. For example, in the study conducted by Gagnon and Bumpus (2016) one of the program buy-in items read "I am "bought-in" to the *Step One Leadership Program*. Following these design recommendations in a study also utilizing the FCPC, Gagnon, Garst, and Stone (2015a) modified the items to read "I am "bought-in" to the *CARE Now Program*." This planned modification of the items allows for the FCPC to be utilized in differing programs while maintaining the desired items and factors.

In the pilot study of the FCPC, Gagnon and Bumpus (2016) utilized Exploratory Factor Analysis (EFA) to determine which factors the items best fit. However, the use of EFA is problematic for several reasons. Specifically, the basic premise underpinning EFA is that it allows data to drive decisions rather than theory (capitalizing on chance relationships between variables), which is reflective of a pseudo-scientific approach to empirical research. In other words, EFA is designed to uncover theory and typically confirms assumptions, where 'true' scientific inquiry is intended to disconfirm or falsify theory (Popper, 1981). EFA approaches seem relatively amenable to confirmation bias, where the researcher(s) does not disconfirm their findings, rather they take an "oh that makes sense" approach (i.e., confirmation bias) and modify their findings accordingly (i.e., Type 1 error). Additionally, in the pilot study, the authors utilized composite scores

144

(e.g., items averaged and transformed into one factor based upon means) to explore hypotheses. The challenge with composite score factors is that they assume that each item within a factor perfectly contributes (e.g., factor loading of 1.00); however, as evidenced in the pilot study this assumption was not accurate as loadings ranged from .62 to .95. In other words, a composite factor (as compared to a latent factor) potentially compounds type 1 and type 2 error within a study as individual item contributions (or lack thereof) may be obscured or suppressed. As Gagnon and Bumpus's (2016) study of the FCPC was intended as a pilot, in combination with the relatively low sample size ($N = 28$) and lack of contemporary missing data procedures (e.g., the authors utilized listwise deletion), the findings were interpreted with a high degree of caution for inclusion within the current study.

A follow-up study conducted by Gagnon et al. (2015a) explored the measurement validity of the FCPC through a CFA, and produced an alternative model, where the three factors proposed by Gagnon and Bumpus (2016) were reduced to two factors; specifically, the pro-fidelity beliefs factor and program buy-in factor were merged into a single factor. However, as Gagnon et al. (2015a) indicated, "Facilitators ($N = 121$) from three distinct university programs were recruited to complete the FCPC at the conclusion of their various experiential education programs" (p. 4), the FCPC was investigated with differing programs with potentially varied missions and goals. This combination of data could be the cause for the reduction in factors. While the FCPC was designed to be a global measure of factors that contribute to implementation (Gagnon & Bumpus, 2016), it was not designed to be merged and analyzed with other program sources. More

145

specifically, while CFA was the appropriate analysis choice to determine the psychometric reliability and validity of the FCPC, an additional test was also necessary to determine if there were differences in measurement quality within the three groups tested. Referred to as *measurement invariance*, this test determines "the extent to which items or subtests have equal meaning across groups of examinees" (French & Finch, 2006, p. 379). In the study conducted by Gagnon et al. (2015a), the lack of a measurement invariance test indicates it is possible that one of the three groups contributed to measurement issues within the data set and correspondingly influenced the CFA results (e.g., the merging of the buy-in and pro-fidelity beliefs factors). As such in the current study, the original three-factor, 17-item FCPC as proposed by Gagnon and Bumpus (2016) was utilized.

Analyses

Data Preparation and Transformation. Prior to exploration of hypothesized relationships, the parental respondent data were prepared for an examination of their measurement properties and later multi-level-modeling. First, parental reports of outcomes were linked to facilitators by matching the reported last name of each child with that of each facilitator through an examination of the six site rosters. For example, in site ABCD, a parent reported the last name of their child as Smith, their child's dates of camp attendance, and other related demographic information in the online questionnaire. This response was then linked to site-specific rosters listing all available campers by last name and related demographic detail (e.g., Smith, Male, aged 14, week 4) for that corresponding facilitator.

146

After parental reports of their child's growth and counselors were linked, the child growth data were group mean centered. *Group mean centering* is the "transformation of a variable by taking each score and subtracting from it the mean of the scores (for that variable) for the group to which the score belongs" (Field, 2013, p. 875-876). Group mean centering is generally utilized when examining if "effects" are related to a group (Kreft & De Leeuw, 1998). More specifically, group mean centering removes group-specific (i.e., facilitator) variances and covariances for single level analyses. As it is theorized within this study that variance may occur at multiple levels, group mean centering allows for a CFA, without the added complication of potential "nesting" (Hoffman, 1997). This process was accomplished in three steps utilizing SPSS 24 software: (1) parental respondents were sorted by their child's linked facilitator, (2) assigned a mean for each score (i.e., response to each question on the PPDO scale) within that facilitator, and finally (3) this mean information was subtracted from their original score to create a group mean centered score. For example, in this study Facilitator McDonald's group reported a mean score of 3.10 to the question "As a result of attending Camp ABCD my son or daughter has a good mental attitude." This score is then subtracted from the Camper Smith's reported score of 4, creating a new group-mean centered score of .90 for Camper Smith, with the group mean now equaling zero.

After the variables were group mean centered, the data were screened for missingness to determine if they were Missing Completely at Random (MCAR) or Missing Not at Random (MNAR) utilizing Little's test of MCAR in SPSS 24 software (Little, 1988). Descriptive tests indicated that complete information was available for 549

parental respondents (85.92% of total sample) and no item (i.e., question) had a level

greater than 14.1% of missing values (11.9% to 14.1%). The non-significant results of

Little's test of MCAR [$\chi^2(243) = 245.377$, $p = .445$] indicated that the data was MCAR,

indicating that the use of a FIML technique would be appropriate to manage missing data

for preliminary psychometric testing of the CSD scale. This non-significant result also

demonstrates that within this study, the values of variables are not related in terms of

missingness (Little, 1988).

Parental Reported CSD Psychometric Testing. After preparation for analyses,

the data were analyzed to examine the psychometric properties of the CSD, specifically

for their validity and reliability through confirmatory factor analysis (CFA) in EQS 6.3

software. In this study, a multi-stage CFA was utilized, where the original measurement

model was specified with all items orientated towards one of the four theorized factors

(see table 5 for operational definitions). Then, through multiple stages, the model was re-

specified for issues such as error, items with poor unidimensionality, and items with

unusually high or low error covariances (Brown, 2015). Beyond the item level analyses,

the quality of model fit was examined with fit indices including the comparative fit

indices (CFI), Bentler-Bonnett non-normed fit indices (N-NFI), the root means square

error of approximation (RMSEA), and the Yuan-Bentler Scaled Chi-Square (Hu &

Bentler, 1999; Marsh, Hau, & Wen, 2004). The CFA results indicated that eight

respondents were contributing to multivariate kurtosis within the data set and as such

they were removed from further analyses. Additionally, due to evidence of non-normality

(e.g., items were negatively skewed) embedded within the data set (Yuan, Lambert, and

Fouladi Coefficient = 240.0372, Normalized Estimate = 120.6810) robust estimation techniques were utilized. The robust methods function is typically utilized "when the researcher is faced with non-normality" (Byrne, 2006, p. 138) in their data, indicated by the results of the normalized estimate being larger than 10.00 (suggesting data is high in skewness and/or kurtosis). Beyond removal of these outliers and the use of robust estimation techniques, the CFA results indicated that two items should also be removed from later analyses (i.e., "*takes initiative/is a self-starter*," Exploration Factor; "*is more helpful about the future*," Attitude Factor) due to relatively poor factor loadings in comparison to other items within those factors. Additionally, due to high levels of correlation (e.g., $r = .8 - .9$) between the four first-order factors, a second order factor was created, Camp Skill Development (CSD), to more effectively capture level of between-factor shared variance and to mitigate collinearity between first-order factors issues in later model testing. More simply, the decision to migrate the PPDO to a second order factor (e.g., the CSD) was driven by the results of an analysis with all first-order factors highly correlated. The final fit indices of the preliminary model were acceptable for later analysis: Yuan-Bentler $\chi^2(100) = 352.194$, $p \le .001$, RMSEA = .065 (90%, CI .058, .073), CFI = .983, and N-NFI = .980. In figure eight, the relationships between the second order factor (CSD) and the four first order factors (i.e., responsibility, exploration, self-regulation, and attitude) are presented. Descriptive information regarding the discriminant validity of the first order factors is provided in Table 6 in the form of between factor correlations and squared AVE scores. Additionally, information regarding the convergent validity of the group-mean centered CSD is provided in Table 7 including

Table 6.

Evidence of Discriminant Validity of Group Mean Centered CFA of CSD.

Factor	F1	F2	F3	F4
F1. Responsibility	**.922**			
F2. Exploration	.824*	**.938**		
F3. Self-Regulation	.913*	.832*	**.919**	
F4. Attitude	.885*	.911*	.898*	**.912**

* indicates $p \le .001$; **Bold** indicates √**AVE**

factor loadings, reliability in the form of Raykov's Rho (ϱ) and AVE scores. The aggregate of this information indicates, when group-mean centered at one level for MLM, the CSD is a psychometrically valid and reliable instrument. More simply, the CFA results indicated that when group level variance is removed due to the group mean centering process, the CSD is a psychometrically reliable instrument.

Facilitator Characteristics and Traits.

FCPC Psychometrics and Transformation. Due to the small number of facilitators ($N = 64$) within the study, the FCPC items were transformed into composite variables utilizing XYZ steps. First, the three factors embedded within the FCPC were examined for internal consistency utilizing Cronbach's Alpha and through an examination of inter-item correlations. The results of this analysis indicated that the pro-fidelity factor consisted of three items ($\alpha = .742$, $M = 4.233$ ($SD = .910$), the buy-in factor consisted of four items ($\alpha = .725$, $M = 6.481$, $SD = .502$), and the perceived preparedness factor consisted of six items ($\alpha = .845$, $M = 5.970$, $SD = .611$).

Facilitator SJT Responses. The SJTs in this study were developed based upon the results of a content analysis of Subject Matter Expert (SME) responses to scenarios they could encounter while providing summer camp programs (See Table 2). SMEs and
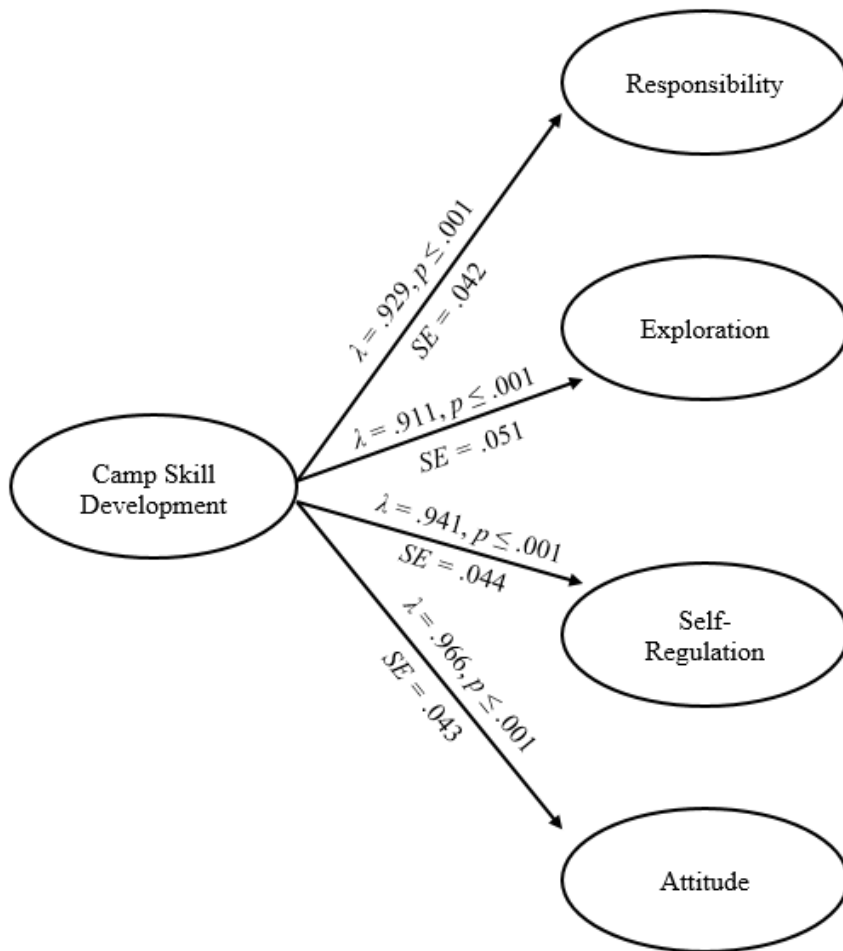
Figure 8. Presentation of Group-Mean Centered CSD Second Order Measurement Model. *Note:* Error terms, individual items, and constant (utilized for FIML simulation) excluded for parsimony of presentation.

later raters then condensed and ranked responses from low scores (indicating pro-implementation responses) to higher scores (indicating more anti-implementation responses). Facilitators within this study then completed the six "ranked" SJTs as part of their overall training (See Appendix A).

   Facilitator Characteristics. To examine hypothesized relationships two facilitator variables were transformed. To examine hypotheses H_{4a}, H_{4b}, and H_5, facilitator gender was "dummy-coded" where females equaled zero and males equaled one. Additionally,

for hypothesis H$_6$, a dichotomized experience variable was created from the continuous

facilitator experience variable; specifically, within SPSS 24, the grand mean centered

experience ($M = 0$, $SD = 1.056$) variable was examined for its 50% cutoff point, to

establish novice (equaling zero) and non-novice facilitator groups (equaling one).

Multi-Level Model Data Transformation. Within multi-level-modelling

research, there remains a lack of congruity on how to manage missing data. For instance,

Lüdtke, Robitzsch, and Grund (2017) highlighted how despite advancements in missing

data management there is insufficient attention to missing data analysis and management

in the context of MLM. More specifically, there is a lack of agreement on how to

simulate and/or impute missing data when nesting is hypothesized (e.g., when data may

be influenced at multiple levels), as is the case with this study. To address this challenge,

the parent respondent data were further transformed for later hypothesis testing. As the

results of the group mean centered CFA of the CSD indicated psychometric reliability

and validity when group level variance is removed (see Table 6 and 7) the variables of

interest were transformed from raw to grand mean centered scores for MLM analysis. In

the context of MLM group mean centering is done to determine how an individual is

different within a group, grand mean centering is done to see how an individual is

different from all groups (Field, 2013). In this study, grand mean centered variables were

created utilizing the "mean-center" utility within SPSS 24. After grand mean centered

variables were created, missing data were then addressed. As Little's test of MCAR

indicated the data were MCAR, missing data were generated through the expectation-

maximization (EM) algorithm. According to Kline (2011):

in the E (expectation) step, missing observations are imputed by predicted scores

in a series of regressions in which each incomplete variable is regressed on the

remaining variables for a particular case. In the M (maximization) step, the whole

imputed data set is submitted for ML estimation. These two steps are repeated

until a stable solution is reached across the M steps. (p. 59)

Specifically, in this study, EM was utilized within EQS 6.3 software to generate (i.e.,

impute) missing data based upon the previously developed grand mean centered scores.

Having earlier established the validity of the CSD model through CFA, an MLM

exploring hypothesized relationships (See Table 4) was then conducted.

