November 2015

Implementing Universal Social And Emotional Learning Programs: The Development, Validation, And Inferential Findings From The Schoolwide SEL Capacity Assessment

Cheyne A. LeVesseur
University of Massachusetts Amherst

Implementing Universal Social And Emotional Learning Programs: The Development, Validation, And Inferential Findings From The Schoolwide SEL Capacity Assessment

A Dissertation Presented

By

CHEYNE ALLEN LEVESSEUR

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2015

School Psychology
College of Education

Implementing Universal Social And Emotional Learning Programs: The Development, Validation, And Inferential Findings From The Schoolwide SEL Capacity Assessment

A Dissertation Presented

By

CHEYNE ALLEN LEVESSEUR

Approved as to style and content by:

_____
Sara A. Whitcomb, Co-Chair

_____
Jennifer Randall, Co-Chair

_____
John Hintze, Member

_____
Christopher Overtree, Member

_____
Christine B. McCormick, Dean
College of Education

DEDICATION

I dedicate this to my mother.

ACKNOWLEDGEMENTS

Thank you Sara, Jennifer, John, and Chris for your support. I couldn't be more grateful.

ABSTRACT

IMPLEMENTING UNIVERSAL SOCIAL AND EMOTIONAL LEARNING PROGRAMS: THE DEVELOPMENT, VALIDATION, AND INFERENTIAL FINDINGS FROM THE SCHOOLWIDE SEL CAPACITY ASSESSMENT

SEPTEMBER 2015

CHEYNE ALLEN LEVESSEUR, B.A., OAKLAND UNIVERSITY

M.Ed., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Sara A. Whitcomb and Professor Jennifer Randall

In order to effectively transport universal social and emotional learning (SEL) programs into natural settings, it is important to understand implementation barriers that may hinder the likelihood of successful outcomes (Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005). The current study is primarily based on the notion that within the planning phase of implementation, few technically adequate assessment measures targeting both organizational capacity (OC) and provider characteristics (PC) for SEL programming actually exist. The purpose is to extend the SEL implementation assessment literature by developing a new rating scale designed to measure SEL implementation barriers (School SEL Capacity Assessment [SSCA]) and make preliminary inferences regarding the current state of SEL implementation. In order to satisfy our objectives, we evaluated the psychometric quality of the SSCA using the Rasch Rating Scale model. In all, the data are encouraging and provide promising validity evidence of internal structure for the OC and PC scales within the SSCA. Within reason, the SSCA met the qualifications for accurate measurement. Findings from the

Rasch analysis helped us analyze survey response differences among subgroups, which included participants' stage-of-implementation, years of professional experience, grade level, and social economic status. Generally speaking, teachers who are maintaining an SEL program found it easier to endorse items on the SSCA, suggesting that they have more OC and PC attributes. Contrary to what we would expect, years of professional experience is not related to teachers' level of OC and PC attributes. Because they're so few teachers in several of the grade levels, any comparison would be tenuous and we have opted to not to include the analysis in our results. As anticipated however, the results show that teachers in high SES schools have significantly more OC attributes, but their level of PC attributes are not affected by their schools SES. Limitations of this study as well as directions for future research are discussed.

Keywords: Social and emotional learning, implementation, organizational capacity, provider characteristics, rating scale, assessment, Rasch Rating Scale model

TABLE OF CONTENTS

Page

LIST OF TABLES

Tables Page

LIST OF FIGURES

Figures Page

CHAPTER 1

INTRODUCTION

Statement of the Problem

The treatment of mental health problems has historically focused on curing dysfunction and the sequelae of psychological problems, which tends to be reactive and costly to society (Mrazek & Haggerty, 1994; Seligman & Csikszentmihalyi, 2000). Socioeconomic indices have estimated that the total economic costs associated with mental illness was $148 billion in 1996 (an estimated $274 billion in 2015) and the total economic costs associated with substance abuse was $511 billion in 1999 (an estimated $727 billion in 2015) (World Health Organization, 2003; Miller & Hendrie, 2008). Although the economic costs are large, when we consider that one third of the population will face similar issues, the cost of suffering is immeasurable.

The general malaise is further compounded by the fact that laws at the federal level have been largely slow to adopt comprehensive systems for addressing mental health problems (Mrazek & Haggerty, 1994). The reluctance among policy makers is unfortunate considering it is estimated that 12 to 22 percent of our nation's children under the age of 18 will suffer from one or more mental health disorders and will be in need of services. (Greenburg, Bumbarger, & Domitrovich, 2001).

Along the same lines, a report from the U.S. Government Accountability Office (2012) revealed that children who are at a higher risk of mental health problems and are also receiving psychotropic medication—such as those on Medicaid, State Children's Health Insurance Programs, and in foster care—are rarely given counseling or therapy to complete their care. Despite the concern for the mental health of our nation's children,

our society's actions seem to prevent prevention programming, which is all too alarming considering the economic benefits yield $2-$20 for every dollar spent on prevention (Greenburg et al., 2001; Kauffman, 1999). This provides clear evidence that universal prevention efforts are not just important, but imperative.

Fuchs, Mock, Morgan, and Young (2003) suggest that schools should begin the prevention process before a problem manifests. A proactive approach that focuses on universal prevention, risk reduction, and early intervention can be considered least restrictive to a student in comparison to providing treatment after a problem has become a part of the student's behavioral repertoire (Yell, 2011). These theoretical underpinnings can be applied within a tiered model of social-emotional support and prevention in a practical school setting. A multilevel prevention system consists of universal (also referred to as primary or schoolwide), targeted, and individualized prevention (Fuchs et al., 2005). In theory, all students should be provided universal support and if identified as at-risk, more intensive interventions should be provided to accommodate their needs.

Social and Emotional Learning

Greenberg et al., (2003) published an article on the benefits of social and emotional competency that was the direct result of the work of the APA Task Force on Prevention: Promoting Strength, Resilience, and Health in Young People. The ineffective nature of many prevention and health promotion efforts of the 1990's concerned practitioners, researchers and policy-makers. For example, the national D.A.R.E. program was not only limited by its categorical approach to prevention but also failed to effectively ameliorate drug use among our youth (Greenberg et al., 2003). As a consequence, the term social and emotional learning (SEL) was introduced to address

underlying causes of targeted problem behaviors while also supporting academic achievement. In sum, the goal was to defragment the piecemeal categorical prevention programming efforts and develop a unified conceptual framework.

SEL is conceptualized as how we work with people, manage our own emotions, and live productive and healthy lives (Merrell, 2000). SEL has been broadly defined by the Collaborative for Academic, Social, and Emotional Learning ([CASEL]; see www.CASEL.org) as "the process of acquiring the skills to recognize and manage emotions, develop caring and concern for others, make responsible decisions, establish positive relationships, and handle challenging situations effectively" (CASEL, 2003, p. 1). CASEL (2003) has identified five skills or competencies that include:

1. Self-awareness—accurately assessing one's feelings, interests, values, and strengths; maintaining a well-grounded sense of self-confidence.
2. Self-management—regulating one's emotions to handle stress, control impulses, and persevere in overcoming obstacles; setting and monitoring progress toward personal and academic goals; expressing emotions appropriately**.**
3. Social awareness—being able to take the perspective of and empathize with others; recognizing and appreciating individual and group similarities and differences; recognizing and using family, school, and community resources.
4. Relationship skills—establishing and maintaining healthy and rewarding relationships based on cooperation; resisting inappropriate social pressure; preventing, managing, and resolving interpersonal conflict; seeking help when needed**.**

5. Responsible decision-making—making decisions based on consideration of ethical standards, safety concerns, appropriate social norms, respect for others, and likely consequences of various actions; applying decision-making skills to academic and social situations; contributing to the well-being of one's school and community.