Multi-Level-Modeling. The goal of this study was to examine the effectiveness

of SJTs, the FCPC and/or facilitator characteristics as a method to predict implementation

quality and corresponding parent perceptions of outcomes. In this study, multi-level-

modeling (MLM) was utilized to examine relationships among facilitator SJT responses,

facilitator characteristics, FCPC scores, and parent reported outcomes (i.e., CSD scores).

As noted earlier, much of the research exploring implementation quality fails to

acknowledge the multiple levels with which implementation quality takes place.

However, "at present, multilevel statistical models afford the best means to represent the

data structures that typically arise when multiple fidelity indicators are used to

characterize the manner in which providers deliver an intervention to recipients" (Zvoch,

2012, p. 549).

At a foundational level, *MLM* is a linear procedure where the "structure of the

data is explicitly considered" (Field, 2013, p. 880). More specifically, the process of

Table 7.

Group Mean Centered CFA Results of CSD Second Order Model.

Factor/Item	SD	λ	ϱ	AVE
Camp Skill Development			.966	.878
Responsibility	-	.929		
Exploration	-	.911		
Self-Regulation	-	.941		
Attitude	-	.966		
Exploration			.967	.881
…participates in new learning experiences	1.393	.925		
…is curious about new topics and subjects	1.425	.933		
…seeks challenges beyond his / her comfort zone	1.451	.949		
…is willing to try new experiences	1.461	.947		
Responsibility			.966	.851
…takes responsibility for his / her own actions	1.241	.909		
…takes care of his / her own things	1.258	.911		
…shares work responsibilities	1.288	.935		
…follows through when asked to do something	1.234	.925		
…follows directions	1.265	.931		
Self-Regulation			.942	.844
…properly handles success and failure	1.293	.950		
…manages disappointment well	1.291	.944		
…deals effectively with conflict	1.181	.860		
Attitude			.952	.832
…doesn't get frustrated easily	1.195	.828		
…has a good mental attitude	1.306	.945		
…has a generally "positive" view on life	1.354	.939		
…shows a positive attitude when around others	1.373	.931		

Note: *indicates first order factor*; *λ*: standardized coefficient (factor loading); **ϱ**: Raykov's Rho; **AVE**: Average Variance Extracted.

MLM examines data at multiple levels. In this study, two levels of data were hypothesized within the data, the facilitator (level 2) and the parent reported outcomes (level 1) (see Figure 9). MLM is frequently utilized within educational settings where students represent a level (level one), classrooms and/or teachers represent another (level two), and schools may represent another (level three) (Hoffman, 1997). In this example as each classroom may provide a unique environment, MLM allows for differentiation

between classrooms, specifically determining if there are unique effects and relationships present due to the classroom environment, thus allowing for between classroom comparison. Moreover, as each school may provide a unique setting, MLM allows for between- school comparisons. MLM is advantageous over single level analyses such as ANOVA and linear regression as it takes full advantage of these nested designs. Hox and Roberts (2011) described the issues associated with using ANOVA and regression for nested designs, "historically the problem of analyzing data from individuals nested within groups was solved by moving all variables by aggregation… to one single level, followed by some standard single level approach" (p. 4) such as a regression or ANOVA procedure. In these single level processes, much of the variation that would be present is either absorbed into one level (and thus hidden) or misinterpreted, leading to a higher chance of a type I error (e.g., saying there is an effect present when there is not one) and/or a type II error (e.g., saying there is no effect when there is one).

As the purpose of this study was to examine the effectiveness of SJTs, the FCPC and/or facilitator characteristics as a potentially cost-saving method to predict implementation quality and corresponding parent perceptions of outcomes, MLM allows for this differentiation, whereas a single level model would group the influences and potentially suppress or magnify effects. Another consideration on why the use of MLM in this study is appropriate is offered by Hoffman (1997), "Hierarchical relationships occur when at one level of analysis influence or are influenced by variables at another level of analysis" (p. 724). As many of the hypotheses (See Table 4) suggest, a central tenant of this study is that the individual camper is nested within a facilitator, highlighting the

interplay between variables within this study. Furthermore, MLM allows for uneven sample sizes and missing data between facilitators whereas single level measures (e.g., RMANOVA) do not (Field, 2013). A visual example of the potential levels in this study is presented in Figure 9.

Results

Data were collected from two groups in this study: (1) parents reporting on their child's growth resulting from the camp experience and (2) facilitators responsible for providing the camp experience, linked to a specific camper through a parent reported child last name and corresponding match to a specific facilitators' roster.

Sample Descriptives

Parents. As mentioned earlier, parent respondents were solicited through an email one week following the conclusion of their child's camp experience and sent a reminder one week afterwards (i.e., reminder was sent 14 days' post-camp experience). Initially 1,607 parents were sent an email through email management software (i.e., Mail Chimp and ProClass) to participate in the study, of these emails eight were rejected due to a spam filter on the respondents' email address. To incentivize participation respondents were offered entry into a drawing for one of three $100.00 Amazon gift cards, one for every month of the study. To further increase engagement within the study winners were announced via a social media post on the organizational Facebook page including a photo of the past month's winner, demonstrating that someone won the drawing. The combination of these strategies resulted in a total of 613 respondents indicating a 36.72% response rate to the questionnaire. Further matching of parent

156
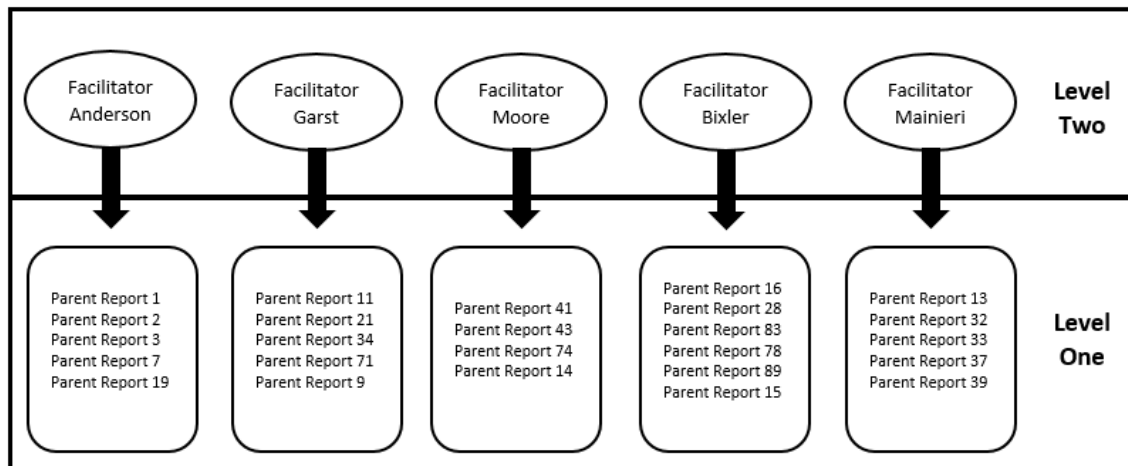
Figure 9. Example of proposed MLM.

reported data to facilitators resulted in removal of 21 respondents to the main study, as these data were not reportable within the context of the study research questions (e.g., matching facilitator and parent reports). Additional screening for multi-variate outliers and respondents who completed less than 50% of the questionnaire resulted in 4 and 5 respondents, respectively, being removed from the sample, indicating a sample of 583 parent respondents.

Parental respondents in this study primarily identified as female ($n = 467$, 83.2%), married ($n = 449$, 80.3%), white ($n = 497$, 88.4%), high-income earners, with 80.5% of the sample reporting annual household income above the national median of $56,516 dollars per year [($M = \$137,733.05$, *Mdn* $= \$125,000$, $SD = \$69,166.14$) see Federal Reserve Bank of St. Louis, 2016], and well-educated, with 75.9% ($n = 422$) of the sample reporting a Bachelors' degree or higher. Parents reported an average 2.52 years ($SD = 2.80$ years) attending summer camp themselves (range 0 - 15 years), with 33.9% ($n = 184$) of respondents indicating no prior camp attendance. Parents were also asked to provide descriptive information regarding their child. Parents indicated that children were

primarily attending camp for the first time (n = 410, 71.2%, M = 1.45 years, SD = .867 years), primarily identified as male (n = 338, 57.9%), were majority white (n = 492, 84.8%), and ranged in age from 7 to 19 years, with an average age of 12.55 years (SD = 2.59 years, Mdn = 12 years). Additional parent and child descriptive information is provided in Table 8.

Facilitators. Data were collected from facilitators at three time points: (1) 1 week prior to training, (2) immediately following training completion, and (3) at the end of the eight-week summer season. To address potential missing data necessary for hypothesis testing, facilitators were asked to provide demographic information. Specifically, if a facilitator did not complete the pre-training measure, data describing their demographic information (e.g., gender, age, education level) would still be available for later analysis. Within the three measurement occasions, all facilitators completed the pre-training measure (N = 64), 61 of 64 completed the post-training measure (95.31%), and 57 of 64 completed the end-of-summer measure (89.06%). Facilitators primarily identified as white (n = 42, 65.6%), with African American (n = 13, 20.3%), Multiple Race (n = 7, 10.9%), and Hispanic or Latino Origin (n = 2, 3.1%) representing the remainder of the sample. Facilitators were evenly split between genders (female = 32, 50%; male = 32, 50%) and ranged in age from 18 to 29 years with an average age of 20.88 years (Mdn = 21 years, SD = 1.95 years). The majority of facilitators were current undergraduate students (n = 56, 87.1%) with an average of 2.91 years of time (SD = 1.36) as a college student; the remaining 8 facilitators (12.5%) reported current possession of a Bachelors'

Table 8.

Additional Parent and Child Descriptive Information (N = 584)

Variable	Descriptive Information			
Parent Ethnic Group	African American (*n* = 43, 7.7%)	White (*n* = 497, 88.4%)	Hispanic or Latino Origin (*n* = 7, 1.2%)	
	Asian Origin (*n* = 4, .7%)	Native American (*n* = 1, .2%)	Multiple Race (*n* = 7, 1.2%)	
Parent Relationship Status	Single (*n* = 21, 3.8%)	Divorced (*n* = 31, 5.5%)	Widowed (*n* = 4, .7%)	Married (*n* = 448, 80.3%)
	Re-Married (*n* = 28, 5%)	Separated (*n* = 11, 2%)	Long-Term Relationship (non-married) (*n* = 15, 2.7%)	
Parent Education Level	Less than High School (*n* = 2, .4%)	High School (or equivalent) (*n* = 19, 3.4%)	Some College (No Degree) (*n* = 52, 9.4%)	Technical Degree or Certification (*n* = 17, 3.1%)
	Associates Degree (*n* = 44, 7.9%)	Bachelors' Degree (*n* = 218, 39.2%)	Masters' Degree (*n* = 156, 28.1%)	Doctorate (MD, Ph.D., or J.D.) (*n* = 48, 8.6%)
Child Ethnic Group	African American (*n* = 47, 8.1%)	White (*n* = 492, 84.8%)	Hispanic or Latino Origin (*n* = 11, 1.9%)	
	Asian Origin (*n* = 5, .9%)	Asian (Indian or Arabic Origin) (*n* = 2, .3%)	Multiple Race (*n* = 23, 4%)	

degree or higher. Facilitators reported an average of 1.63 years facilitating, managing, and/or leading groups (SD = 1.94 years, Range = 0 to 8 years).

MLM Intraclass Correlation Testing

Prior to exploring hypothesized relationships, an examination of intraclass correlations (ICC) was conducted. ICCs represent a test "of dependency in data within the same context" (Field, 2013, p. 877), or the degree to which "nesting" is present within data. If an ICC is equal to zero it indicates that "no group differences exist for the variables of interest. People within the same group are as different from each other on

these variables as people across groups are" (Kreft & De Leeuw, 1998, p. 4). Conversely, the closer an ICC gets to 1.0 there is a greater probability that group does influence the variable of interest (Field, 2013). More simply, if the ICC equals one then all variance is at the higher level and if it equals zero then all variance is at the lower level (Hoffman, 1997). As such, ICC's were examined for the variables of interest in EQS 6.3 software. In table 9, the ICCs are presented for the CSD outcome variables. As indicated within table 9, CSD ICCs ranged from .012 to .067, the question of course being "how small is small." (Byrne, 2006, p. 398). While many ICCs in this study are low, the "worst" ICC accounting for only 1.2% of variance at level two (% of level two variance = .012 * 100), they are not zero, thus, indicating the possibility of between level effects. Additionally, some MLM research indicates that "if the ICC values are nonzero (< .05 or .10), then multilevel" (Cho, Lee, Moore, Norman, & Ramshaw, 2017, p. 7) modeling should be performed. Further, participant gender indicated a large ICC (ICC = .590); however, this high ICC for gender only indicates that there were more female participants than male participants in the sample.

Multilevel Confirmatory Factor Analysis

Prior to examining the eight multi-part hypotheses (See Table 4 for summary) a CFA of the full MLM was conducted. The results of which indicated acceptable model fit Bentler-Liang $\chi^2(200) = 656.050$, $p \leq .001$, RMSEA = .094 (90%, CI .086, .102), CFI = .963, and N-NFI = .956. Further testing of reliability through Raykov's Rho (ϱ) indicated acceptable level of internal reliability across first-order factors, second-order factors, level one (child), and level two (facilitator) (i.e., ϱ = .943-.986). Further, the results of

Table 9.

Intraclass Correlations of Level One Variables

Variable	ICC
Exploration	
…participates in new learning experiences	.033
…is curious about new topics and subjects	.044
…seeks challenges beyond his / her comfort zone	.022
…is willing to try new experiences	.045
Responsibility	
…takes responsibility for his / her own actions	.036
…takes care of his / her own things	.030
…shares work responsibilities	.024
…follows through when asked to do something	.067
…follows directions	.049
Self-Regulation	
…properly handles success and failure	.029
…manages disappointment well	.024
…deals effectively with conflict	.022
Attitude	.012
…doesn't get frustrated easily	
…has a good mental attitude	.028
…has a generally "positive" view on life	.016
…shows a positive attitude when around others	.015
Participant Gender	.590

squared AVEs scores at both levels indicated acceptable levels of discriminant validity

($\sqrt{\text{AVE}}$ = .914 - .973); however, correlations between factors were not available due to

the second order nature of the CSD factor. The aggregate of the information indicates that

the CSD is psychometrically valid at two levels, and thus appropriate for hypothesis

testing.

Multilevel Hypothesis Testing

      To explore the eight multi-part hypotheses (See Table 4) a multilevel model was

examined with all hypothesized relationships embedded. The MLM indicated acceptable

levels of fit: Bentler-Liang χ^2(457) = 730.069, $p \leq$ .001, RMSEA = .051 (90%, CI .044,

.058), CFI = .975, and N-NFI = .970. In the sections below, the results of each hypothesis are shared.

Hypothesis One. In the first hypothesis, it was hypothesized that facilitator pro-implementation beliefs would have a positive effect on parent perceptions of programmatic outcomes. The results indicated that at level two, facilitator pro-implementation beliefs had a significant and negative effect on parent perceptions of program outcomes (β = -.405, p = .054, SE = .030), indicating that for every-one unit increase in pro-implementation beliefs, CSD outcomes decreased by .405, suggesting that, hypothesis one was not supported; moreover, the exact opposite of the hypothesized effect was uncovered

Hypothesis Two. In the second hypothesis, it was hypothesized that facilitator buy-in would have a positive effect on parent perceptions of program outcomes. The non-significant MLM results indicate that this hypothesis was not supported (β = .037, p = .778, SE = .033)

Hypothesis Three. In the third hypothesis, the effect of facilitator perceived competence (e.g., FCPC preparedness) on parent perceptions of program outcomes was investigated. The non-significant MLM results indicated that this hypothesis was not supported (β = -.262, p = .137, SE = .038)

Hypothesis Four. Hypothesis four consisted of two dimensions: first that male facilitators would have better parent perceptions of program outcomes, and second, that facilitator gender would not influence perceived implementation quality (e.g., FCPC dimensions). The results of the MLM suggest that facilitator gender did not have a

162

Table 10.

MLM CFA Results

Factor/Item	L1 λ	L2 λ	L1 ϱ	L2 ϱ	L1 AVE	L2 AVE
Camp Skill Development			.966	.980	.876	.924
Responsibility	.935	.979				
Exploration	.904	.947				
Self-Regulation	.941	.928				
Attitude	.963	.990				
Exploration			.967	.986	.880	.947
…participates in new learning experiences	.924	.975				
…is curious about new topics and subjects	.933	.980				
…seeks challenges beyond his / her comfort zone	.948	.961				
…is willing to try new experiences	.948	.976				
Responsibility			.967	.970	.852	.868
…takes responsibility for his / her own actions	.913	.825				
…takes care of his / her own things	.916	.957				
…shares work responsibilities	.938	.974				
…follows through when asked to do something	.932	.953				
…follows directions	.917	.942				
Self-Regulation			.943	.962	.848	.894
…properly handles success and failure	.952	.987				
…manages disappointment well	.943	.948				
…deals effectively with conflict	.865	.900				
Attitude			.953	.964	.836	.871
…doesn't get frustrated easily	.840	.799				
…has a good mental attitude	.948	.969				
…has a generally "positive" view on life	.938	.976				
…shows a positive attitude when around others	.928	.976				

Note: *indicates first order factor*; λ: standardized coefficient (factor loading); ϱ: Raykov's Rho; **AVE**: Average Variance Extracted; L1: Level One; L2: Level Two

meaningful influence on parent perceptions of program outcomes ($\beta = .133$, $p = .402$, *SE* = .046), nor on pro-fidelity beliefs ($\beta = .176$, $p = .170$, *SE* = .257). Additionally,

facilitator gender did not influence facilitator buy-in ($\beta = .208$, $p = .097$, *SE* = .145), but it

did significantly influence perceived preparedness ($\beta = .260$, $p = .026$, *SE* = .156);

specifically, this indicates that male facilitators rated themselves as more prepared. However, in aggregate these results indicate that hypothesis four was not supported.

Hypothesis Five. The effect of gender on SJT response choice was explored in hypothesis five, specifically it was hypothesized that facilitator gender would not have a meaningful influence on SJT response choice. In five of the six SJTs there was no significant effect of facilitator gender on SJT response; however, in the adaptation scenario, males were more likely to select anti-implementation responses ($\beta = .304$, $p = .015$, $SE = .309$), thus, indicating only partial support for hypothesis five.

Hypothesis Six. Hypothesis six consisted of five dimensions, specifically that experience measured as a dichotomous variable would have no effect on (1) facilitator buy-in, (2) facilitator perceived preparedness, (3) facilitator pro-fidelity beliefs, (4) SJT response choice, and (5) parent perceptions of program outcomes. The majority of results indicate that as a dichotomous variable, facilitator experience had no significant effect on: (1) facilitator buy-in, (2) facilitator perceived preparedness, (3) facilitator pro-fidelity beliefs, or (4) parent perceptions of program outcomes. However, the results did indicate that facilitator experience as a dichotomous variable had a negative effect on the adherence scenario ($\beta = .585$, $p \leq .001$, $SE = .387$), indicating that as facilitators shifted from novice to non-novice they were more likely to select a pro-implementation choice within the adherence SJT. However, as a dichotomous variable, facilitator experience did not have a significant influence on the remaining SJTs.

Hypothesis Seven. Similar to hypothesis six, hypothesis seven consisted of five dimensions. However, within this hypothesis facilitator experience was measured

continuously and hypothesized to have no effect on: (1) facilitator buy-in, (2) facilitator perceived preparedness, (3) facilitator pro-fidelity beliefs, (4) SJT response choice, and (5) parent perceptions of program outcomes. The MLM results suggest that as a continuous variable facilitator experience did not have a significant influence on (1) facilitator buy-in, (2) facilitator perceived preparedness, (3) facilitator pro-fidelity beliefs, or (4) parent perceptions of program outcomes. However, the results suggested, that as a continuous variable, facilitator experience may influence SJT response choice. Specifically, as facilitators increased in experience, they were more likely to select a pro-implementation choice within the quality of delivery ($\beta = .358$, $p = .037$, $SE = .079$) and program differentiation ($\beta = .171$, $p = .028$, $SE = .181$) SJTs.

Hypothesis Eight. Within hypothesis eight, the effect sizes of SJT response choice were compared to the three dimensions of the FCPC to determine which had a greater effect on parent perceptions of program outcomes. The results of this hypothesis are mixed. Specifically, neither the full six-item SJT nor three-dimension FCPC had a significant effect on program outcomes. However, the quality of delivery SJT and pro-implementation dimension of the FCPC did have significant effects on CSD outcomes. Specifically, as noted in hypothesis one, as pro-implementation beliefs increased CSD outcomes decreased ($\beta = -.405$, $p = .054$, $SE = .030$); additionally, as facilitators selected more pro-implementation choices on the quality of delivery SJT, CSD outcomes increased as theorized ($\beta = -.562$, $p = .037$, $SE = .079$). The results of this test indicate that one dimension of the SJTs developed for this study were better predictors of outcomes than the FCPC in terms of their larger effect size. However, further testing of

these effects sizes did not yield statistically significant differences, indicating a lack of support for hypothesis eight.

Discussion

The goal of this study was to examine the effectiveness of SJTs, the FCPC, and/or facilitator characteristics as a potentially cost-saving method to predict implementation quality and corresponding outcomes. In the sections below the implications of this study are discussed, limitations are explored, and future research directions are established. The finding of facilitator pro-implementation beliefs having a negative effect on program outcomes was surprising as it is counter to the findings of prior studies utilizing the same measure (e.g., Gagnon et al., 2015a), and of independent studies examining pro-implementation attitudes (e.g., Mihalic & Irwin, 2003). There are many potential reasons for this opposite effect including measurement issues (FCPC scores were only examined for internal reliability and "composited"), unknown confounds to camp programming, and a limited sample size. However, an arguably simpler reason for this opposite effect is that within residential summer camp, pro-fidelity beliefs may simply be lower due to the "be flexible" culture imparted to many residential summer camp facilitators.

The null effects of facilitator buy-in and perceived competence (i.e., preparedness) on program outcomes were surprising given the evidence suggesting that these constructs should positively predict outcomes within LER contexts (e.g., Gagnon & Bumpus, 2016; Morgan et al., 2016). However, there is a non-LER body of research (e.g., Wehby et al., 2012) that suggests otherwise, demonstrating that buy-in does not influence implementation quality or outcomes. While in the instance of this study it is possible that

facilitator buy-in may not influence outcomes within residential summer camp, a more likely cause is the limited range within FCPC responses in this study. Specifically, for facilitator buy-in, the non-centered mean score was 6.481 ($SD = .502$) with a range of 5 to 7 measured on a 1-7 scale. In this instance, higher scores indicate higher buy-in, thus, a likely explanation for the lack of effect of buy-in on outcomes was the lack of variance within the variable. A similar "floor" effect was indicated within the perceived competence variable ($M = 5.970$, $SD = .610$, range 4.5 to 6.67). This indicates a likely measurement issue within these constructs, where future studies should explore how to create variance within these measures. Additionally, two further explanations for the variance issues within these constructs may be: (1) there could be a degree of social desirability where facilitators want to be perceived as competent and "bought-in" to their organizations and (2) the organizations could truly be fostering a culture where high buy-in and preparedness are normative, thus the limited range should be expected.

The lack of effect of facilitator gender on program outcomes ($H4_a$) and perceived implementation quality ($H4_b$), while counter to the proposed hypotheses, was not surprising given the aggregate of implementation research suggesting that gender does not have a meaningful effect on implementation (Baker et al., 2010; Dolcini et al., 2014). This does not indicate that the science on gender's effect is "settled," rather it indicates that the null effect facilitator gender within this study supports some of the prior research into the effect of gender (e.g., Berkel et al., 2011). In future research, the effect of gender should be examined at a more hierarchical level, specifically how an organizational leadership's and program trainer's gender may influence facilitator's implementation

167

quality. Studies examining the race and/or ethnicity of program facilitators have indicated that when trainees are the same race and gender of trainers, better program outcomes can ensue (e.g., My Brother's Keeper, Whitehouse.gov, 2016). While facilitator gender did not have a meaningful influence on buy-in or pro-fidelity beliefs, it did on perceived competence. Specifically, male facilitators reported higher levels of preparedness than their female counterparts. Outside of the implementation literature, there is an ample body of research indicating that males both tend to overrate their ability and preparedness, but also that this overrating is socially expected (e.g., expectancy value models, see Wigfield, Tonks, & Eccles, 2004). In future research, the effect of gender on perceived preparedness should be compared to actual competency assessments to determine the relationships between the two, and corresponding effect on implementation quality and program outcomes.

In hypothesis five, the effect of gender on SJT response choice was examined, as a preponderance of SJT related literature suggests SJTs are less vulnerable to gender bias than other measures of personality (e.g., Weekley & Ployhart, 2006). This hypothesis was supported in five of the six SJTs; however, within the adherence SJT, male facilitators tended to select more anti-implementation responses than their female counterparts. In future investigations of implementation, the SJTs utilized within this study should be examined for any potential female- or male-centered bias.

In multipart hypotheses six and seven, the influence of facilitator experience was examined utilizing two approaches common to implementation research and experience, (H_6) as a dichotomous variable (e.g., novice and non-novice facilitator) and (H_7) as a

168

continuous variable, in number of years. Facilitators reported their level of experience in number of hours utilizing an index provided below the question; this number was then transformed into a "year-based" number. For example, if a facilitator reported six months of experience, their year-based number would be .5 (i.e., 6/12=.5). In the case of hypotheses six and seven, the majority of both hypotheses were supported, indicating a null effect on (1) facilitator buy-in, (2) facilitator perceived preparedness, (3) facilitator pro-fidelity beliefs, and (4) program outcomes. However, the results indicated less support for H_{6d} and H_{7d}, specifically facilitator experience had a positive effect on some of the SJT responses. This suggests that facilitators with more experience are likely to engage in more pro-implementation activities when presented with a situation rather than the more abstract questions embedded within the FCPC, and is congruent with SJT research outside of the field of implementation (e.g., Whetzel & McDaniel, 2009).

Beyond the lack of effect of experience on many of the variables within this study, it is also likely that the effect of facilitator experience is reduced due to factors outside of the facilitator's control. For instance, it may be that organizational leaders are pairing their more challenging participants with their more experienced facilitators, thus suppressing the effect of experience on the variables of interest. Similar "triage" occurs in the healthcare fields; for instance, the Cleveland Clinic is renowned for its cardiac unit, paradoxically it also has one of the highest mortality rates in the country. Does this mean this reputation as a renowned clinic is undeserved? The answer is likely more nuanced, Cleveland Clinic's reputation attracts many of the sickest patients, thus mortality may be more likely within this group. As such, in future investigations a deeper assessment of

participant skill level may uncover differing results regarding the relationship between experience and implementation. Furthermore, while both versions of the experience variable influenced SJT response choices, they influenced different SJTs. This suggests a measurement issue remains embedded within the experience variable. In future research exploring the effect of facilitator experience, there should be more investigation into what constitutes "good" experience, non-linear experience growth, and what does not constitute experience (Desimone & Lee-Hill, 2017; Weekley & Ployhart, 2005).

In the final hypothesis, the effects sizes of the FCPC and SJT were compared. The results indicated that the pro-implementation component of the FCPC negatively predicted outcomes and was incongruent with much of the theoretical development in this area, specifically, facilitator pro-implementation beliefs should positively predict outcomes (Gagnon & Bumpus, 2016). Conversely, the results also suggested that the quality of delivery SJT positively predicted CSD outcomes. The incongruence in findings suggests that as noted within discussion of hypothesis one, a measurement issue is likely to blame or a culture of flexibility is present within the sites examined in this study.

Limitations

While many of the study limitations were captured early in the discussion, a few warrant a deeper examination. The programs delivered across the six sites had a regimented and schedule-centric design with clearly communicated outcomes; however, data regarding actual adherence to this schedule was not captured as part of this study. This additional data could provide more context on why many of the implementation variables did not meaningfully influence outcomes. An added challenge for this study

was all information gathered was self-reported and required facilitators to provide their name. While explicitly communicated to facilitators that confidentiality would be maintained, a degree of social desirability may have influenced results. Further, the effect of training on facilitator ability was not explored as a part of this study, creating a potential gap within the data that could illustrate implementation-centric training and the effect on outcomes and later implementation quality. While this study had a high level of response from facilitators, the sample size was near the minimum necessary to conduct a MLM (see Kreft & De Leuw, 1998). Further testing of sample size to explore power levels necessary to detect an effect indicated that a sample of 62 facilitators was inadequate ($\lambda = 7.11$) where a necessary λ level would have been 16.24 to detect an effect indicating an approximate sample of 130 facilitators may have uncovered additional effects.

Additionally, education level and age of facilitators were not incorporated into the model due to their narrow ranges (2-3 years of college, 18-20 years of age), a more diverse sample in terms of age and education may produce differing results. Finally, both samples (i.e., parents and facilitators) were homogenous in terms of reported ethnicity, indicating again a more diverse sample could have produced differing effects.

Conclusion

While the findings of this study were mixed, this is one of the first to examine implementation in a multilevel format within LER and the broader social sciences (Zvoch, 2012), indicating a "proof-of-concept" for future research. There are many exciting potential avenues to examine regarding implementation prediction in the future

171

(video-based SJTs, peer, supervisor reports), the bottom line is "does implementation

matter?" Unfortunately, this study does not bring much clarity to this question. Indeed,

the quote of Neil deGrasse Tyson rings true, "In science, when human behavior enters the

equation, things go nonlinear. That's why Physics is easy and Sociology is hard." More

simply, more work remains.

CHAPTER FIVE: DISCUSSION, IMPLICATIONS, FUTURE DIRECTIONS, AND LIMITATIONS

The purpose of this dissertation was to explore how a facilitator's traits, characteristics, skills, and behaviors relate to implementation quality through three related papers. To this end three outcomes were achieved: (1) in chapter two, a framework of macro and micro level factors that influence implementation quality was presented, (2) in chapter three, new measures to predict facilitator implementation quality were developed, and (3) in chapter four, through a multi-level model, the relationships among facilitator traits, characteristics, skills, behaviors, outcomes, and implementation quality were explored. This chapter discusses the results of this dissertation, broader study limitations are shared, and implications and future directions for practice and research.