For children, social and emotional competence represents a protective mechanism that may interrupt a risk cycle by improving adaptive response to risk (Pianta, 1998). Developing social and emotional competence helps children negotiate diverse contexts within their developmental level (Greenberg et al., 2003). For example, as Greenberg et al. (2003) point out, children with social and emotional competence show a reduction in substance abuse, antisocial behavior, and improved mental health, school nonattendance, and academic performance. More recently, Durlak, Weissberg, Dymnicki, Taylor, and Schellinger (2011) published a meta-analysis of 213 school-based, universal social and emotional learning programs in which students demonstrated significantly improved social and emotional skills, attitudes, behavior, and an 11-percentile-point gain in academic achievement.

Durlak et al. (2011) concluded that two key variables moderate student outcomes: the use of four recommended practices (i.e., SAFE practices) and quality implementation. SAFE is an acronym for sequenced, active, focused, and explicit practices. As indicated by the authors, previous research has supported the use of SAFE practices, demonstrating that a combination of these practices equate to more effective programming.

In more detail, sequencing implies a coordinated and progressive set of activities that provide ample opportunity to connect the steps. Active is akin to the adage,

"practice makes perfect", and involves active participation; which requires students to be focused, involving sufficient time and attention. Lastly, programming should be explicit, with clear and specific learning objectives. Moreover, effective programming is only possible through quality implementation, the second key moderator. Evidence-based programs are necessary but ultimately, insufficient. Both SAFE practices and implementation moderate positive student outcomes, and although evidence-based interventions may be efficacious, the effectiveness of an intervention in natural settings can only be achieved through quality implementation.

Implementation Science

To improve universal prevention service delivery, one must have an understanding of the tenets of implementation science. Historically, implementation science has suffered from a lack of consensus among terminology and definition of the construct, a lack of research identifying a systems framework of moderating and mediating variables, and minimal development of assessment tools (Schulte, Easton, & Parker, 2009).

Implementation is referred to as implementation quality, which is often used interchangeably with and similarly defined as treatment integrity, fidelity, adherence and procedural reliability; all of which have been subjected to oversimplified definitions (Sanetti & Kratochwill, 2009). The most prevalent definition—the degree to which treatment is delivered as intended—has been revised to capture the complex nature of this construct. Fixsen et al, (2005, pg. 5) interpret implementation as "a specified set of activities designed to put into practice an activity or program of known dimensions." While the definition incorporates a more complex systems-level approach, it may lack a

few important specific details, such as the dimensions of quality of delivery and the intervention as received and utilized by consumers (Lichstein, Riedel, & Greive, 1994).

Combining Fixsen et al., (2005) updated definition with Lichstein et al., (1994) proposed implementation model broadens the conceptualization by incorporating the dimensions of quality of delivery and the intervention as received. A proposed definition is as follows: Implementation is defined as a specified set of activities designed to ensure essential program components are delivered with sufficient quantity and quality, and received and utilized appropriately by consumers.

Currently, our understanding of implementation science is still in its infancy but recent advances have identified contextual factors that moderate and mediate the delivery of prevention programs. Durlak and DuPre (2008) proposed a multilevel ecological perspective that delineates contextual factors affecting the implementation process. In brief, five categories of associated variables have been identified: Community Level Factors (e.g., policy); Provider Characteristics (e.g., self-efficacy); Characteristics of the Innovation (e.g., adaptability); the Prevention Delivery System (i.e., Organizational Capacity; e.g., leadership); and the Prevention Support System (e.g., training).

In order to understand implementation barriers, it is important to understand the school and provider's current level of program use, comfort with program, their current willingness to change their teaching behavior if needed, and their perceptions of support related to program use (Elder, Ayala, & Harris, 1999). Unfortunately, principals and other educational leaders (i.e., isolated from the classroom) tend not to have a reliable understanding of actual classroom practice and not all teachers are ready to implement universal prevention programming (Elder et al., 1999); Hall & Loucks, 1977)

As such, Prochaska and DiClemente's (1982) Transtheoretical, or Stages-of-Change (SOC) Model, is proposed as a heuristic for understanding providers' thoughts and behaviors related to change regarding classroom and schoolwide implementation efforts and reform. The SOC model provides a diagnostic framework that represents a temporal dimension for behavior change, ranging from precontemplation (i.e., not intending to take action) to maintenance (i.e., has already taken action). While the SOC Model is typically associated with individual behavior change, there is some limited evidence to suggest that it can be used to support providers' schoolwide implementation efforts (Prochaska, 2000; Prochaska, 2006). While several universal SEL programs have been found to be efficacious, there is still much to learn about program effectiveness. To effectively transport universal SEL programs into natural settings, it is important to understand implementation barriers that may hinder the likelihood of successful outcomes (Fixsen et al., 2005).

Purpose and Significance

The current study is primarily based on the notion that within the planning phase of universal SEL implementation, few technically adequate assessment measures exist, targeting both Provider Characteristics and Organizational Capacity, for SEL programming (see Teacher SEL Beliefs Scale; Brackett, Reyes, Rivers, Elbertson, & Salovey, 2012). The purpose of the current study is to extend the SEL implementation assessment literature by developing a new instrument that targets all professional school staff and evaluates their perceived need, benefits, self-efficacy, and skill proficiency for SEL programming, and school staff belief about the capacity for the school's universal

prevention and intervention delivery system, general organizational factors, practices/processes, staffing, and the support system.

The proposed instrument uses Durlak and DuPre's (2008) multilevel ecological perspective as a frame for items developed for the survey. The survey will be used to identify implementation barriers that are alterable (e.g., training) rather than those that are distal (e.g., federal policy) and less malleable. Because of this rationale, not all of the contextual factors identified by Durlak and DuPre (2008) were used in the construction of the survey. Instead, those factors that were more malleable to local reform efforts and had the potential to undergo immediate change were deemed more appropriate for the purpose of this instrument. For the sake of parsimony, the five categories proposed by Durlak and DuPre (2008) were consolidated into two principal components: Provider Characteristics and Organizational Capacity.

The Current Study

The objective of the current study is twofold: (1) to extend the SEL implementation assessment literature by developing a new rating scale designed to measure SEL implementation barriers; and (2) make preliminary inferences regarding the current state of SEL implementation.

The current instrument—the Schoolwide SEL Capacity Assessment (SSCA)—has been developed according to the constructive approach to measurement proposed by Wilson (2005). The framework, construct modeling, is an instrument development process that uses four building blocks to guide the construction of an instrument to measure the SEL implementation barriers construct. The development cycle consists of (1) construct mapping, (2) item design, (3) outcome space, and the (4) measurement

model. The development cycle is a generative and cyclical process that not only leads to instrument development but also supports the evaluative or validation process (Wilson, 2005; Smith & Smith, 2007).

Instrument design was facilitated through a construct mapping process in which conceptualization of the construct was investigated through a literature review. Durlak and DuPre's (2008) meta-analysis on factors affecting the implementation process provided a framework for the SEL implementation barriers construct. Although most items in the SSCA target factors affecting implementation, as specified by Durlak and DuPre's (2008) meta-analysis, one item addresses Prochaska and DiClemente's (1982) SOC Model, with the intent to identify schools and providers' current level of universal SEL program use and current stage within the change cycle.

Within the first objective of this study, the aim was to evaluate the psychometric quality of the SSCA using the Rasch Rating Scale model (i.e., quality control procedures; Wright & Masters, 1982). Using the SEL implementation barriers framework employed for this study and findings from the Rasch analysis, the second objective of this study—making preliminary inferences regarding the current state of universal SEL implementation—were addressed. Within the second objective, the aim was fourfold:

1. Assess differences between participants "stage of change" within the SOC model (Prochaska & DiClemente, 1982).
2. Assess differences between participants with different years of professional experience.
3. Assess differences between participants' grade level at which they currently work.
4. Assess differences between schools' social economic status.

CHAPTER 2

LITERATURE REVIEW

Introduction

A comprehensive mission for schools is to create a challenging learning environment that encourages high expectations for success. Schools may carry out this mission by fostering safe, orderly, caring, and supportive environments that emphasize the social, emotional, physical, and intellectual development of each child. This mission may be supported by the use of schoolwide prevention service delivery for all students, such as the use of social and emotional learning (SEL). Yet, the current impact of schoolwide SEL and other prevention programming has been hindered by the lack of research addressing how to effectively transport programs into school settings.