Discussion

This dissertation sought to continue the shift of LER towards evidence-based design and practice informed by contemporary implementation research. Given the mandate of many LER professionals to improve the quality of life for the communities and constituents they serve, this is an important goal reflecting the desired outcomes of the prevention, social, and health sciences. As noted throughout chapters two, three, and four, the study of implementation is complex. This complexity is highlighted within the framework (see Figure 4) and measurements (see Table 1) produced within chapter two, where at the facilitator level at least 23 unique facilitator characteristics, traits, or behaviors can influence implementation quality. Notably, this framework is based on a

"siloing" of these facilitator characteristics, traits, or behaviors, potentially failing to test for exponential combinations of additional factors. Furthermore, as noted within the findings of chapter 4, there are unique combinations of facilitator characteristics, traits, or behaviors that can influence implementation quality, program outcomes, and interestingly, other dimensions of facilitator characteristics, traits, or behaviors. While potentially off-putting, these findings only further illustrate the importance of capturing the 23 unique facilitator characteristics, traits, and behaviors for later analyses. In other words, it may be difficult to conduct the analyses with the highlighted characteristics, traits, or behaviors when analyzing relationships, but it may be impossible to capture them later. Thus, the new framework and measurement recommendations could serve as a strong starting point for understanding the facilitator role and relationship with implementation quality. However, the best way to eat an elephant is one bite at a time (Hogan, 2011); more simply, this study indicated that at a broader level, the study of factors influencing implementation quality can be done, but with a high degree of recognition of the multitude of factors that can influence implementation quality.

Within chapter three, the development of the SJTs indicated surprising SME responses to implementation challenges. Specifically, when SMEs were faced with challenges to implementation quality across all six dimensions (see Table 2), SMEs almost universally recommended changing the program plan, thus compromising implementation quality. While this finding highlights a degree of internal consistency across SJTs, it also indicates the need for greater promotion of the importance of implementation quality within LER. Conversely, this finding could indicate that within

the LER context of this study, residential summer camp, the implementation components

explored may not matter. Furthermore, as noted in chapter four, the narrow response

ranges to SJTs by less experienced facilitators suggests that this "adapt and overcome"

mentality is also embedded within newer program facilitators. More simply, when faced

with maintaining a program plan or modifying it, facilitators within this study almost

universally chose modification. However, the findings of chapter 4 also indicated in some

instances SJT responses reflect more pro-implementation choices and higher CSD scores

(i.e., programmatic outcomes) occurred, supporting the broader implementation

literature. More simply, implementation quality related behaviors do matter.

Furthermore, as illustrated the null effect of both measures of facilitator

experience on implementation quality and program outcomes reflects prior

implementation research (e.g., Desimone & Lee-Hill, 2017; Domitrovich et al., 2015;

Lillehoj et al., 2004). In fact, the supposed beneficial effect of higher levels of facilitator

experience on program outcomes and implementation quality (Dusenbury et al., 2005)

continued to be elusive within this study. Indeed, as indicated within chapter four,

facilitator experience was negatively associated with program buy-in and pro-

implementation SJT choices. This may indicate that experienced facilitators need

differing levels of involvement and/or training when providing programs to mitigate their

influence on implementation quality (in the form of SJT responses) and program buy-in.

As noted by Durlak and DuPre (2008), Berkel et al. (2011), and Moore et al.

(2013), much of the research regarding implementation quality has taken place in school-

based contexts, rather than community or family settings. The reasons for this skew

towards school-based research is likely due to a combination of funders (a) requiring

implementation assessment as part of program studies (Wanless & Domitrovich, 2015b),

(b) a relatively longer history of implementation research within school settings (e.g.,

treatment integrity, Gresham, 1989), and (c) a longer history of evidence-based practice

also taking place within educational settings. Additionally, as education is arguably under

more constant pressure to demonstrate value than many community based programs, a

natural consequence may be a deeper recognition of the importance and influence of

implementation quality as part of outcome achievement. As much of the research

underpinning implementation science is grounded within educational contexts, the

challenge may be greater for program designers and providers outside of the educational

context; specifically, in their contextual understanding of the factors that contribute

to/detract from implementation quality within their non-school programs. Is it appropriate

to suggest that the same factors that detract from or contribute to implementation quality

within school contexts would be replicated within community-based contexts? Although

some evidence suggests in-school time and out-of-school time programs can influence

similar outcomes, suggesting a degree of cross-over between the two settings, there is

less evidence indicating that the same programs work as well within and outside of

school time. This contrast suggests that there may be differing influences to

implementation quality in either setting and concurrently differing levels of importance to

the components compromising implementation quality. In future studies, investigation

into why implementation quality influences outcomes in differing contexts could

highlight how in some programs it may be the context that drives participant growth more

than that of an evidence-based program. In the context of LER research this difference could manifest in a comparison of a multi-day small group backpacking organization, a multi-week after-school youth sport program, and a semester long in-school program where the same substance abuse program is provided, but the differing context drives outcomes more than the quality of implementation. In other words, what are the central facilitator characteristics that drive change across the three programs and how can program stakeholders and designers manipulate future program iterations to better capture these characteristics to better improve the improve both programmatic and participant outcomes?

Limitations

The primary limitation of this study related to the low level of SMEs to develop the SJTs and correspondingly the relatively low number of facilitators to implement and test them. While outlier screening did not demonstrate that facilitators had an undue influence on study results, a larger sample may have produced greater variation in responses. The small sample of facilitators also was problematic for the development of level two variables, including the FCPC constructs. While reliability estimates were similar to prior studies utilizing the FCPC, the use of composites rather than latent variables could have compromised study findings; however, the low sample size precluded other approaches. The high levels of correlation among first order CSD factors could have suppressed relationships among variables; however, squared AVE scores indicated that the CSD first order factors were discriminant enough for later relationship testing. Finally. this study only utilized two data points to establish relationships, parent

177

reports of outcomes and program facilitators. Responses from program observations, supervisor evaluations, child-reported outcomes, and peers may indicate other results (Durlak, 2015). The challenge with implementation science is not merely what we understand, rather it is that there is so much to understand. This study, while narrow, uncovered over 120 variables that could influence a facilitator's implementation quality suggesting there are likely exponentially more at the broader levels and paradoxically illustrates the necessity of continuing the investigation to ensure that LER programs are provided with the highest degree of evidence.

Practitioner Implications and Future Directions

Within the context of LER programs, there is a consistent and necessary pressure to demonstrate value and maximize resource use. An implementation-focused LER organization can aid in both regards. Implementation assessment uncovers why programs succeed and fail, and in effect captures the "magic" of programs. As LER organizations trend towards implementation-focused assessment and development, they will be able to identify the factors that best promote positive outcomes for those they are charged with serving. More simply, implementation focused culture helps to uncover the essential ingredients of an intervention or program, and then ensures that a facilitator can provide them (or provides actionable data to why they would not). As highlighted in figure 10, there are necessary (e.g., semisweet chocolate pieces) and optional (e.g., pecans) ingredients to deliver a product. In the example provided in figure 10, this necessary and optional approach can highlight the resources needed to provide a program to achieve the desired outcome. Further, implementation assessment can highlight the importance of

delivering a program as designed, where if the oven is too warm, the cookies will burn, and if the butter is melted rather than softened, the cookies will go flat. Interestingly though, the only way to uncover which components of a recipe or program are necessary is to try it as designed. As such, this study highlights how practitioners charged with program design and delivery should assess implementation to determine what factors, components, and/or ingredients best contribute to program outcomes and implementation quality. Additionally, as noted in chapter three, some of the SJTs produced for this study did predict program outcomes and some facilitator behaviors. They may act as a training and/or screening tool when selecting staff for programs or training them for specific roles.

Researcher Implications and Future Directions

The current challenge with much of the research underpinning the critical importance of implementation and its components to the achievement of program outcomes is "this is an argument based primarily on conclusions drawn from the *absence*, rather than *presence*, of empirical evidence" (Moore et al., 2013, p. 149). In other words, much of implementation science research suggests that when programs are implemented as designed they tend to have better outcomes; however, this direction of research does not effectively capture the exceptions well. In what circumstances does the maintenance of implementation quality *hurt* outcomes? The findings within this study indicate that when facilitators are more experienced, they are less bought-in, although this effect did not translate to program outcomes.

CHOCOLATE CHIP COOKIES

PREP: 25 MINUTES BAKE: 8 MINUTES PER BATCH Oven 375°

½ cup shortening
½ cup butter or margarine
½ cup granulated sugar
1 cup packed brown sugar
½ teaspoon baking soda
2 eggs
1 teaspoon vanilla
2½ cups all-purpose flour
1 12-ounce package (2 cups) semisweet chocolate pieces
1½ cups chopped walnuts, pecans, or hazelnuts (filberts) (optional)

1. In a large mixing bowl beat the shortening and butter or margarine with an electric mixer on medium to high speed for 30 seconds. Add the granulated sugar, brown sugar, and baking soda. Beat mixture till combined, scraping sides of bowl occasionally. Beat in the eggs and vanilla till combined. Beat in as much of the flour as you can with the mixer. Stir in remaining flour (see photo 1, right). Stir in chocolate pieces and, if desired, nuts.
2. Drop dough by rounded teaspoons 2 inches apart on an ungreased cookie sheet (see photo 2, right). Bake in a 375° oven 8 to 10 minutes or till edges are lightly browned. Transfer cookies to a wire rack and let cool. Makes about 60 cookies.

Nutrition Facts per cookie: 93 cal., 5 g total fat (1 g sat. fat), 11 mg chol., 29 mg sodium, 12 g carbo., 0 g fiber, 1 g pro.
Daily Values: 1% vit. A, 0% vit. C, 0% calcium, 3% iron

Figure 10. Chocolate Chip Cookie Recipe (Better Homes and Gardens, 1996).

In many LER settings, satisfaction with the program or service is often measured for internal program and/organizational improvement. Additionally, within the residential summer camp industry the percentage of returning campers is often used to gauge the efficacy of camp processes and programs. These measures of satisfaction and percentage of returning campers also fall under a sub-dimension of implementation quality, participant responsiveness (Berkel et al., 2011; Ennett et al., 2011). While in many cases these measures may not have been intended as criteria to assess implementation quality, they do represent preliminary evidence of LER research to assess implementation quality, and potentially the option of retrospective assessments of some dimensions of implementation quality. In future studies, these measures should be considered as elements of implementation assessment.

This study builds upon the recommendation of Gagnon and Bumpus (2016) who suggested, "An important goal of future research in this domain would be to examine systematically the conditions under which general beliefs about the importance of fidelity are, or are not, predictive of fidelity to a specific program design" (p. 21). More

specifically, this study provided additional evidence regarding the usefulness of the

FCPC as a quality measure predicting implementation quality outside of college and

youth-at-risk settings (Gagnon et al., 2015a). However, the FCPC has yet to be

implemented outside of the control of the FCPC designers. Future investigations utilizing

the FCPC may produce alternative results. Additionally, some SJT responses were

associated with dimensions of the FCPC, indicating that the shared variance between the

two measures may in effect "cancel out." More specifically, future investigations should

examine the discriminant validity of the FCPC and SJT to address potential crossover

between the measures.

APPENDICES

Complete Facilitator Questionnaire

Thank you for completing the following survey. If you have any questions, please ask the survey administrator. Please complete all eight pages of this survey. All responses will be kept confidential and used for research purposes only.

1. What is your gender? **(Circle One)**

Male   Female   Non-Binary   Other **(Fill-In)**

_____

2. What is your age in years? **(Fill-In)** _____

3. What is your ethnic group **(Circle One)**

White   Asian Origin   Pacific Islander   Black, African American

Other **(Fill-In)**   East Asian   Multiple Race   Hispanic or Latino Origin

_____   (Indian)

4. How many years have you been in college? **(Fill-In)** _____

5. Do you already have a bachelor's degree? **(Circle One)**   No   Yes

6. What is your Last Name? **(Fill-In)** _____
(Fill-in the same as you did on the pre-tests).

7. In number of years please estimate your experience level in facilitating groups (a table is provided below to help you estimate your hours). Facilitation involves the planning, guiding, and management of a group to achieve a goal. This could include, but is not limited to: leading activities for children and/or youth, training staff, teaching fitness classes, leading group trips, etc …
(Fill-in Years Here) _____

1 week =	40 hours
1 month =	160 hours
1 year =	1920 hours

8. Please describe your level of skill facilitating groups from beginner to expert **(Circle One)**

Beginner 1   2   3   4   5   6   7 **Expert**

9. Which site are you working at **(Circle One)**

Sewee   Hannon Adventures   4-H

Wildlife   Voyager   Other (Please Fill-in)

_____

In this portion you will be presented with scenarios commonly encountered during the facilitation of programs. The scenario details may not be an exact fit to your own programs, but we'd still like to know how you would respond in the given scenario. **Read the scenario and then circle the action you would most likely take.**

Circle One	*You are delivering a 90-minute resilience program to a group of eight 13-15 year olds. Your co-facilitator is not adhering to the program plan and mentioned to you beforehand that she knew the teens would not be engaged. Her poor attitude towards the program is now causing the teens to disengage. Please select the strategy you believe would be the most effective to reengage your partner while maintaining program quality?*
1	Remove co-facilitator from program
2	Address co-facilitator's issue(s) in private while group participates in alternative activity
3	Take a group break and visit with co-facilitator in private
4	Introduce unrelated activity
5	It is not correctable, cancel program
6	Revisit & Refocus on program goals with co-facilitator
7	Move forward without addressing co-facilitator
8	Acknowledging and addressing co-facilitator issue in front of participants
9	It is not correctable, keep going.
10	Revisit & Refocus on program goals with entire group including co-facilitator.
11	Working with that particular co-facilitator

Circle One	*You and a co-facilitator are delivering a 4-hour program that promotes cooperation skills in youth-at-risk. After driving for 30 minutes to the site to deliver the program your co-facilitator realizes that he forgot the required program supplies at home. Driving back to get the supplies will cut the time you have to less than 3 hours. Please select what you believe will be the most effective strategy to achieve the program goals and maintain the program plan?*
1	Consider Alternative Plans
2	Proceed with non-equipment-required activities, during which have co-facilitator pick-up equipment
3	Delay program and return to pick up equipment
4	Relate problem to program goals of participants
5	Deliver alternative program
6	Deliver program without supplies
7	Leave participants at program site to pick up equipment
8	Take participants back to pick up equipment
9	Telling the group about forgotten items.