The purpose of this chapter is to provide an overview of the need for SEL programming and bridge the gap between SEL and implementation science. This will be accomplished by providing background information on etiology, schoolwide prevention service delivery, SEL programming, and evaluation and implementation science. The purpose of addressing the foregoing topics is to help build understanding that a need for action exists including, how to prevent the causation of mental health problems, to acknowledge that schoolwide prevention service delivery through SEL can be effective; and how to effectively build capacity, deliver SEL programming, and formatively evaluate and monitor progress. The development of the Schoolwide SEL Capacity Assessment (SSCA) will contribute to implementation science by extending the SEL implementation assessment literature and the preliminary inferences from this study will help practitioners understand implementation variables that are essential to successful

implementation and evaluation. Practitioners will have an evaluation tool to help guide their efforts, allowing them to identify context-specific barriers and drivers, and monitor their progress over time; essentially allowing schools to weaken implementation barriers and strengthen drivers necessary for strong program impact and sustainability.

Psychological Determinants: Risk Versus Protective Factors

Current approaches to understanding the etiology of mental health problems in children have benefited greatly from epidemiological research. Populations most at risk and moderating environmental factors have been identified, and the strength of different risk factors has been studied. Risk factors are of particular interest and this is evidenced by the definition provided by Kazdin, Kraemer, Kessler, Kupfer, and Offord (1997, pg. 376): "a characteristic, experience, or event that, if present, is associated with an increase in the probability (risk) of a particular outcome over the base rate of the outcome in a general (unexpected) population."

Risk and protective factors may best be understood as a strength-weakness continuum (Doll & Lyon, 1998). Risk factors weaken a student's capacity to overcome adversity and may lead to dysfunction. Examples of risk factors include deficits in social abilities, high emotional distress, and exposure to violence or conflict. Protective factors on the other hand strengthen a student's capacity to overcome risk, which may improve their resiliency to adversity. Examples of protective factors include social competency, connectedness to school, and good academic performance. In other words, risk factors are potential precursors to dysfunction and protective factors are potential precursors to resiliency (Coie et al., 1993; Pianta, 1998).

Within a developmental-ecological lens, the strength-weakness continuum of a risk or protective factor may change depending on context; in other words, factors may be a risk in one setting but protective in another (Morrison, Brown, D'incau, O'Farrell, & Furlong, 2006). For example, high sociability may be considered a protective factor during recess but viewed as a risk factor during independent seatwork in the classroom. The foregoing highlights the apparent intricacies of the literature and cautions educators to not think in simplistic terms.

Children may acquire risk factors such as school failure, involvement with antisocial peers, and family poverty that make it more likely they will develop mental health related problems. On the other hand, protective mechanisms and the development of competencies—such as social decision making skills—help children cope and lead more successful lives. Protective and risk mechanisms can stem from biological, psychological, and sociological domains; and can be shaped within many different environments: family, school, community, etc. (Kazdin et al., 1997).

Biological factors increase the likelihood of subsequently developing mental health problems; however, biology is by-and-large out of the reach of prevention efforts in schools. It would be erroneous to assume, however, that biological and environmental factors are independent of each other—in fact, development should be understood as an interplay between a multitude of influences and denotes a diffusion of each through the other, fostering a mutually created system (Coll, 2004).

A position held within the field of behavioral genetics is that the environment plays a substantial role in development; shaping individual differences in aspects of cognition and personality (Bjorklund & Pelligrini, 2002, p.81). Children may develop

characteristics consistent with their genotype in all but the most adverse environments—thus, adverse environments may restrict children's options and thus limit their development. Interestingly, as Bjorklund and Pelligrini (2002) point out, within a normative developmental framework, the correlation of heritability and intelligence is .57 and the effect of environment is only .13. However, even more interesting is that in adverse environments, heritability of intelligence is reduced to .26 and the effect of environment is increased to .23.

These findings suggest that the environment plays a much bigger role in the prevalence of low intelligence scores in impoverished and low social economic status groups. Likewise, the same deduction about intelligence may be applicable in other phenotypes. Mental health disorders are due to reactions to adverse environments and cumulative risk factors that impede a child's normative and adaptive development (Carter & Briggs-Gowan, 2004). Within a developmental-ecological lens, the environment plays a substantial role in the development of our youth and the earlier we intervene, the greater the likelihood we succeed.

Mental Health Issues Amongst our Youth

There is a preponderance of evidence to suggest that our youth are faced with a multitude of risks that lead to mental health problems. Essentially, the identified problem is not one specific mental health disorder; rather, the scope encompasses all mental health related issues. This line of reasoning assumes that mental health is an evaluative term that includes internal and external problems as an outcome.

According to a National Health and Nutrition Examination Survey from the years 2001-2004, 1 of 8 children aged eight to fifteen met a 12-month criteria for ≥1 of 6

Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (Merikangas et al., 2010). In general, there were few ethnic differences, however, poor children had higher rates of most disorders than their more affluent counterparts. These findings provide estimates of the prevalence of DSM-IV defined mental disorders in the general population of the United States; an interesting point in these indexes are that not all children will meet DSM-IV (now DSM-V) criteria for a mental health disorder, but regrettably, up to twice the amount will experience a mental health problem within their K-12 school years (Greenburg et al., 2001).

The sociodemographic correlates of mental health problems vary across gender and age (Merikangas et al., 2010). In general, there is a higher prevalence of internalizing disorders for females and externalizing disorders for males—as children mature, there is an increase in the rates of both disorders. In other words, rates of most disorders tend to be higher as children reach adolescence, and develop into adulthood. As Merikangas et al. (2010) state, comorbidity was less common among youth than in comparable studies of adults, however, comorbidity is still a concern.

The following discussion uses depression as an example to illustrate the social and emotional problems youth face. Keep in mind however, the discussion refers to clinically significant levels of depression, but lesser degrees are concerning and represent a larger web affecting our youth. The lifetime prevalence of depression among adolescents is currently 14% (Substance Abuse and Mental Health Services Administration [SAMHSA], 2006). According to the National Survey on Drug Use and Health, 5.4% to 12.3% of adolescents experienced a major depressive episode (MDE) in the past year. MDE was defined as a period of 2 weeks or longer during which there was

either depressed mood or loss of interest or pleasure and at least four other symptoms that reflect a change in functioning, such as problems with sleep, eating, energy, concentration, and self image.

Alarmingly, the survey showed an association between adolescent depression and the onset of cigarette smoking, alcohol use, and drug use. The occurrence of a MDE within the past year doubled the likelihood that the adolescent used illicit drugs in the past month, as compared to peers who had not (i.e., 21.2 vs. 9.6 percent). Equally discerning is the fact that less than half (i.e., 40.3 percent) received treatment for depression during that time.

The likelihood that an adolescent will receive services for emotional or behavioral problems in the past year is bleak (SAMHSA, 2006). Combined 2005 and 2006 data indicate that 13.3% receive services at a specialty mental health setting, 12% receive services in a school-based setting, and about 3% receive services in a general medical setting (SAMHSA, 2009). Disappointingly, even though the research demonstrates an increase in the prevalence of disorders over time for youth, the percentage of youth aged 12 to 17 who received school-based services decreased concomitantly (Merikangas, et al. 2010; SAMHSA, 2009).

As the foregoing illustrates, our society is called to shift our focus to prevention efforts. In fact, the economic value of effective prevention programs can yield benefits that exceed their costs (Zins & Elias, 2006). For instance, the economic benefits have ranged from $3.14 (i.e., Seattle Social Development Program) per dollar spent per student to upwards of $28.42 (i.e., Caring School Community) (Zins & Elias, 2006). The social benefits on the other hand, ranged from better educational outcomes to reduced crime.

Miller and Hendrie (2008) estimated that a return on investment in school-based prevention services could range between $7.40 and $36 per dollar invested, with a medium estimate of $18.

A Public Health Perspective

In light of the evidence that our youth are faced with a multitude of mental health problems, it is presumably justified to intervene early. In order to build resilience in schools, ecological, risk, and protective factors must be considered. Building resilience implies ameliorating risk and fostering individual competence—which may best be viewed within a public health perspective.