10	Panicking
11	Remove the co-facilitator
12	Cancel program

In this portion you will be presented with scenarios commonly encountered during the facilitation of programs. The scenario details may not be an exact fit to your own programs, but we'd still like to know how you would respond in the given scenario. **Read the scenario and then circle the action you would most likely take.**

Circle One	*You are delivering a program aimed at developing self-efficacy to a group of 8-9 year olds that has only previously been delivered to 14-16 year olds. The 8-9 year olds are having trouble focusing on the activities and are not engaging in the program. Please select what you believe will be the most effective strategy to achieve the program goals and maintain the program plan?*
1	Modify/Adjust program plan for age-appropriate behaviors
2	Modify activities to better fit group
3	Focus on learning objectives rather than plan
4	Continue without acknowledging issues
5	Hire an experienced facilitator
6	Shame the kids
7	Take a break

Circle One	*This morning your colleague called in sick and you were called in to deliver an 8-hour program that develops independence in 16-18 year olds. You recall practicing this program several months ago during staff training, but don't have any more resources than the program guide, which describes each step of the program in great detail. Please select what you believe will be the most effective strategy to achieve the program goals and maintain the program plan?*
1	Revisit program guide and plan
2	Check-in with prior facilitator
3	Cancel program
4	Deliver alternative program
5	Rely on program guide
6	Complain
7	Not utilize program guide, create own
8	Not utilize outside help
9	Be inflexible
10	Adhere strictly to program guide
11	Assess ability and attitude with delivering the program.
12	Rely on group feedback to improve engagement

13	Have a Bad attitude
14	Presenting program poorly, using circumstances as an excuse
15	Following guide step by step in real time
16	Not believing in self
17	Lower expectations of group

In this portion you will be presented with scenarios commonly encountered during the facilitation of programs. The scenario details may not be an exact fit to your own programs, but we'd still like to know how you would respond in the given scenario. **Read the scenario and then circle the action you would most likely take.**

Circle One	*You are delivering a five-hour program to develop leadership behaviors in a group of nine 12-14 year olds. About an hour into the program, several of the participant's mention that they "played the same games" last week in their afterschool program you are not affiliated with. You know that it is important to stick with the program plan as evidenced in your own training with program. Please select what you believe will be the most effective strategy to achieve the program goals and maintain the program plan?*
1	Focus on learning objectives and alternative activities
2	Modify current activities to make them more challenging
3	Utilize prior participant experience as part of facilitation, frame as positive
4	Introduce alternative activity strategies
5	Exclude participants with prior experience
6	Communicate to participants they "have" to do it anyway
7	Ignore prior experience
8	Use back up activities
9	Not having a backup plan

Circle One	*You are providing the second half of a 2 day 16-hour program aimed at improving teamwork quality in nine 11-13 year olds. As part of the program the youth developed a video to show what leadership meant to them. However, as soon as you moved to the video sharing component of the program, the electricity went out. This video was the central outcome of the program. Please select what you believe will be the most effective strategy to achieve the program goals and maintain the program plan?*
1	Determine if/when electricity will return
2	Have participants deliver product in alternative format
3	Provide filler activity until electricity is back

4	Wait for electricity
5	Have participants not share information that is one video
6	End program early
7	Apply situation (flexibility) to leadership and discuss with kids
8	Ensure videos will be seen regardless if it's at camp or on social media
9	Use back up options/plan for video sharing
10	Showing frustration and panic to the group

These next questions relate to you as a program facilitator, leader, or counselor, at the Youth Learning Institute (YLI). Circle the appropriate choice for each question.

	Strongly Disagree	Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Agree	Strongly Agree
I consider myself experienced in facilitating groups	1	2	3	4	5	6	7
I believe in the goals of the Youth Learning Institute (YLI)	1	2	3	4	5	6	7
I follow pre-designed program plans	1	2	3	4	5	6	7
I have enough training to facilitate programs for YLI	1	2	3	4	5	6	7
I trust in the YLI	1	2	3	4	5	6	7
I am "bought-in" to YLI	1	2	3	4	5	6	7
I am prepared to facilitate YLI programs due to my general facilitating experience	1	2	3	4	5	6	7
I feel well trained to facilitate at YLI	1	2	3	4	5	6	7
Deviating from the program plan allows facilitators to meet program goals	1	2	3	4	5	6	7
A program plan limits my ability to facilitate	1	2	3	4	5	6	7
The ability to change a program plan is important to achieve quality outcomes	1	2	3	4	5	6	7
The training I have received has prepared me to facilitate at YLI	1	2	3	4	5	6	7

I would recommend YLI to other groups	**1**	**2**	**3**	**4**	**5**	**6**	**7**
It is important to deliver YLI programs as designed	**1**	**2**	**3**	**4**	**5**	**6**	**7**
I have a high level of experience facilitating groups	**1**	**2**	**3**	**4**	**5**	**6**	**7**

These questions relate to your skill as a counselor or facilitator. Circle the appropriate choice for each question.

	Not like me at all	Not much like me	Somewhat like me	Mostly like me	Very much like me
I have overcome setbacks to conquer an important challenge	**1**	**2**	**3**	**4**	**5**
New ideas and projects sometimes distract me from previous ones	**1**	**2**	**3**	**4**	**5**
My interests change from year to year	**1**	**2**	**3**	**4**	**5**
Setbacks don't discourage me	**1**	**2**	**3**	**4**	**5**
I have been obsessed with a certain idea or project for a short time, but later lost interest	**1**	**2**	**3**	**4**	**5**
I am a hard worker	**1**	**2**	**3**	**4**	**5**
I often set a goal, but later choose to pursue a different one	**1**	**2**	**3**	**4**	**5**
I have difficulty maintaining focus on projects that take more than a few months to complete	**1**	**2**	**3**	**4**	**5**
I finish whatever I begin	**1**	**2**	**3**	**4**	**5**
I have achieved a goal that took years of work to complete	**1**	**2**	**3**	**4**	**5**
I have become interested in new pursuits every few months	**1**	**2**	**3**	**4**	**5**

I am diligent		1	2	3	4	5

Below is a list of worries that YLI counselors may encounter. Please rate your level of worry or concern from 1 (Not concerned or worried at all) to 5 (Extremely concerned or worried) by circling the appropriate choice

I am worried about…	Not at all	Little	Somewhat	A Great Deal	Completely
…working with kids	1	2	3	4	5
…managing child homesickness	1	2	3	4	5
…managing discipline of children	1	2	3	4	5
…housing with children	1	2	3	4	5
…managing children with emotional instability	1	2	3	4	5
…managing with sensitive issues	1	2	3	4	5
Select Three here	1	2	3	4	5
…facilitating intimidating children	1	2	3	4	5
…managing children very different than me	1	2	3	4	5
…interacting with parents	1	2	3	4	5
…becoming "burned out"	1	2	3	4	5
…working in an outdoor setting	1	2	3	4	5
…being away from my social support system	1	2	3	4	5

There are many reasons why camp may be good for youth. Please rate the choices below on how camp may benefit youth.

Camp is good for youth because…	Not at all	Little	Somewhat	A Great Deal	Completely
…it allows youth to be challenged	1	2	3	4	5
…it helps develop communication skills	1	2	3	4	5
…they cooperate better with their peers	1	2	3	4	5
…it grows their self-confidence	1	2	3	4	5

…it broadens their social circles	1	2	3	4	5
…it helps them meet people from different cultures	1	2	3	4	5
…it helps them establish their own identity	1	2	3	4	5
Please select three here	1	2	3	4	5
…it helps them develop personal responsibility	1	2	3	4	5
…it helps them to learn technical skills like swimming	1	2	3	4	5
…it helps to develop independence	1	2	3	4	5
…it exposes them to nature	1	2	3	4	5
…it provides time away from electronics	1	2	3	4	5
…it helps them overcome adversity	1	2	3	4	5
…it helps them to manage stress	1	2	3	4	5
…it helps them to take initiative	1	2	3	4	5

Below is a list of outcomes many YLI counselors expect to achieve as a result of their training. Please rate the outcomes you learned a result of training.

I learned…	Not at all	Little	Somewhat	A Great Deal	Completely
…how to facilitate experiences for youth	1	2	3	4	5
…how to properly supervise youth	1	2	3	4	5
…how to keep youth safe	1	2	3	4	5
…how to provide customer service	1	2	3	4	5
…how to help youth have fun	1	2	3	4	5
…the camp's policies and procedures	1	2	3	4	5
…how to work effectively with fellow staff members	1	2	3	4	5
…how to manage youth behavior	1	2	3	4	5
…how to implement camp policies and procedures	1	2	3	4	5

How do you expect to grow as a YLI counselor this summer? Below please rate the areas where you will grow.

The areas I grew in are	Not at all	Little	Somewhat	A Great Deal	Completely
…learning to work in a different setting	1	2	3	4	5
…learning a new technical skill like lifeguarding	1	2	3	4	5
…gaining experience working with children	1	2	3	4	5
…learning how to interact with a diverse group	1	2	3	4	5
…improving my work ethic	1	2	3	4	5
…gaining experience in goal setting	1	2	3	4	5
…becoming more independent	1	2	3	4	5
…enhancing my enjoyment of the outdoors	1	2	3	4	5
…gaining practical experience for a future career	1	2	3	4	5
…learning to take initiative	1	2	3	4	5
…developing personal responsibility	1	2	3	4	5
…learning to start tasks with little direction	1	2	3	4	5

Appendix B

Facilitator Pre-Training Questionnaire (non-online version)

Thank you for completing the following survey. If you have any questions, please ask the survey administrator. Please complete all pages of this survey. All responses will be kept confidential and used for research purposes only.

What is your gender? **(Circle One)**

Male   Female   Non-Binary   Other **(Fill-In)**

_____

What is your age in years? **(Fill-In)** _____

What is your ethnic group? **(Circle One)**

Caucasian   Asian Origin   Pacific Islander   African American
Other **(Fill-In)**   East Asian (Indian)   Multiple Race   Hispanic or Latino
_____   Origin

How many years have you been in college? **(Fill-In)** _____ Circle Here if Not Applicable

Do you already have a bachelor's degree? **(Circle One)**   No   Yes

What is your Last Name? **(Fill-In)** _____
(Fill-in the same as you did on the prior questionnaires).
At what camp are you working? **(Circle One)**

Sewee   4-H   Hannon   Wildlife   Voyager
                                                            Adventures

In **number of weeks** please estimate your experience level in facilitating groups (a table is provided below to help you estimate your hours). Facilitation involves the planning, guiding, and management of a group to achieve a goal. This could include, but is not limited to: leading activities for children and/or youth, training staff, teaching fitness classes, leading group trips, etc…
(Fill-in Weeks Here) _____

1 week =	40 hours
1 month =	160 hours
1 year =	1920 hours

Please describe your level of skill facilitating groups from beginner to expert **(Circle One)**

Beginner 1   2   3   4   5   6   7 **Expert**

On the following page you will be presented with 6 scenarios that describe situations you are likely to encounter at some point as a YLI counselor. Carefully read the scenario, and then select the response based on how you would likely respond.

Scenario 1
You are delivering a 90-minute outdoor resilience program to a group of eight 13-15 year olds. Your co-facilitator is not adhering to the program plan and mentioned to you beforehand that he/she knew the teens would not be engaged. Her poor attitude towards the program is now causing the teens to disengage. Please respond with the strategy you believe would be the most effective to reengage your partner while maintaining program quality?

Scenario 2
You and a co-facilitator are delivering a 4-hour program that promotes cooperation skills in youth-at-risk. After driving for 30 minutes to the site to deliver the program your co-facilitator realizes that he forgot the required program supplies at home. Driving back to get the supplies will cut the time you have to less than 3 hours. Please respond with what you believe will be the most effective strategy to achieve the program goals and maintain the program plan?

Scenario 3
You are delivering a program aimed at developing self-efficacy to a group of 8-9 year olds that has only previously been delivered to 14-16 year olds. The 8-9 year olds are having trouble focusing on the activities and are not engaging in the program. Please respond with what you believe will be the most effective strategy to achieve the program goals and maintain the program plan?

Scenario 4
This morning your colleague called in sick and you were called in to deliver an 8-hour program that develops independence in 10-12 year olds. You recall practicing this program several months ago during staff training, but don't have any more resources than the program guide, which describes each step of the program in great detail. Please respond with what you believe will be the most effective strategy to achieve the program goals and maintain the program plan?

Scenario 5
You are delivering a five-hour program to develop self-regulative behaviors in a group of nine 12-14 year olds. About an hour into the program, several of the participant's mention that they "played the same games" last week in their afterschool program you are not affiliated with. You know that it is important to stick with the program plan as

evidenced in your own training with program. Please respond with what you believe will be the most effective strategy to achieve the program goals and maintain the program plan?

Scenario 6
You are providing the second half of a 2 day 16-hour program aimed at improving leadership quality in nine 11-13 year olds. As part of the program the youth developed a video to show what leadership meant to them. However, as soon as you moved to the video sharing component of the program, the electricity went out. This video was the central outcome of the program. Please respond with what you believe will be the most effective strategy to achieve the program goals and maintain the program plan?

These next questions relate to you as a program facilitator, leader, or counselor, at the Youth Learning Institute (YLI).

Circle the appropriate choice for each question.

	Strongly Disagree	Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Agree	Strongly Agree
I consider myself experienced in facilitating groups	1	2	3	4	5	6	7
I believe in the goals of the Youth Learning Institute (YLI)	1	2	3	4	5	6	7
I follow pre-designed program plans	1	2	3	4	5	6	7
I have enough training to facilitate programs for YLI	1	2	3	4	5	6	7
I trust in the YLI	1	2	3	4	5	6	7
I am "bought-in" to YLI	1	2	3	4	5	6	7
I am prepared to facilitate YLI programs due to my general facilitating experience	1	2	3	4	5	6	7
I feel well trained to facilitate at YLI	1	2	3	4	5	6	7
Deviating from the program plan allows facilitators to meet program goals	1	2	3	4	5	6	7
A program plan limits my ability to facilitate	1	2	3	4	5	6	7
The ability to change a program plan is important to achieve quality outcomes	1	2	3	4	5	6	7

The training I have received has prepared me to facilitate at YLI	**1**	**2**	**3**	**4**	**5**	**6**	**7**
I would recommend YLI to other groups	**1**	**2**	**3**	**4**	**5**	**6**	**7**
It is important to deliver YLI programs as designed	**1**	**2**	**3**	**4**	**5**	**6**	**7**
I have a high level of experience facilitating groups	**1**	**2**	**3**	**4**	**5**	**6**	**7**

These questions relate to your skill as a counselor or facilitator. Circle the appropriate choice for each question.

	Not like me at all	Not much like me	Somewhat like me	Mostly like me	Very much like me
I have overcome setbacks to conquer an important challenge	**1**	**2**	**3**	**4**	**5**
New ideas and projects sometimes distract me from previous ones	**1**	**2**	**3**	**4**	**5**
My interests change from year to year	**1**	**2**	**3**	**4**	**5**
Setbacks don't discourage me	**1**	**2**	**3**	**4**	**5**
I have been obsessed with a certain idea or project for a short time, but later lost interest	**1**	**2**	**3**	**4**	**5**
I am a hard worker	**1**	**2**	**3**	**4**	**5**
I often set a goal, but later choose to pursue a different one	**1**	**2**	**3**	**4**	**5**
I have difficulty maintaining focus on projects that take more than a few months to complete	**1**	**2**	**3**	**4**	**5**
I finish whatever I begin	**1**	**2**	**3**	**4**	**5**
I have achieved a goal that took years of work to complete	**1**	**2**	**3**	**4**	**5**
I have become interested in new pursuits every few months	**1**	**2**	**3**	**4**	**5**
I am diligent	**1**	**2**	**3**	**4**	**5**

Below is a list of worries that YLI counselors may encounter. Please rate your level of worry or concern from 1 (Not concerned or worried at all) to 5 (Extremely concerned or worried)

I am worried about…	Not at all	Little	Somewhat	A Great Deal	Completely
…working with kids	**1**	**2**	**3**	**4**	**5**
…managing child homesickness	**1**	**2**	**3**	**4**	**5**
...managing discipline of children	**1**	**2**	**3**	**4**	**5**
…housing with children	**1**	**2**	**3**	**4**	**5**
…managing children with emotional instability	**1**	**2**	**3**	**4**	**5**
…managing with sensitive issues	**1**	**2**	**3**	**4**	**5**
…facilitating intimidating children	**1**	**2**	**3**	**4**	**5**
…managing children very different than me	**1**	**2**	**3**	**4**	**5**

…interacting with parents	1	2	3	4	5
…becoming "burned out"	1	2	3	4	5
…working in an outdoor setting	1	2	3	4	5
…being away from my social support system	1	2	3	4	5

Below is a list of reasons people choose to become YLI counselors. Please rate the reasons you may have selected to work for the Youth Learning Institute

I chose to work at YLI because…	Not at all	Little	Somewhat	A Great Deal	Completely
Practical experience related to my career interest/college major	1	2	3	4	5
I wanted to work with youth	1	2	3	4	5
I grew up going to camp	1	2	3	4	5
I want to become a camp director as my career	1	2	3	4	5
The YLI camp is close to my home	1	2	3	4	5
The YLI camp is located where I want to live	1	2	3	4	5
It meets an internship requirement	1	2	3	4	5

Below is a list of outcomes many YLI counselors expect to achieve as a result of their training. Please rate the outcomes you expect as a result of training.

I expect to learn…	Not at all	Little	Somewhat	A Great Deal	Completely
…how to facilitate experiences for youth	1	2	3	4	5
…how to properly supervise youth	1	2	3	4	5
…how to keep youth safe	1	2	3	4	5
…how to provide customer service	1	2	3	4	5
…how to help youth have fun	1	2	3	4	5
…the camp's policies and procedures	1	2	3	4	5
…how to work effectively with fellow staff members	1	2	3	4	5
…how to manage youth behavior	1	2	3	4	5
…how to implement camp policies and procedures	1	2	3	4	5

Camp is good for youth because…	Not at all	Little	Somewhat	A Great Deal	Completely
…it allows youth to be challenged	1	2	3	4	5
…it helps develop communication skills	1	2	3	4	5
…they cooperate better with their peers	1	2	3	4	5
…it grows their self-confidence	1	2	3	4	5
…it broadens their social circles	1	2	3	4	5
…it helps them meet people from different cultures	1	2	3	4	5
…it helps them establish their own identity	1	2	3	4	5
…it helps them develop personal responsibility	1	2	3	4	5
…it helps them to learn technical skills like swimming	1	2	3	4	5
…it helps to develop independence	1	2	3	4	5
…it exposes them to nature	1	2	3	4	5
…it provides time away from electronics	1	2	3	4	5
…it helps them overcome adversity	1	2	3	4	5

…it helps them to manage stress	1	2	3	4	5
…it helps them to take initiative	1	2	3	4	5

How do you expect to grow as a YLI counselor this summer? Below please rate the areas where you will grow.

The areas I expect to grow in are	Not at all	Little	Somewhat	A Great Deal	Complet ely
…learning to work in a different setting	1	2	3	4	5
…learning a new technical skill like lifeguarding	1	2	3	4	5
…gaining experience working with children	1	2	3	4	5
…learning how to interact with a diverse group	1	2	3	4	5
…improving my work ethic	1	2	3	4	5
…gaining experience in goal setting	1	2	3	4	5
…becoming more independent	1	2	3	4	5
…enhancing my enjoyment of the outdoors	1	2	3	4	5
…gaining practical experience for a future career	1	2	3	4	5
…learning to take initiative	1	2	3	4	5
…developing personal responsibility	1	2	3	4	5
…learning to start tasks with little direction	1	2	3	4	5

Parent Questions (non-online version)

Descriptives
P1   How did you hear about this survey?
P2   Which camp did your child most recently attend?
P3   Approximately how many weeks was your child's camp session for the summer of 2016?
P4   Approximately what week did your child start camp?
P4A   What is your child's last name
P5   How many of your children will attend (or have attended) camp this year?
P6   How many years has your child attended this camp?
P7   What is your child's age in years?
P8   What is your child's gender?
P9   What ethnic and/or racial group is your child?
P10   What grade has your child most recently completed?
P11   What type of school does your child attend?
P12   Have you already completed this survey for another child?
P13   What is your gender?
P14   What is your ethnic/racial group?
P15   What is your annual household income in dollars per year?
P16   What is your current relationship status?
P17   What is your current education level?
P18   How many years did you attend camp as a child?
P19   If you attended camp as a child which of the following style of camp did you attend?
P20   What is your zip code?
Ways You Learned About Camp
P21   How did you learn about the overnight camp your child is attended?
Reasons for Sending Your Child to Camp
P22   Which of the following best describes your PRIMARY reason for sending your child to camp?

Parental Satisfaction with Camp
P23   Please rate your satisfaction with the following elements of the camp program.-Camp marketing
P24   Please rate your satisfaction with the following elements of the camp program.-Customer service
P25   Please rate your satisfaction with the following elements of the camp program.-Camp fee (value)
P26   Please rate your satisfaction with the following elements of the camp program.-Camp website

P27   Please rate your satisfaction with the following elements of the camp program.-Camp staff

P28   Please rate your satisfaction with the following elements of the camp program.-Camp registration process and paperwork

P29   Please rate your satisfaction with the following elements of the camp program.-Administrative personnel

P30   Please rate your satisfaction with the following elements of the camp program.-Check-in process

P31   Please rate your satisfaction with the following elements of the camp program.-Check-out process

P32   Please rate your satisfaction with the following elements of the camp program.-Your child's overall camp experience

P33   Please rate your satisfaction with the following elements of the camp program.-Camp buildings and grounds

Parental Observations of Developmental Outcomes of Camp Experiences - RPT

Format with "Before" questions asked prior to "After" questions

P34   My child... : Before Camp-…takes responsibility for his / her own actions.

P35   My child... : Before Camp-…takes care of his / her own things.

P36   My child... : Before Camp-…takes initiative / is a self-starter.

   P37   My child... : After Camp-…takes responsibility for his / her own actions.

   P38   My child... : After Camp-…takes care of his / her own things.

   P39   My child... : After Camp-…takes initiative / is a self-starter.

P40   My child... : Before Camp-…participates in new learning experiences.

P41   My child... : Before Camp-...is curious about new topics and subjects.

P42   My child... : Before Camp-…seeks challenges beyond his / her comfort zone.

P43   My child... : Before Camp-…is willing to try new experiences.

   P44   My child... : After Camp-…participates in new learning experiences.

   P45   My child... : After Camp-...is curious about new topics and subjects.

   P46   My child... : After Camp-…seeks challenges beyond his / her comfort zone.

   P47   My child... : After Camp-…is willing to try new experiences.

P48   My child... : Before Camp-…handles success and failure.

P49   My child... : Before Camp-…adapts to change.

P50   My child... : Before Camp-…manages disappointment well.

P51   My child... : Before Camp-…deals effectively with conflict.

   P52   My child... : After Camp-…handles success and failure.

   P53   My child... : After Camp-…adapts to change.

   P54   My child... : After Camp-…manages disappointment well.

   P55   My child... : After Camp-…deals effectively with conflict.

P56   My child... : Before Camp-...is a team player.

P57   My child... : Before Camp-…shares work responsibilities.

P58   My child... : Before Camp-…cooperates and works well in a group.

P59   My child... : Before Camp-…follows through when asked to do something.

P60   My child... : Before Camp-…follows directions.

P61   My child... : After Camp-…is a team player.
P62   My child... : After Camp-…shares work responsibilities.
P63   My child... : After Camp-…cooperates and works well in a group.
P64   My child... : After Camp-…follows through when asked to do something.
P65   My child... : After Camp-…follows directions.

P66   My child... : Before Camp-…has a good mental attitude.
P67   My child... : Before Camp-...has a generally "positive" view on life.
P68   My child... : Before Camp-…is more hopeful about the future.
P69   My child... : Before Camp-…shows a positive attitude when around others.
P70   My child... : After Camp-…has a good mental attitude.
P71   My child... : After Camp-...has a generally "positive" view on life.
P72   My child... : After Camp-…is more hopeful about the future.
P73   My child... : After Camp-…shows a positive attitude when around others.
P74   My child... : Before Camp-…participates in a discussion.
P75   My child... : Before Camp-…communicates well with others.
P76   My child... : Before Camp-…shares thoughts and ideas verbally.
P77   My child... : After Camp-…participates in a discussion.
P78   My child... : After Camp-…communicates well with others.
P79   My child... : After Camp-…shares thoughts and ideas verbally.
P80   My child... : Before Camp-…listens to the opinions of others.
P81   My child... : Before Camp-…gets along with people around him/her.
P82   My child... : Before Camp-…is aware of other's needs in social situations
P83   My child... : After Camp-…listens to the opinions of others.
P84   My child... : After Camp-…gets along with people around him/her.
P85   My child... : After Camp-…is aware of other's needs in social situations
P86   My child... : Before Camp-…tries to find solutions to problems.
P87   My child... : Before Camp-…tries to find answers to questions.
P88   My child... : Before Camp-…asks questions.
P89   My child... : After Camp-…tries to find solutions to problems.
P90   My child... : After Camp-…tries to find answers to questions.
P91   My child... : After Camp-…asks questions.

P92   My child... : Before Camp-…makes good decisions.
P93   My child... : Before Camp-…considers choices before making a decision.
P94   My child... : Before Camp-…sets priorities.
P95   My child... : Before Camp-…sets goals for himself / herself.
P96   My child... : After Camp-…makes good decisions.
P97   My child... : After Camp-…considers choices before making a decision.
P98   My child... : After Camp-…sets priorities.
P99   My child... : After Camp-…sets goals for himself / herself.
P100   My child... : Before Camp-…manages his/her emotions.
P101   My child... : Before Camp-…doesn't get frustrated easily.
P102   My child... : After Camp-…manages his/her emotions.

P103   My child... : After Camp-…doesn't get frustrated easily.

Parental Anxiety Associated with Outdoor Experiences (subscale of the PAOEO)

P104   I told my child to stay away from bullies at camp.

P105   I was worried about being away from my child while he/she was at camp.

P106   I talked with my child about things that worried me about camp.

P107   I am afraid that my child will get lost outside in nature.

P108   I am afraid of wild animals or insects outside in nature.

P109   I am afraid of my child getting hurt if he/she plays outside in nature.

P110   I am concerned when my child gets dirty after playing outdoors

P111   I am afraid my child may be harmed by strangers outside.

P112   I worry that my child will be hurt by bullies if he/she plays outside.

P113   I worry about my child getting too much sun.

P114   I am afraid my child may be abducted when spending time outside.

Overparenting (subscale of the PAOEO)

P115   I make important decisions for my child.

P116   I intervene in settling disputes with my child's classmates or friends.

P117   I intervene in settling disputes with my child's teacher, coach, or youth program leader.

P118   I have told my child that he/she needs my support to succeed in life.

P119   I regularly call or text my child to check in with them.

P120   I try to protect my child from negative influences.

P121   If something doesn't work out for my child, I do what I can to fix it.

P122   when something goes wrong in my child's life, I jump in to take care of it.

P123   when my child is engaged in an important task or project, I do some of it for them.

P124   I manage most important decisions in my child's life.

P125   I solve any crisis or problem my child might have.

References

Aarons, G. A., Glisson, C., Hoagwood, K., Kelleher, K., Landsverk, J., & Cafri, G. (2010). Psychometric properties and U.S. national norms of the Evidence-Based Practice Attitude Scale (EBPAS). *Psychological Assessment, 22*(2), 356-365.

Arai, S., & Pedlar, A. (2003). Moving beyond individualism in leisure theory: A critical analysis of community and social engagement. *Leisure Studies, 22*, 185-202.

Baker, C. N., Kupersmidt, J. B., Voegler-Lee, M. E., Arnold, D. H., & Willoughby, M. T. (2010). Predicting teacher participation in a classroom-based, integrated preventive intervention for preschoolers. *Early Childhood Research Quarterly, 25*, 270-283.

Barrera, M., Berkel, C., & Gonzalez-Castro, F. (2016). Directions for the advancement of culturally adapted preventive interventions: Local adaptations, engagement, and sustainability. *Prevention Science, Early Online Copy*, 1-9.

Barrett, G. V., Doverspike, D., & Young, C. M. (2010). The special case of public sector police and fire selection. In J.C. Scott & D.H. Reynolds (Eds.), *Handbook of workplace assessment: Evidence-based practices for selecting and developing organizational talent* (pp. 437-462). San Francisco, CA: Jossey-Bass.

Bechar, S., & Mero-Jaffe, I. (2014). Who is afraid of evaluation? Ethics in evaluation research as a way to cope excessive evaluation anxiety: Insights from a case study. *American Journal of Evaluation, 35*(3), 364-376.

Berk, M., & McGivern-Moon, L. (2016). Effects of a facilitated fee waiver program on participation in youth sports programs. *Journal of Park and Recreation Administration, 34*(3), 99-105.

Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011). Putting the pieces together: An integrated model of program implementation. *Prevention Science, 12*, 23-33.

Better Homes and Gardens. (1996). Cookies. In J. D. Darling (Ed.) (11th ed.), *Better Homes and Gardens: New Cook Book* (180-202). Des Moines, IA: Meredith Corporation.

Bishop, D. C., Pankratz, M., M Hansen, W. B., Albritton, J., Albritton, L., & Strack, J. (2013). Measuring fidelity and adaptation: Reliability of a instrument for school-based prevention programs. *Evaluation and the Health Professions*, 1-27.

Borrelli, B., Sepinwall, D., Bellg, A. J., Breger, R., DeFrancesco, C., Sharp, D. L., …Orwig, D. (2005). A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *Journal of Consulting and Clinical Psychology, 73*(5), 852-860.

Bronfenbrenner, U. (1994). Ecological models of human development. In M. Gauvain, & M. Cole. (Eds.)., *Readings on the development of children* (2nd ed.). New York, NY: Freeman.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York,   NY: The Guilford Press.

Browne, L. P., Garst, B. A., & Bialeschki, M. D. (2011). Engaging youth in

    environmental sustainability: Impact of the Camp 2 Grow program. *Journal of*

    Park and Recreation   Administration, 29(3), 70-85.

Bryant, A., & Charmaz, K. (2007). *The SAGE handbook of grounded theory.* Thousand

    Oaks,   CA: SAGE Publications.

Bruening, J. E., Clark, B. S., & Mudrick, M. (2015). Sport-based youth development in

    practice: The long-term impacts of an urban after-school program for girls.

    Journal of Park and   Recreation Administration, 33(2), 87-103.

Byrne, B. (2006). *Structural equation modeling with EQS.* (2nd ed.). New York, NY:

    Psychology Press.

Caldwell, L. L., Younker, A. S., Wegner, L. Patrick, M. E., Vergnani, T., Smith, E. A., &

    Flisher, A. J. (2008). Understanding leisure-related program effects by using

    process data in the HealthWise South Africa project. *Journal of Park &*

    Recreation Administration,   26(2), 146–162.

Camire, M., Forneris, T., Trudel, P., & Bernard, D. (2011). Strategies for helping coaches

    facilitate positive youth development through sport. *Journal of Sport Psychology*

    in Action, 2, 92-99.

Cantu, A. M., Hill, L. G., Becker, L. G. (2010). Implementation quality of a family-

    focused preventive intervention in a community-based dissemination. *Journal of*

    Children's Services, 5(4), 18-30.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A

    conceptual framework for implementation fidelity. *Implementation Science, 2*,

    Article No. 40.