The principles of prevention, such as addressing risk and protective factors, are at the core of a public health framework. Applied in schools, a multilevel prevention system focuses on three tiers of increasingly intense support; which is often referred to as response-to-intervention (RTI) or multi-tiered systems of supports (Strein, Hoagwood, & Cohn, 2002). The public health framework of prevention is often displayed as a triangle as depicted in Figure 1.

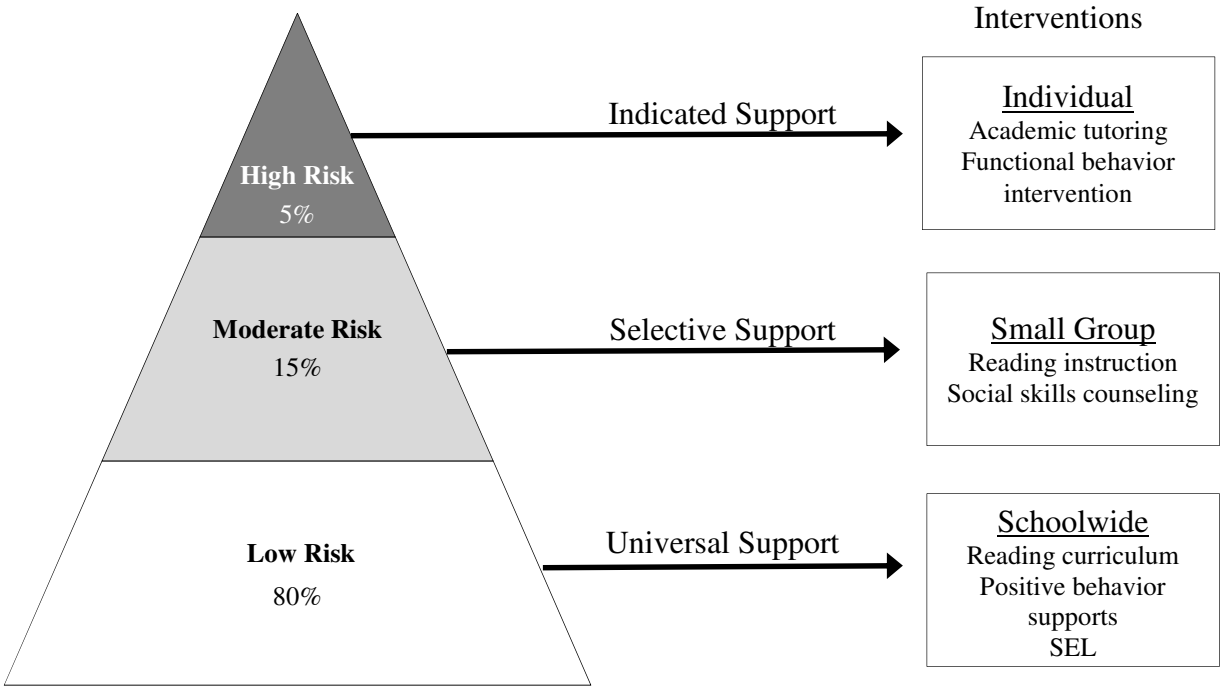

Figure 1. A Public Health Framework of Prevention

As risk increases, support increases. Tier 1, or universal systems, are schoolwide and referred to as early intervention, and should prevent around 80 percent of all problems (Strein et al., 2002). A schoolwide reading or positive behavior program given to all students, regardless of risk, is an example of the first level of support provided. Tier 2, or selective prevention, identifies students with some risk and provides targeted support, and should combat around 15 percent of remaining problems (Strein et al., 2002). Small group reading instruction and social skills counseling are examples of the second tier of support provided to some students. Lastly, Tier 3, or indicated prevention, identifies students at-risk and attempts to reduce problem behaviors that have already manifested (Strein et al., 2002). This level of intensive support should be provided to less than 5 percent of students, such as individual academic instruction and functional behavior assessments.

The public health framework is a successful and systematic process for providing preventive support services and places valuable resources in the hands of all students, increasing intensity with student needs (Strein et al., 2002). Legally speaking, schools have become increasingly motivated to ensure that students are successful, as evidenced by the No Child Left Behind Act of 2001, a federal law that calls for increased accountability for student performance (No Child Left Behind, 2002). Although resources are limited—thus providing a rationale for using resources more effectively—schools are excellent settings for providing prevention and intervention services. In fact, of the students who do receive mental health services, some estimate that as high as 75% receive services in a school setting (Rones & Hoagwood, 2000).

Social Emotional Learning

Although it is difficult for educators and related staff to change systemic factors that are conducive to risk processes, such as biological factors, school programming can provide important protective—psychological and sociological—factors that may buffer those processes (Wang, Haertel, & Walberg, 1997). A study conducted by Wang, Hsu, Lin, Cheng, and Lee (2010) researched risk behaviors and concluded that emotion regulation was an important explanatory variable for maladjustment and resilience. Eight out of the 11 most influential constructs influencing learning, involved social and emotional factors.

The authors concluded that direct intervention in social emotional competency was the most promising and potentially the most effective avenue for reform (Wang et al., 1997). They recommended teaching emotional regulation skills. The argument taken in this study is that schoolwide SEL will counteract many risk processes and provide

protective factors that may improve overall mental health—ultimately playing a role in preventing dysfunction in our youth (Seligman & Csikszentmihalyi, 2000).

The Collaborative for Academic, Social and Emotional Learning (CASEL) defines SEL as a "process for helping children and even adults develop the fundamental skills for life effectiveness" (CASEL, 2011). Further, CASEL has identified five competencies for SEL that address cognitive, affective, and behavioral skills that are critical for success, which are depicted and defined in Figure 2. One cannot argue that optimal functioning and fundamental skills for life effectiveness are not two sides of the same coin. Simply stated, prevention science and SEL go hand-in-hand.

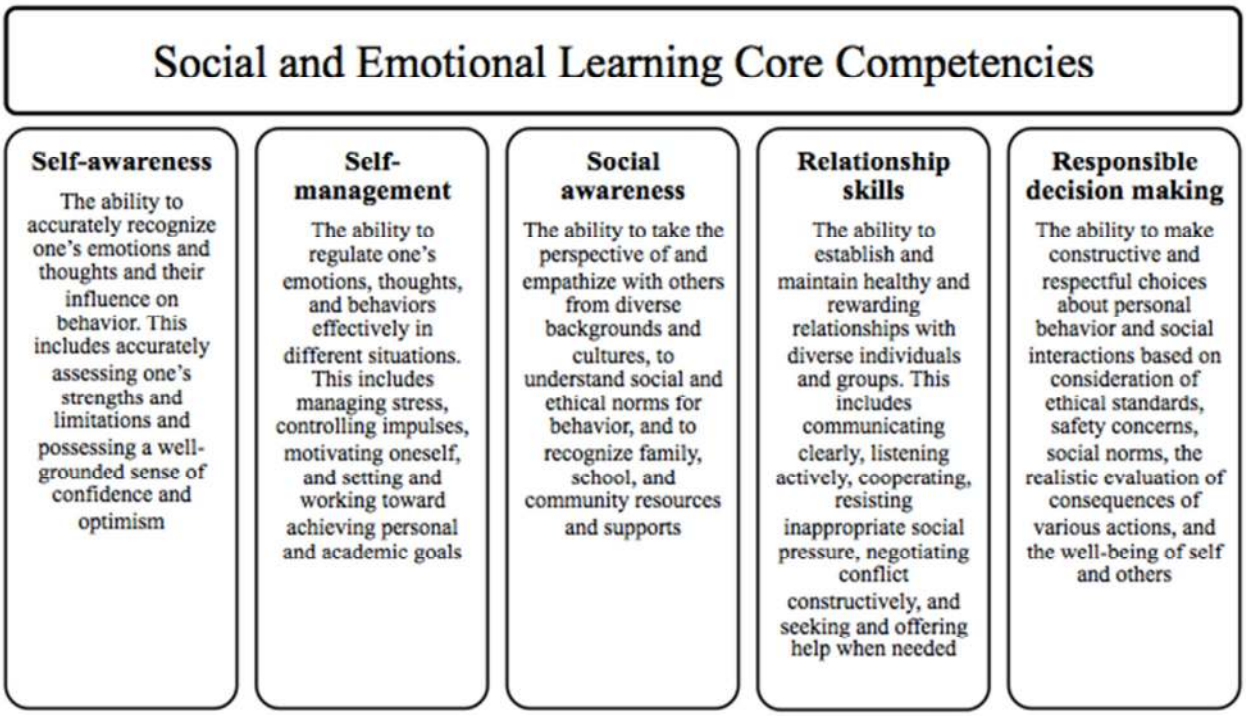

Figure 2. SEL Competencies and Definitions

There is evidence to suggest that students display problem behaviors as early as kindergarten, however, many are not identified until their middle school years (Merrell, 2000). In light of the evidence, best practice would be to intervene early, or in this case,