Castro, F. G., Barrera, M., & Martinez, C. R., Jr. (2004). The cultural adaptation of

    prevention interventions: Resolving tensions between fidelity and fit. *Prevention*

    Science, 5(1), 41-45

CDC. (2011). *Program evaluation tip sheet: Reach and Impact*. National Center for

    Disease Prevention and Health Promotion, Division for Heart Disease and Stroke

    Prevention. Retrieved from:

    https://www.cdc.gov/dhdsp/programs/spha/docs/reach_impact_tip_sheet.pdf

Century, J., & Cassata, A. (2016). Implementation research: Finding common ground

    on what, how, why, where, and who. *Review of Research in Education, 40*(1),

    169-215.

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of

    implementation: A foundation for shared language and accumulation of

    knowledge. *American Journal of Evaluation, 31*(2), 199-218.

Chan, D., & Schmitt, N. (2002). *Situational judgement and job performance*. Human

    Performance, 15(3), 233-254.

Cho, H., Lee, H-W., Moore, D., Norman, W. C., & Ramshaw, G. (2017). A multilevel

    approach to scale development in sport tourist nostalgia. *Journal of Travel*

    Research, 1-13.

Crompton, J. L. (2016). Using the benefits continuum, enterprise funds, and cost finding to implement best pricing practices. *Journal of Park and Recreation Administration, 34*(4), 82-101.

Curry, L., & Nunez-Smith, M. (2015). *Mixed methods in health sciences research: A practical primer.* Thousand Oaks, CA: SAGE Publications.

Cyr, L. F. (2008). Facilitation competence: A catalyst for effective Extension work. *Journal of Extension, 46*(4), Article 4RIB2. Retrieved from http://www.joe.org/joe/2008august/rb2.php.

Dane, A., & Schneider, B. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*(1), 23-45.

Dariotis, J. K., Bumbarger, B. K., Duncan, L. G., & Greenberg, M. T. (2008). How do implementation efforts relate to program adherence? Examining the role of organizational, implementer, and program factors. *Journal of Community Psychology, 36*(6), 744-760.

Desimone, L. M., & Lee-Hill, K. (2017). Inside the black box: Examining mediators and moderators of a middle school research intervention. *Educational Evaluation and Policy Analysis, Early Online Publication*, 1-26.

Dolcini, M., M., Catania, J., A., Gandelman, A., Ozer, E. M. (2014). Implementation a brief evidence-based HIV intervention: A mixed methods examination of compliance fidelity. *Translational Behavioral Medicine: Practice, Policy, and Research, 4*(4), 424-433.

Domitrovich, C. E., Gest, S. D., Gill, S., Jones, D., Sanford-DeRousie, R. (2009).

    Individual factors associated with professional development training outcomes of

    the head start REDI program. *Early Education and Development, 20*(3), 402-430.

Domitrovich, C. E., & Greenberg, M. T. (2000). The study of implementation: Current

    findings from effective programs that prevent mental disorders in school-aged

    children. *Journal of   educational and psychological consultation, 11*(2), 193-

    221.

Domitrovich, C. E., Pas, E. T., Bradshaw, C. P., Becker, K. D., Keperling, J. P., Embry,

    D., D., Ialongo, N. (2015). Individual and school organizational factors that

    influence implementation of the PAX Good Behavior Game intervention.

    Prevention Science, 16, 1064-1074.

Drasgow, F., Nye, C. D., & Tay, L. (2010). Indicators of quality assessment. In J.C. Scott

    & D.H. Reynolds (Eds.), *Handbook of workplace assessment: Evidence-based*

    practices for   selecting and developing organizational talent (pp. 27-59). San

    Francisco, CA: Jossey-Bass.

Duerden, M. D., & Witt, P. A. (2012). Assessing program implementation: What it is,

    why it's important, and how to do it. *Journal of Extension, 50*(1), Article 1FEA4.

    Retrieved from http://www.joe.org/joe/2012february/a4.php

Dufrene, B. A., Noell, G. H., Gilbertson, D. N., & Duhon, G. J. (2005). Monitoring

    implementation of reciprocal peer tutoring: Identifying and intervening with

    students who do not maintain accurate implementation. *School Psychology*

    Review, 34(1), 74-86.

Durlak, J. A. (1998). Why program implementation is important. *Journal of Prevention and Intervention in the Community, 17*(2), 5-18.

Durlak, J. A. (2010). The importance of doing well in whatever you do: A commentary on the special section, "Implementation research in early childhood education." *Early Childhood Research Quarterly, 25*(3), 348-357.

Durlak, J. A. (2015). Studying program implementation is not easy, but it is essential. *Prevention Science, 16*(8), 1123-1127.

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

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Research*, *18*(2), 237-256.

Dusenbury, L., Brannigan, R., Hansen, W., Walsh, J., & Falco, M. (2005). Quality of implementation: Developing measures crucial to understanding the diffusion of preventive interventions. *Health Education Research, 20*(3), 308-313.

Eames, C., Daley, D., Hutchings, J., Hughes, J. C., Jones, K., Martin, P., & Bywater, T. (2007). The leader observation tool: A process skills treatment measure for the Incredible Years parenting program. *Child: Care, Health, and Development, 34*(3), 391-400.

Eccles, M. P., & Mittman, B. S. (2006). Welcome to implementation science. *Implementation Science*, 1(1).

Elder, J. P., Lytle, L., Sallis, J. F., Rohm-Young, D., Steckler, A., Simons-Morton, D., Stone., E.,…Ribisl, K. (2007). A description of the social-ecological framework used in the trial of activity for adolescent girls (TAAG). *Health Education Research, 22*(2), 155-165.

Elliott, D. S., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science, 5*, 47-53.

Ennett, S. T., Haws, S., Ringwalt, C. L., Vincus, A. A., Hanley, S., Bowling, J. M., & Rohrbach, L. A. (2011). Evidence-based practice in school substance use prevention: Fidelity of implementation under real-world conditions. *Health Education Research, 26*(2), 361-371.

Ewert, A. (1983). *Outdoor adventure and self-concept: A research analysis.* Eugene, OR: Center for Leisure Studies, University of Oregon.

Fagan, A. A., Hanson, K., Hawkins, J. D., & Arthur, M. W. (2008). Bridging science to practice: Achieving prevention program implementation fidelity in the community youth development study. *American Journal of Community Psychology, 41*, 235-249.

Federal Reserve Bank of St. Louis. (2016). *Real median income in the United States* (Report No. MEHOINUSA672N). St. Louis, MO: Federal Reserve Bank of St. Louis.

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). Thousand Oaks, CA: SAGE Publications.

Fixsen, D., Blasé, K., Metz, A., & Van Dyke, M. (2013). Statewide implementation of

    evidence-based programs. *Exceptional Children, 79*(2), 213-230.

Fixsen, D. L., Blasé, K. A., Naoom, S. F., & Wallace, F. (2009). Core implementation

    components. *Research on Social Work Practice, 19*(5), 531–540.

Floyd, R. G., Phaneuf, R. L, & Wilczynski, S. M. (2005). Measurement properties of

    indirect assessment methods for functional behavioral assessment: A review of

    research. *School Psychology Review, 34*(1), 58-73.

Fox, S., Spector, P. E., Goh, A., & Bruursema, K. (2007). Does your coworker know

    what you're doing? Convergence of self- and peer-reports of counterproductive

    work behavior. *International Journal of Stress Management, 14*(1), 41-60.

Frantz, I., Stemmler, M. Hahlweg, K., Pluck, J., & Heinrichs, N. (2015). Experiences in

    disseminating evidence-based prevention programs in a real-world setting.

    Prevention Science, 16, 789-800.

Fredericksen, P., & London, R. (2000). Disconnect in the hollow state: The pivotal role

    of organizational capacity in community-based development organizations. *Public

    Administration Review, 60*(3), 230–239.

French, B. F., & Finch, W. H. (2006). Confirmatory factor analytic procedures for the

    determination of measurement invariance. *Structural Equation Modeling: A

    Multi- Disciplinary Journal, 13*(3), 378-402.

Gagnon, R. J. (2014). Exploring the relationship between the facilitator and fidelity.

    Journal of Outdoor Recreation, Education, and Leadership, 6(2), 183-186.

Gagnon, R. J., & Bumpus, M. F. (2016). Fidelity and its importance to experiential and outdoor education. *Journal of Outdoor Recreation, Education, and Leadership 8*(1), 10-25.

Gagnon, R. J., Garst, B. A., & Stone, G. A. (2015a). *Measuring self-reported fidelity in recreation: Exploring the effectiveness of the facilitator characteristics and program contributions scale.* Proceedings from the Northeastern Recreation Research Symposium 2015. Annapolis, MD: NERR.

Gagnon, R. J., Franz, N., Garst, B. A., & Bumpus, M. F. (2015b). Factors impacting program delivery: The importance of implementation research in Extension. *Journal of Human Sciences and Extension, 3*(2), 68-82.

Garst, B. A., & Gagnon, R. J. (2016a). A structural model of camp director practices and outcomes: Does intention toward program outcomes matter? *Journal of Park and Recreation Administration*, 34(4), 51-65. doi: 10.18666/JPRA-2016-V34-I4-7293

Garst, B. A., & Gagnon, R. J. (2016b). *Parent Perceptions Study Executive Report.* Clemson University Youth Learning Institute.

Gearing, R. E., El-Bassel, N., Ghesquiere, A., Baldwin, S., Gillies, J., & Ngeow, E. (2011). Major ingredients of fidelity: A review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Review, 31*, 79-88.

Gillis, H. L., Gass, M. A., & Russell, K. C. (2008). The effectiveness of Project Adventure's behavior treatment management programs for male offenders in

residential treatment. *Residential Treatment for Children and Youth, 25*(3), 227-247.

Glover, T., Graham, T., Mock, S., Mannell, R., Carruthers, A., & Chapeskie, A. (2013). *Canadian Summer camp research project: Parent Perception of changes in children after returning home from camp.* (Phase 3). Waterloo, Ontario: University of Waterloo.

Goodman, R. M. (2000). Bridging the gap in effective program implementation: From concept to application. *Journal of Community Psychology, 28*(3), 309-321.

Gresham, F. (1989). Assessment of treatment integrity in school consultation and pre-referral intervention. *School Psychology Review, 18*(1). 37-50.

Gresham, F. (2009). Evolution of the treatment integrity concept: Current status and future directions. *School Psychology Review, 38*(4), 533-540.

Hansen, W. B. (1996). Pilot test results comparing the All Stars program with seventh grade D.A.R.E.: Program integrity and mediating variable analysis. *Substance Use and Misuse, 31*(10), 1359-1377.

Hansen, W. B. (2014). Measuring Fidelity. In Z. Sloboda & H. Petras (Eds.), *Defining prevention science*. (pp. 335-359). New York, NY: Springer.

Hecht, M. L., & Miller-Day, M. A. (2010). "Applied" aspects of the Drug Resistance Strategies Project. *Journal of Applied Communication Research, 38*(3), 215-229.

Henderson, K. A. (2016). Leisure and the Academy: Curricula and Scholarship in Higher Education. In G. J., Walker, D. Scott, & M. Stodolska (Eds.), *Leisure matters:*

The state and future of leisure studies, (pp. 345-352). State College, PA: Venture

Publishing.

Henderson, K. A., Bialeschki, M. D., & James, P. A. (2007). Overview of camp research.

Child and Adolescent Psychiatric Clinics of North America, 16, 755-767.

Hill, L., Maucione, K., & Hood, B. K. (2007). A focused approach to assessing program

fidelity. *Prevention Science*, *8*(1), 25-34.

Hoffmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear

models. *Journal of Management, 23*(6), 723-744.

Hogan, B. (2011). *How do you eat an elephant? One bite at a time*. Ft. Lauderdale, FL:

Llumina Press.

Howard, G. S., & Dailey, P. R. (1979). Response-shift bias: A source of contamination of

self-report measures. *Journal of Applied Psychology, 64*(2), 144-150.

Hox, J. J., & Roberts, J. K. (2011). Multilevel analysis: Where we were and where we

are. In J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel modeling*.

(pp. 3-14).

Hoy, W. K., & Feldman, J. A. (1987). Organizational health: The concept and its

measure. *Journal of Research and Development in Education, 20*(4), 30-38.

Hsieh, H-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis.

Qualitative Health Research, 15(9), 1277-1288.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity

to underparameterized model misspecification. *Psychological Methods*, *3*, 424-

453.

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence of subjective performance evaluation in education. *Journal of Labor Economics, 26*(1), 101-136.

James Bell Associates. (2009, October). *Evaluation brief: Measuring implementation fidelity.* Arlington, VA: James Bell Associates.

James, L. A., & James, L. R. Integrating work environment perceptions: Explorations into the measurement of meaning. *Journal of Applied Psychology, 74*(5), 739-751.

Jelalian, E., Foster, G. D., Sato, A. F., Berlin, K., S., McDermott, C., & Sundal, D. (2014). Treatment adherence and facilitator characteristics in a community based pediatric weight control intervention. *International Journal of Behavioral Nutrition and Physical Activity, 11*(17), 1-8.

Johnson, E., Mellard, D., Fuchs, D., & McKnight, M. (2006). Responsiveness to intervention (RTI): How to do it. Lawrence, KS: National Research Center on Learning Disabilities.

Johnson, K. W., Ogilvie, K. A., Collins, D. A., Shamblen, S. R., Dirks, L. G., Ringwalt, C. L., & Norland, J. J. (2010). Studying implementation quality of a school-based prevention curriculum in frontier Alaska: Application of video-recorded observations and expert panel judgment. *Prevention Science, 11*, 275-286.

Jorde-Bloom, P. (1996). The quality of work life in NAEYC accredited and nonaccredited early   childhood programs. *Early Education and Development, 7*, 301–317.

Kaczynski, A. T., & Henderson, K. A. (2007). Environmental correlates of physical activity: A review of evidence about parks and recreation. *Leisure Sciences, 29*(4), 315-354.

Kam, C., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the PATHS curriculum. *Prevention Science, 4*(1), 55-63.

Kleiber, D. A., Walker, G. J., & Mannell, R. C. (2011). *A social psychology of leisure* (2ⁿᵈ ed.). State College, PA: Venture Publishing.

Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling* (3ʳᵈ Ed.). New York, NY: The Guilford Press.

Kogan, J. R., Holmboe, E. S., & Hauer, K. E. (2009). Tools for direct observation and assessment of clinical skills of medical trainees: A systematic review. *Journal of the American Medical Association, 302*(12), 1316-1326.

Kreft, I., & De Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: SAGE Publications.

Lamb, T. (2005). The retrospective pretest: An imperfect but useful tool. *The evaluation exchange, 11*(2), 18-19.

Larsen, T., & Samdal, O. (2007). Implementing Second Step: Balancing fidelity and program adaptation. *Journal of Education and Psychological Consultation, 17*(1), 1-29.

Lau, A. S. (2006). Making the case for selective and directed cultural adaptations of

    evidence-based treatments: Examples from parent training. *Clinical Psychology:*

    Science and Practice, 13(4), 295-310.

Lefebvre, R. C., & Flora, J. A. (1988). Social marketing and public health intervention.

    Health Education Quarterly, 15(3), 299-315.

Lesesne, C. A., Lewis, K. M., Poindexter-White, C., Green, D. C., Duffy, J. L., &

    Wandersman, A. (2008). Promoting science-based approaches to teen pregnancy

    prevention: Proactively engaging the three systems of the interactive systems

    framework. *American Journal of Community Psychology, 41*, 379-392.

Lillehoj, C. J., Griffin, K. W., & Spoth, R. (2004). Program provider and observer ratings

    of school-based preventive intervention implementation: Agreement and relation

    to youth outcomes. *Health Education and Behavior, 31*(2), 242-257.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with

    missing values. *Journal of the American Statistical Association, 83*(404), 1198-

    1202.

Little, M. A., Sussman, S. Sun, P., & Rohrbach, L. A. (2013). The effects of

    implementation fidelity in the Towards No Drug Abuse dissemination trail.

    Health Education, 113(4),   281-296.

Long, J. & Hylton, K. (2012). Reviewing research evidence and the case of participation

    in sport and physical recreation by black and minority communities. *Leisure*

    Studies, 33(4), 379-   399.

Lüdtke, O., Robitzsch, A., & Grund, S. (2017). Multiple imputation of missing data in multilevel designs: A comparison of different strategies. *Psychological Methods, 22*(1), 141-165.

Macmillan, R. (1998). Approaches to leadership: What comes with experience? *Educational Management Administration Leadership, 26*, 179–184.

Mainieri, T. L., & Anderson, D. M. (2015a). Exploring the "black box" of programming: Applying systematic implementation evaluation to a structured camp curriculum. *Journal of Experiential Education, 38*(2), 144-161.

Mainieri, T. L., & Anderson, D. M. (2015b). Exploring the postcamp civic outcomes and supporting program features of the Teens Leading and Connecting Program. *Journal of Park and Recreation Administration, 33*(2), 54-72.

Mainieri, T. L. (2016, February). *Using counselor implementation journals to explore the processes at work in two girl scout camps.* Paper and poster presented at the annual meeting of the American Camp Association Research Symposium, Atlanta, GA.

Malloy, M., Acock, A., DuBois, D. L., Vuchinich, S., Silverthorn, N., Ji, P., & Flay, B. R. (2015). Teachers' perceptions of school organizational climate as predictors of dosage and quality of implementation of a social-emotional and character development program. *Prevention Science, 16*, 1086-1095.

Marsh, H., Hau, K., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers

in overgeneralizing Hu and Bentler's (1999) findings, *Structural Equation Modeling, 11*, 320–341.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2007). Situational judgement tests, response instructions, and validity: A meta-analysis. *Personnel Psychology, 60*, 63-91.

McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgement tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9* (1/2), 103-113.

McGrew, J. H., Bond, G. R., Dietzen, L., & Salyers, M. (1994). Measuring the fidelity of implementation of a mental health program model. *Journal of Consulting and Clinical Psychology, 62*(4), 670-678.

McPhail, S. M., & Stelly, D. J. (2010). Validation strategies. In J.C. Scott & D.H. Reynolds (Eds.), *Handbook of workplace assessment: Evidence-based practices for selecting and developing organizational talent* (pp. 671-710). San Francisco, CA: Jossey-Bass.

Meyers, D. C., Durlak, J. A., & Wandersman, A. (2012). The quality implementation framework: A synthesis of critical steps in the implementation process. *American Journal of Community Psychology, 50*, 462-480.

Mihalic, S. F., & Irwin, K. (2003). Blueprints for Violence Prevention: From research to real-world settings-factors influencing the successful replication of model programs. *Youth Violence and Juvenile Justice, 4*(1), 307-329.

Mihalic, S., Irwin, K., Elliot, D., Fagan, A., & Hansen, D. (2004). Blueprints for violence
prevention (NCJ 204274). Washington, DC: Office of Juvenile Justice and
Delinquency Prevention.

Molloy, L. E., Moore, J. E., Trail, J., Van Epps, J. J., & Hopfer, S. (2013). Understanding
real-world implementation quality and "active ingredients" of PBIS. *Prevention
Science, 14*, 593-605.

Moore, J. E., Bumbarger, B. K., & Rhoades-Cooper, B. (2013). Examining adaptations of
evidence-based programs in natural contexts. *Journal of Primary Prevention, 34*,
147-161.

Morgan, C., Sibthorp, J., & Browne, L. P. (2016). Moving beyond outcomes: An applied
example of implementation evaluation in a youth recreation program. *Journal of
Park and Recreation Administration, 34*(4), 66-81.

Mowbray, C., Holter, M., Teague, G., & Bybee, D. (2003). Fidelity criteria:
Development, measurement, and validation. *American Journal of Evaluation, 24*,
315-340.

My Brother's Keeper. (2016). *My Brother's Keeper 2016 progress report: Two years of
expanding opportunity and creating pathways to success*. Washington, DC:
Author. Retrieved from:
https://www.whitehouse.gov/sites/whitehouse.gov/files/images/MBK-2016-
Progress-Report.pdf

Neiltyson. (2016, February 5). *In science, when human behavior enters the equation, things go nonlinear. That's why Physics is easy and Sociology is hard.* [Twitter post]. Retrieved from https://twitter.com/neiltyson/status/695759776752496640

Nobel, O. B., Zbylut, M. L., Fuchs, D., Campbell, K., Brazil, D., & Morrison, E. (2006). Leader experience and the identification of challenges in a stability and support operation (Technical Report 1186). *United States Army Research Institute for the Behavioral and Social Sciences*, 1–38.

Oh, C., Richardson, R. B., & Lacher, G. (2016). Economic decision making in leisure and recreation. In G. J., Walker, D. Scott, & M. Stodolska (Eds.), *Leisure matters: The state and future of leisure studies*, (pp. 17-24). State College, PA: Venture Publishing.

Pankratz, M. M., Jackson-Newsom, J., Giles, S. M., Ringwalt, C. L., Bliss, K., & Bell, M. L. (2006). Implementation fidelity in a teacher-led alcohol use prevention curriculum. *Journal of Drug Education, 36*(4), 317-333.

Pas, E. T., Waasdorp, T. E., & Bradshaw, C. P. (2015). Examining contextual influences on classroom-based implementation of positive behavior support strategies: Findings from a randomized controlled effectiveness trial. *Prevention Science, 16*, 1096-1106.

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage Publications.

Pawson, R., & Tilley, N. (1997). Realistic evaluation. London: SAGE Publications.

Peeter, H., & Lievens, F. (2005) Situational judgement tests and their predictiveness of

    college students' success: The influence of faking. *Educational and Psychological*

    Measurement, 65(1), 70-89.

Perepletchikova, F., Hilt, L. M., Chereji, E., & Kazdin, A. E. (2009). Barriers to

    implementing treatment integrity procedures: Survey of treatment outcome

    researchers. *Journal of Consulting and Clinical Psychology, 77*(2), 212-218.

Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in

    psychotherapy research: Analysis of the studies and examination of the associated

    factors. *Journal of Consulting and Clinical Psychology*, *75*(6), 829-841.

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A*

    practical guide. Malden, MA: Blackwell Publishing.

Pettigrew, J. P., & Gagnon, R. J. (2016, May). *Do implementer perceptions predict pro-*

    fidelity beliefs? Poster presented at the annual meeting of the Society for

    Prevention Research, San Francisco, CA.

Popper, K. (1981). Science, pseudo-science, and falsifiability. *Scientific thinking*, 92-99.

Potter, J. W., & Levine-Donnerstein, D. (1999). Rethinking validity and reliability in

    content analysis. *Journal of Applied Communication Research, 27*(3), 258-284.

Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity

    building. *American Journal of Evaluation, 29*(4), 443-459.

Riley, B. L., Taylor, S. M., & Elliott, S. J. (2001). Determinants of implementing heart

    health promotion activities in Ontario public health units: A social ecological

    perspective. *Health Education Research, 16*(4), 425-411.

Ringwalt, C., Pankratz, M., Gottfredson, N., Jackson-Newsom, J., Dusenbury, L., Giles, S., Currey, D., & Hansen, B. (2009). The effects of students' curriculum engagement, attitudes toward their teachers, and perception of their teachers' skills on school-based prevention curriculum outcomes. *Journal of Drug Education, 39*(3), 223-237.

Rohrbach, L. A., Graham, J. W., & Hansen, W. B. (1993). Diffusion of a school-based substance abuse prevention program: Predictors of program implementation. *Preventative Medicine, 22*(2), 237-260.

Rohrbach, L. A., Gunning, M., Sun, P., & Sussman, S. (2010). The project Towards No Drug   Abuse (TND) dissemination trial: Implementation fidelity and immediate outcomes. *Prevention Science, 11*, 77-88.

Russell, T. L. (2010). Knowledge and skill. In J.C. Scott & D.H. Reynolds (Eds.), *Handbook of   workplace assessment: Evidence-based practices for selecting and developing   organizational talent* (pp. 141-164). San Francisco, CA: Jossey-Bass.

Sackett, D. L., Rosenberg, W. C., Gray, J., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal, 312*(7023), 71-72.

Sackett, P. R. (2010). Final thoughts on the selection and assessment field. In J.C. Scott & D.H. Reynolds (Eds.), *Handbook of workplace assessment: Evidence-based practices for selecting and developing organizational talent* (pp. 757-777). San Francisco, CA: Jossey-Bass.

Schippmann, J. S. (2010). Competencies, job analysis, and the next generation of

modeling. In J.C. Scott & D.H. Reynolds (Eds.), *Handbook of workplace*

assessment: Evidence-based practices for selecting and developing

organizational talent (pp. 671-710). San Francisco, CA: Jossey-Bass.

Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data

management in counseling psychology. *Journal of Counseling Psychology, 57*,

1-10.

Sibthorp, J., Paisley, K., Gookin, J., & Ward, P. (2007). Addressing response shift bias:

Examining the use of retrospective pretests in recreation research and evaluation.

Journal of Leisure Research, 39(2), 295-315.

Schmit, M. J., & Strange, J. M. (2010). Assessment for supervisory and early leadership

roles. In J.C. Scott & D.H. Reynolds (Eds.), *Handbook of workplace assessment:*

Evidence-based practices for selecting and developing organizational talent (pp.

365-394). San Francisco, CA: Jossey-Bass.

Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., &

Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement

of implementation fidelity. *Administration and Policy in Mental Health and*

Mental Health Services Research, 38, 32-43.

Sherer, M., Maddux, J. E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R.

W. (1982). The self-efficacy scale: Construction and validation. *Psychological*

Reports, 51(2), 663-671.

Sloboda, Z., Dusenbury, L., & Petras, H. (2014). Implementation science and the

effective delivery of evidence-based prevention. In Z. Sloboda & H. Petras (Eds.),

Defining prevention science. (pp. 293-314).

Spoth, R., Guyll, M., Lillehoj, C. J., Redmond, C., & Greenberg, M. (2007). PROSPER

study of evidence-based intervention implementation quality by community-

university partnerships. *Journal of Community Psychology, 35*(8), 981-999.

Stein, M., Berends, M., Fuchs, D., McMaster, K., Sáenz, L., Yen, L., Fuchs, L., &

Compton, D. (2008). Scaling up an early reading program: Relationships among

teacher support, fidelity of implementation, and student performance across

different sites and years. *Educational Evaluation and Policy Analysis*, *30*(4), 368-

388.

Substance Abuse and Mental Health Services Administration. (2002). *Finding the*

balance: Program fidelity and adaption in substance abuse prevention. A state-of-

the-art review. (Report 277-99-6203). Retrieved from

http://www.csun.edu/sites/default/files/FindingBalance1.pdf

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation

data. *American Journal of Evaluation, 27*(2), 237-246.

Tucker, A. R., & Rheingold, A. (2010). Enhancing fidelity in adventure education and

adventure therapy. *Journal of Experiential and outdoor Education, 33*, 258–273.

Ungar, M. (2012, September). *Camps help make children resilient*. Camping Magazine.

US Census. (2015). *Educational Attainment in the United States* (Publication No. P20-

578). Retrieved from:

www.census.gov/content/dam/Census/library/publications/2016/demo/p20-578.pdf

Van Scotter, J., Motowidlo, S. J., & Cross, T. C. (2000). Effects of task performance and contextual performance on systemic rewards. *Journal of Applied Psychology, 85*(4), 526-535.

Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L., Blachman, M., Dunville, R., & Saul, J. (2008). Bridging the gap between prevention research and practice: The Interactive Systems Framework for dissemination and implementation. *American Journal of Community Psychology, 41*(3-4), 171–181.

Wandersman, A., Imm, P., Chinman, M., & Kaftarian, S. (2000). Getting To outcomes: A results-based approach to accountability. *Evaluation and program planning, 23*, 389-395.

Wanless, S. B., & Domitrovich, C. E. (2015a). Readiness to implement school-based social-emotional learning interventions: Using research on factors related to implementation to maximize quality. *Prevention Science, 16*, 1037-1043.

Wanless, S. B., Rimm-Kaufman, S. E., Abry, T., Larsen, R. A., & Patton, C. L. (2015b). Engagement in training as a mechanism to understanding fidelity of implementation of the responsive classroom approach. *Prevention Science, 16*, 1107-1116.

Weekley, J. A., & Ployhart, R. E. (2005). Situational judgement: Antecedents and relationships with performance. *Human Performance, 18*(1), 81-104.

Weekley, J. A., & Ployhart, R. E. (2006). *Situational judgement tests: Theory, measurement, and application*. J. A. Weekley & R. E. Ployhart. Mahwah, NJ: Lawrence Erlbaum Associates.

Wehby, J. H., Maggin, D. M., Moore-Partin, T. C., & Robertson, R. (2012). The impact of working alliance, social validity, and teacher burnout on implementation fidelity of the good behavior game. *School Mental Health, 4*(1), 22-33.

Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgement tests: An overview of current research. *Human Resource Management Review, 19*, 188-202.

Whetzel, D. L., & Reeder, M. C. (2016). Why some situational judgement tests fail to predict job performance (and others succeed). *Industrial and Organization Psychology, 9*(1), 71-77.

Wigfield, A., Tonks, S.., & Eccles, J. S. (2004) Expectancy value theory in cross-cultural perspective. *Big Theories Revisited: Research on Sociocultural Influences on Motivation and Learning*, 4, 165-198.

Witt, P. A., & Caldwell, L. L. (2010). *The rationale for recreation services for youth: An evidenced based approach.* Ashburn, VA: The National Recreation and Parks Association.

Williams-White, S., Keonig, K., Scahill, L. (2007). Social skills development in children with autism spectrum disorders: A review of the intervention research. *Journal of Autism Development Disorders, 37*, 1858-1868.

Williford, A. P., Sanger-Wolcott, C., Vick-Whittaker, J., & Locasale-Crouch, J. (2015). Program and teacher characteristics predicting the implementation of Banking

Time with preschoolers who display disruptive behaviors. *Prevention Science, 16*, 1054-1063.

Zief, S., Henke, J., Knab, J., & Zaveri, H. (2011). *Planning evaluations designed to meet scientific standards: Communicating key components of the plan for a rigorous and useful evaluation of a teenage pregnancy prevention program* (HHS Office of Adolescent Health Task No. HHSP23337017T). Retrieved from: https://www.hhs.gov/ash/oah/oah-   initiatives/ta/ta-planning-evaluations.pdf

Zollo, M., & Gottschalg, O. (2004). *When does experience hurt? The confidence-competence paradox*. Fontainebleau, France: INSEAD.

Zvoch, K. (2012). How does fidelity of implementation matter? Using multilevel models to detect relationships between participant outcomes and the delivery and receipt of treatment. *American Journal of Evaluation. 33*(4), 547-565.