prevent the problem and its sequelae from manifesting. The rationale is that by increasing protective factors, children acquire the competence to buffer the deleterious effects of risk factors as they age. In essence, schoolwide SEL programming can be implemented to support a key period of development, building competence early, reducing risk and deficits, and promoting competence throughout students' school tenure.

The research base for SEL curricula and programs is expanding, and more research continues to come forward demonstrating that SEL programs are able to promote resiliency, decrease emotional and behavioral problems in students, and contribute to improved academic performance (Zins & Elias, 2006). Payton et al. (2008) conducted a meta-analysis of approximately 700 evaluation studies of SEL programs for children from preschool to high school, and concluded that behavioral and academic benefits are observed across the board. Similarly, a meta-analysis published by Durlak, Weissberg, Dymnicki, Taylor, and Schellinger (2011), using 213 studies involving universal SEL interventions for all grade levels, indicated observed improvements in behavior, attitudes, and academic performance; and concurrent decreases in conduct problems and emotional distress.

Zins and Elias (2006) provide a list of examples of SEL outcomes related to success in school and life. Each outcome is grouped into three indices: attitudes, behaviors, and performance. Some of the outcomes for attitudes consist of higher sense of self-efficacy, better sense of community, and improved coping with school stressors. Outcomes of behavior include more prosocial behavior, reductions in aggression, disruption, and interpersonal violence, and less drug, tobacco, and alcohol use and delinquent behavior. In regards to performance, some outcomes include increased math,

language arts, and social studies skills, increased achievement over time, and better problem solving and planning.

One such universal preventative program aimed at teaching SEL from grades K-2 is the program *Strong Start: A Social and Emotional Learning Curriculum*. As the authors Merrell, Parisi, and Whitcomb (2007) point out, the *Strong Kids* curriculum is practical, easy-to-use, low cost, and teacher friendly. In a time when schools are expected to do more with less, an SEL program that is able to fit well into the normal routines of schooling and harness the instructional scope and sequence of most classrooms, may provide the best means to solve two problems at one time.

The Need for Implementation Science

Prevention science has made much headway, producing efficacious mental health programs and models that can increase the chances of life success, such as the inception of SEL during the turn of the century (Domitrovich & Greenberg, 2000). If used in conjunction with a decision-making framework, such as RTI, schools will be better able to identify students in need and adjust the intensity and nature of support depending on the student's responsiveness. One important step in the decision-making process is implementation, and although Durlak et al. (2011) indicates that teachers are able to initially implement SEL programs, there is still much to learn about the durability and sustainability of these programs.

An SEL program not only has to be efficacious, but it also has to be formatted in a way that allows it to be implemented as intended, to ensure that it can be adapted and is compatible with current practices. For example, the Durlak et al. (2011) study highlighted that implementation problems are more common for some intervention

formats. Although multicomponent interventions that supplement classroom programming with schoolwide and family components (i.e., broader ecological focus) are needed to address mental health problems in our youth, compared to teacher-led programs, multicomponent programs are more likely to have implementation problems (Durlak, et al. 2011; Tolan, Guerra, & Kendall, 1995).

The literature indicates that SEL programs that encounter implementation barriers during program implementation observe attenuated results. Specifically, Durlak et al. (2011) reported significantly reduced effects in several outcome variables. Programs without implementation barriers yielded significant mean effects in skills, attitudes, social behavior, conduct problems, emotional stress, and academic performance. On the converse, programs with implementation barriers yielded significant mean effects in only two categories: attitudes and conduct problems. Most striking is that implementation barriers hindered positive outcomes in academic performance.

In a time when schools are held accountable for students' academic performance above all, the use of limited resources must be justified. If implementation barriers are not addressed prior to program implementation, universal SEL programs will not have their intended effects, preventing the justification needed to support them over time. The difference between addressing implementation barriers and not, is a mean effect size for academic performance of 0.33 and 0.14, respectively (Durlak et al., 2011). In order to implement evidence-based multicomponent SEL programs that result in academic success, implementation science matters.

Implementation science is still in its infancy, but there is plenty of research substantiating its importance (Domitrovich & Greenberg, 2000). Implementation is

synonymous with several other terms such as treatment integrity, adherence, and program reliability. The most common definition, the degree to which treatment is delivered as intended, has been criticized for its inability to capture the complexity of implementation (Hagermoser-Sanetti & Kratochwill, 2009).

On the one hand, Fixsen et al. (2005, pg. 5) sufficiently broad interpretation may suffice: "Implementation is defined as a specified set of activities designed to put into practice an activity or program of known dimensions." The strength in this definition lies in its focus on activities intended to build capacity or systems-level variables rather than just focus on individual providers' delivery. Although a broad definition can be more encompassing, it may lack specific details that are deprived of specific dimensions; such as those championed by Lichstein et al. (1994), which include dimensions of quality of delivery and the intervention as received. For the sake of being comprehensive, a proposed definition is as follows: Implementation is defined as a specified set of activities designed to ensure essential program components are delivered with sufficient quantity and quality, and utilized appropriately by receivers. The benefit of this definition is its emphasis on systems-level variables (i.e., activities for delivery), program delivery (i.e., sufficient quantity and quality), and treatment efficacy (i.e., utilized appropriately).

Until recently, implementation science has not always been recognized as an important variable in applied settings (Cochrane & Laux, 2008). Highly controlled research studies demonstrate positive outcomes for prevention programs, but when implemented in the real world, the same outcomes are not always observed. Research studies control contextual variables to strengthen cause-and-effect arguments, but often

times, the real world context in which prevention programs are implemented and the level of control over certain barriers (e.g. time, space, materials), can be quite different, leading to attenuated outcomes (Yeaton & Sechrest, 1981).

It has been argued that the gap between what we expect and what we observe is caused by these differences in facsimiles. In other words, the reduced magnitude in outcomes can be attributed to implementation barriers in natural settings (Durlak & DuPre, 2008). The planning stage of implementation, prior to actual delivery, is a necessary step for minimizing implementation barriers and supporting the development of implementation drivers to promote successful outcomes that align with the literature (Fixsen et al., 2005).

Before exploring how to evaluate and understand the multiple interacting ecological factors that moderate SEL implementation, it is important to understand teachers' current knowledge, perceptions, and practices. Buchanan, Gueldner, Tran, and Merrell's (2009) survey results indicated that teachers believe in the importance of SEL and that schools should take an active role, but current academic demands may preclude its use.

Over 96% of teachers regarded SEL as important and believe it enhances academic outcomes, but less than half (45.5%) are currently implementing SEL programs. Only 29.5% have received training via a workshop and over a third (37.5%) were not satisfied with their level of training. Surprisingly, less than half (44.3%) reported that they would like consultation support with program implementation when needed. Lastly, teachers reported time as the largest barrier to implementation, and only one third reported that it would be very feasible to implement SEL one period per week.

The foregoing highlights that most teachers may perceive SEL as important but only half are implementing. Part of the problem may be because implementation barriers are preventing delivery. Although teachers recognize that SEL can improve students' academic success, training, time, and academic demands appear to be major barriers. As school SEL reforms occur and mandates are put in place, policy makers are advised to pay extra attention to assessing readiness and providing implementation support.

Evaluation of Implementation during Program Installation

Payton et al. (2008) state that a well-designed evaluation of implementation is an important feature of quality SEL programming. In order to improve program delivery, we should focus our efforts on evaluating those activities intended to support the delivery of a program. It is vital that education professionals expand their thinking to a systems-level approach (Durlak & DuPre, 2008; Fixsen et al., 2005). A systems-level approach is essential because convergent evidence demonstrates that implementation is affected by multiple interacting ecological factors. That is, their effects on each other are reciprocal and multiplicative. In order to effectively collect, analyze, and use data to examine process, integrity, and efficacy, all interacting ecological factors should be evaluated (Fitzpatrick, Sanders, & Worthen, 2004).

There are different types of evaluation, but formative evaluation in this context, refers to performance feedback on current barriers and drivers. Although the evaluation of program delivery during initial implementation is addressed most often in the literature, the use of assessment tools for understanding implementation barriers should be attended to prior to initial implementation (i.e., program installation) (Fixsen et al., 2005). Program installation is in essence, the planning stage.

The stages of the implementation process include exploration and adoption, program installation, initial implementation, full operation, innovation, and sustainability. With an emphasis on the installation stage, all stages are depicted in Figure 3. After an evidence-based SEL program has been explored and adopted, there are activities that need to be adequately addressed. Resources should be allocated, structural supports within the organization should be attended to, and staff should be prepared to deliver the SEL program during initial implementation.

In spite of the need to identify implementation barriers, there is a shortage of feasible and efficient assessment tools that can be used during program installation (Schulte et al., 2009). On the converse, there are a sufficient number of existing tools and instruments used to measure the social and emotional well-being of school youth, such as The *Social-Emotional Resilience and Assets Scales*, *Social Skills Improvement System Rating Scales*, and the *Behavioral and Emotional Rating Scales*. These tools are appropriate for school youth, monitor change over time, are scientifically sound, and practical to administer (Haggerty, Elgin, & Woolley, 2011). The said instruments are effective at evaluating treatment efficacy during implementation but they do not address systems-level implementation variables.

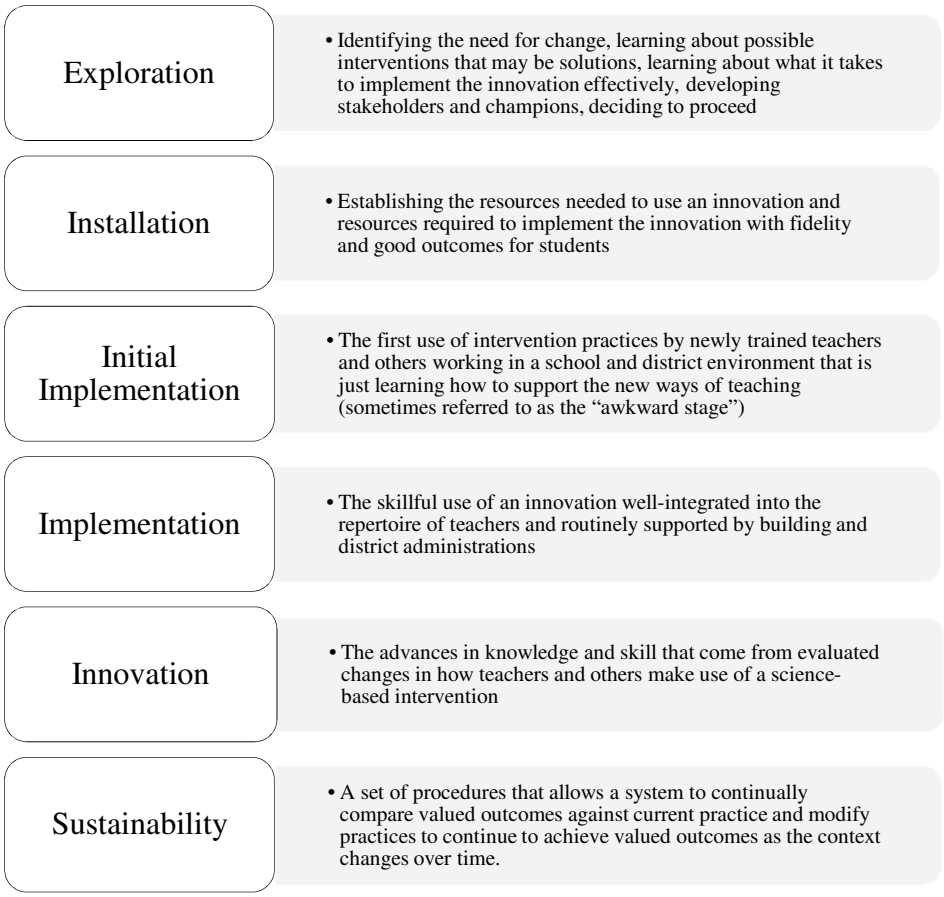

Figure 3. Descriptions of Implementation Stages. Adopted from Fixsen et al. (2005).

Denham, Ji, and Hamre (2010) however take it a step further by recognizing the importance of addressing characteristics of school contexts; for example, they provide a list of tools used to evaluate contexts that promote SEL. Aspects of the classroom that are important in this regard include effective classroom management, instructional support, healthy relationships, social and emotional skills instruction, opportunities for social and emotional skill practice, and prosocial norms and behaviors. Examples of such tools include the *Arnette Caregiver Interaction Scale*, *Assessment Profile For Early Childhood Programs*, and *Early Childhood Environment Rating Scale*. These instruments are effective at evaluating characteristics of school contexts affected by

program delivery and support treatment efficacy of SEL programs; but again, they do not address systems-level variables that support program installation.

Brackett et al. (2012) developed and validated one of the more promising measures, *the Teacher SEL Beliefs Scale*—specifically addressing teachers' beliefs about SEL and helping to determine school readiness for SEL programming. The authors' goal was to develop a measure that could help administrators gauge Provider Characteristics, such as attitude, preparedness, and openness. The Brackett et al. (2012) study resulted in three reliable subscales measuring teachers' comfort with teaching SEL, commitment to learning about SEL, and perceptions about whether their school culture supports SEL.

The *Teacher SEL Beliefs Scale* addresses teacher confidence because it is associated with quality implementation, improved classroom management during lessons, and positive attitudes. Secondly, commitment was addressed because it affects how well teachers integrate SEL into classroom and school routines. Third, teachers' belief about the importance of SEL for student success was addressed because teachers who consider the development of SEL competencies as an important variable for academic success are more likely to implement with quality and devote important resources such as time, and effectively integrate SEL into daily routines. Lastly, the perceptions about school culture and extent that staff supports new SEL innovations were addressed. For example, leadership, especially principal support, was a key variable in the measure because it is associated with sustainability and durability.

As indicated by Brackett et al. (2012), there are few published assessment tools measuring teachers' beliefs about SEL. The authors identified five studies: four made inferences about teachers' SEL beliefs and the fifth assessed teachers' attitudes about one

specific SEL intervention. None of the five studies resulted in a validated measure for practitioners to use as a readiness measure. The Teacher SEL Beliefs Scale is a promising measure in that it fills a much needed void in SEL implementation and is technical adequate.

In sum, there is a paucity of published SEL assessment tools for measuring Provider Characteristics and no assessment tools have been published that addresses Organizational Capacity. To build on the momentum of developing measures that help determine school readiness for SEL programming, a multilevel ecological perspective was employed for the *Schoolwide SEL Capacity Assessment* (SSCA). The SSCA is different from the Teacher SEL Beliefs Scale because it incorporates a systems-level approach by essentially evaluating multiple interacting ecological factors, specifically, Provider Characteristics and Organizational Capacity.

The proceeding section specifically addresses the aforementioned ecological factors and how they are incorporated into the SSCA, which also touches on how the SSCA is different from surveys like the *Teacher SEL Beliefs Scale* that only attend to Provider Characteristics (Brackett et al., 2012). For instance, Durlak and DuPre (2008) have provided a seminal article that incorporates a multilevel ecological perspective, which is helpful when developing readiness measures for program installation. We reviewed each variable and assessed how applicable they were for program installation, and then proceeded to include them in the item development for the SSCA.

A Multilevel Ecological Perspective Applied to the SSCA

Durlak and DuPre's (2008) determined that five categories of associated variables influence an intervention delivery system: Provider Characteristics; the Prevention

Delivery System; the Prevention Support System; Community Level Factors; and Characteristics of the Innovation. After reviewing Durlak and DuPre's (2008) meta-analysis, it was determined that for the purpose of expanding on the notion of a multilevel ecological perspective and developing a measure that helps determine school readiness for SEL programming, four of the foregoing components were collapsed into two evaluative terms: Provider Characteristics and Organizational Capacity. Provider Characteristics retained its original structure. Organizational Capacity included the remaining components—with the exception of some of the community level factors and none of the characteristics of the innovation factors. In brief, the decision not to include some of the components/factors was made because they were deemed less applicable for program installation. The components are subsequently described within their respective evaluative terms and the components not included in the SSCA are described under Other Components.

Organizational Capacity

The Prevention Delivery System is synonymous with Organizational Capacity (Durlak & DuPre, 2008). The variables identified include general organizational features, specific organizational practices and processes, and specific staffing considerations. Fixsen et al. (2005) addresses these variables in-depth, referring to them as implementation drivers, with an emphasis on the organization and leadership. Although the implementation drivers go beyond the prevention delivery system, their seminal work provides a well thought out and descriptive account of Organizational Capacity.

Organizations with a high level of Organizational Capacity encourage innovation and work systematically from adoption to sustainability, essentially cultivating quality

implementation efforts. As stressed by Kress and Elias (2006), a systems level perspective encourages educators to focus on the process and not just the content of SEL programming. The process emphasizes building SEL-focused learning communities that have a shared commitment and collaborate well together. Power et al. (2005) suggests we should shift from the prevailing hierarchical model of implementation (i.e., top-down) and build a partnership-based model (i.e., vertical), ultimately strengthening commitment and collaboration.

While ensuring the Prevention Delivery System is in good working order, we also need to focus on the Prevention Support System (Durlak & DuPre, 2008); this category is similar to the competency driver proposed by Fixsen et al. (2005). Both models address training, coaching, and technical assistance as a prime objective, and both recognize that they occur after the other categories or drivers have been secured. Payton et al. (2000) suggest that quality SEL programs should include training, and ongoing feedback to assist providers in successfully implementing programs with integrity. Investing energy into providing teacher assistance and feedback will increase the integrity of an intervention's implementation and ensure that an intervention has the intended effect (Gresham, 1989).

In general, the goals of training are to prepare providers, allow them to acquire knowledge and develop skills. Skill acquisition and mastery is accomplished through direct and explicit instruction, modeling, role-playing, rehearsal, and constructive feedback (Archer & Hughes, 2011). In contrast to training, technical assistance refers to support provided after initial implementation. The goal is to ensure providers maintain self-efficacy and continue to acquire the appropriate skills, teach with fidelity, and sustain

the practice over time. The supports provided during implementation are especially important for new staff, which has the potential to jeopardize schoolwide SEL reform efforts.

Fixsen et al. (2005) take the notion of technical assistance one step further by addressing both technical and adaptive supports. Technical support provides prescribed solutions to barriers based on a pre-determined program (e.g., Strong Start) or model (e.g., Positive Behavior Intervention Supports). For example, technical support may address acceptable adaptations to a model and may be solved by providing resources and materials. On the converse, adaptive support provides solutions around cultural and individual challenges (i.e., beliefs and values) that may arise as a result of changing behaviors and philosophies through the engagement of an SEL program. Both technical and adaptive supports encourage sustainability and durability of SEL programming.

The community context is also important to consider; both macro and micro level context variables (Bronfenbrenner, 1979). An important community level factor identified in the literature for the SSCA is funding (Durlak & DuPre, 2008). Funding tends to be a barrier for many schools, as they are expected to do more with less. However, financial support is needed for quality implementation, such as the purchase of evidence-based curriculums and continued professional development.

Provider Characteristics

Provider Characteristics can be broken down into four subordinate variables: need and potential benefits of the SEL programming, self-efficacy, and skill proficiency (Durlak & DuPre, 2008). Providers that recognize the need within their community context, believe in the efficacy of the program and themselves, and have the skillset to

implement the program with quality, would likely have improved readiness and ultimately, program delivery. With reference to Organizational Capacity, the relationship between providing adequate training and technical assistance, and teachers' Provider Characteristic attributes, is easily recognized—the two go hand-in-hand. When schools adequately focus on Organizational Capacity, they are in effect improving Provider Characteristics. This is one example of how components within a systems-level perspective work together to improve the whole system.

Other Components

Important Community Level Factors identified in the literature that were not included in the SSCA are the prevention research system, politics, and policy (Durlak & DuPre, 2008). The prevention research system helps educators identify evidence- and research-based best practice by disseminating research studies and professional recommendations. Politics and policy can have a major influence on school reform. For example, politics has had a major thrust in the education accountability movement, resulting in the No Child Left Behind Act of 2001, a federal law that places a great deal of emphasis on formative and summative evaluations (No Child Left Behind, 2002).

A spotlight has been cast on Characteristics of the Innovation, which consists of program adaptability and compatibility (Durlak & DuPre, 2008). For example, effective delivery of a program may rest explicitly on detailed lesson plans that are developmentally appropriate and sequenced according to skills hierarchies (Archer & Hughes, 2011; Becker, 2001). In this regard, adherence to the program content add a level of quality control and ensure children are exposed to critical elements, and appropriate flexibility and compatibility ensures it meets the children's and the

organizations contextual needs.

Adaptability has been defined as the degree to which a program can be modified to meet provider, organizational, and community needs; and is synonymous with program flexibility. A debate continues between fidelity and adaptability, with some championing deviations to help ensure compatibility, while others argue that fidelity is correlated with positive outcomes (Sanetti & Kratochwill, 2005). Compatibility addresses contextual fit or appropriateness and describes how well the program fit the organization's reform efforts, mission, objectives, and existing practices.

Sanetti and Kratochwill (2009) argue that "accepting flexibility within fidelity may be essential for implementation of evidence-based interventions to be sustained across the various contexts and cultures in which they are implemented'' (p. 452). The argument taken here is that adaptable programs tend to be implemented with more quality. The converse would be a program that must be delivered as is, which may hinder compatibility. However, adaptability has limits and deviating too far may jeopardize the impact of an evidence-based SEL program.

Although the aforementioned components are important to consider when implementing schoolwide SEL programs, schools cannot easily change them. When deciding which components to include in the SSCA, a decision was made to only include those that are malleable by local school efforts. The prevention research system, politics, and policy are difficult to change and require efforts beyond the intent of the SSCA.

Transtheoretical Model

Oftentimes administrators lack a clear understanding of teachers' current use of SEL, comfort, and willingness to change their current routine (Elder et al., 1999). To

better support teachers and ameliorate implementation barriers, the *Transtheoretical Model of Behavior Change* and its applications beyond modifying health behaviors will be discussed (Prochaska & DiClemente, 1982). The Transtheoretical Model is also known by the term Stages-of-Change (SOC) because change is a process involving progress through a series of stages: precontemplation (not ready), contemplation (getting ready), preparation (ready), action, and maintenance. The SOC is not a theory but a diagnostic framework, and it is proposed as a heuristic for understanding teachers' change regarding schoolwide SEL implementation efforts.

When developing a readiness measure for SEL program implementation, the expectation is that teachers would potentially implement a SEL program. Nonetheless, teachers may be at different stages regarding their willingness to change. Their Provider Characteristics, and similarly, their schools Organizational Capacity, may influence teachers' willingness to change. When teachers progress through the series of stages, it may be because they demonstrate more Provider Characteristic attributes, and the same could be true for their schools' Organizational Capacity. As schools reduce barriers and increase drivers, it may facilitate teachers' progression through the stages, going from not ready to taking action. Teachers SOC may reflect present levels of systems-level factors, which interact to either augment or hinder their progress through the stages.

The current research project included an item that captured current practices of the participants by framing it within the SOC model, with the intent to identify teachers' current level of universal SEL program use and current stage within the change cycle. It is expected that teachers who identify with higher levels within the SOC model, will also endorse higher levels of Provider Characteristics and Organizational Capacity. As

schools evaluate their readiness to implement with the SSCA, they may be able to expect teachers SOC to mirror their levels within the two components. This is in essence, the heart of a systems-level approach—evaluating multiple interacting ecological factors. The SOC Model helps frame schools understanding of the importance of evaluation during program installation by highlighting that teachers' progress through change can be influenced by improving multiple interacting ecological factors.

Conclusion

Although there is a preponderance of evidence to suggest that our youth are faced with many risk factors, prevention programming—especially SEL—has the potential to introduce protective mechanisms that may ameliorate mental health problems and help children lead more successful lives—socially, emotionally, and academically. The foregoing discussion demonstrates that poor implementation can moderate SEL outcomes, suggesting that evidence-based programs are necessary but ultimately, insufficient.

In order to ensure quality implementation, sustainability, and durability of schoolwide SEL programming, we should focus our efforts on evaluating those activities intended to support the delivery of a program. In doing so, not just program delivery but treatment efficacy may also be improved. Although there is a shortage of feasible and efficient assessment tools for evaluating implementation barriers during program installation, Durlak and DuPre's (2008) multilevel ecological perspective provides a framework for developing a research-based assessment measure that goes beyond the traditional focus of Provider Characteristics and emphasizes system-level variables such as Organizational Capacity.

The development of the SSCA and the current study is unique in that it builds on current research and reflects best practice by incorporating concepts of prevention delivery and uniquely focuses on multiple interacting ecological factors. The study aims to develop a technically sound measurement for school readiness in SEL programming, focusing specifically on Organizational Capacity and Provider Characteristics. In sum, the purpose of the current study is to extend the current SEL implementation assessment literature by developing a new instrument that targets all professional school staff and evaluates their perceived need, benefits, self-efficacy, and skill proficiency for SEL programming, and school staff belief about the school's universal prevention delivery system, general organizational factors, practices/processes, staffing, and the support system.

CHAPTER 3

METHOD

Objectives and Research Questions

The objective of the current study is twofold: (1) to extend the Social and Emotional Learning (SEL) implementation assessment literature by developing a new needs assessment rating scale designed to measure SEL implementation barriers (*School SEL Capacity Assessment* [SSCA]); and (2) make preliminary inferences regarding the current state of SEL implementation. In order to satisfy the first objective, we evaluated the psychometric quality of the SSCA. Specifically, this process involved the investigation of the utility (appropriateness) of the proposed rating scale, the fit of the measurement model, and evidence in support of the validity of the scale

In order to satisfy the second objective, we illuminated teacher perceptions of SEL implementation barriers by evaluating differences, the equivalent of addressing mean differences across groups (e.g., analysis of variance between groups). Specifically, the current study answered the following four research questions:

1. Are there differences across participants' "stage of change" (Transtheoretical Model; Prochaska & DiClemente, 1982)?
2. Are there differences across participants with different years of professional experience?
3. Are there differences across grade level?
4. Are there differences across social economic status?

Setting and Participants

This study was conducted in several school districts in the Midwest and Northeast United States. The participants for the present study were teaching staff from the said districts. Attempts were made to recruit participants based on a sample, representative of the general kindergarten through twelfth grade population of teachers (Snyder & Dillow, 2012). Based on suggestions provided by Linacre (1994), in order to have a sample size large enough to obtain usefully stable item calibrations (±1/2 logit), a minimum sample size ranging from 64 to 144 was needed; our final sample size included 87 participants. Refer to Table 1for the participant demographics.

Recruitment Methods

Prior to initiating the recruitment process, the study outlined below was approved by the Institutional Review Board of the Human Research Protection Office at the University of Massachusetts Amherst. Upon approval, the principal investigator contacted school districts, teachers, and the *SEL Alliance for Massachusetts*, a group of individuals who have made it their mission to champion and introduce SEL into Massachusetts' schools. Prior to data collection, all participants were provided a summary of the research proposal and a written statement of informed consent, privacy, and confidentiality.

The objective was to receive permission from school districts and then recruit schools. Specifically, district administrators, principals, special education administrators, and other contact persons at the target districts or schools were contacted. However, because of low turnout, we used word-of-mouth and email recruiting as an alternative method.

Table 1

Demographics of Participant

Variables		n	%
Participants		87	100%
Gender			
	Female	77	89%
	Male	9	10%
Race/Ethnicity			
	White	78	89%
	Black or African American	2	2%
	Asian	1	1%
	Latino/Hispanic	2	2%
	American Indian or Alaskan Native	1	1%
	Multiracial	1	1%
	Unknown	4	4%
Years of Professional Experience	0-1	1	1%
	2-4	8	9%
	5-7	11	13%
	8-10	10	12%
	10-15	20	24%
	>15	35	40%
	Unknown	1	1%
Grade Level Currently Working			
	Elementary	70	81%
	Middle	5	6%
	High	3	3%
	District-wide	6	7%
	Unknown	2	3%
Social Economic Status			
	High	25	29%
	Middle	42	48%
	Low	17	20%
	Unknown	3	3%
School Size	Small	55	63%
	Average	22	25%
	Large	9	10%
	Unknown	1	2%

As for the latter, the *SEL Alliance for Massachusetts* (SAM) and Michigan's *Integrated Behavior and Learning Support Initiative* (MiBLSi) were contacted. SAM agreed to send out a message via their listserv to over 6,000 members in the state of Massachusetts, which included teachers, SEL providers, hospitals, associations, businesses and nonprofits. MiBLSi agreed to send out a message via their listserv to over 1,000 members in the state of Michigan, which included teachers, social workers, psychologist, and school administrative staff. Although both organizations included professions other than teachers as part of their membership, the email explicitly stated teachers as the primary population of interest.

Procedure

Instrument Development

Given that the constructive approach to measurement proposed by Wilson (2005) is useful for instrument development and supports the evaluative process, we developed and evaluated the Schoolwide SEL Capacity Assessment (SSCA) according to Wilson's construct-modeling framework. The instrument and item pool development included construct modeling, item-design, outcome space, and the measurement model.

We began by conceptualizing the construct and designing a construct map—an approach that is useful for measuring a latent variable so that it may be observable (Wilson, 2005). The construct map is a visual representation of the construct and includes several important features: (a) the essential components that make up the construct; (b) a definition for the components of the construct; and (c) the idea that there is a qualitative order of levels or continuum inherent in the construct.

The conceptualization of the construct was investigated through a literature review. After an extensive review, we determined that Buchanan et al. (2009) survey study examining teachers' knowledge, perceptions, and practices regarding SEL in the classroom and Durlak and DuPre's (2008) meta-analysis on factors affecting the implementation process provided a framework for the SEL implementation barriers construct. The survey study by Buchanan et al. (2009) was useful in helping set the stage for understanding the types of questions and demographic information that were of interest. Durlak and DuPre's (2008) findings were useful in understanding the range of common SEL implementation barriers to include in the SSCA.

After reviewing Durlak and DuPre's (2008) meta-analysis, we determined that for the purpose of the instrument, the construct map would be developed according to two construct components that would encompass many of the five factors identified by the authors. Durlak and DuPre's (2008) findings from 81 reports indicated that there are five factors moderating and mediating implementation quality: community level factors, provider characteristics, characteristics of the innovation, prevention delivery system, and the prevention support system. With the exception of the characteristics of the innovation factor and some aspects of the community level factor, all other factors were used for the development of the SSCA. Characteristics of the innovation, which addresses compatibility and adaptability of the intervention, were not included because they were considered too distal for immediate change. Within the community level factor, issues related to prevention and research, politics, and policy, were also deemed inappropriate, but issues around training were included because it is malleable to local reform efforts.

For the purpose of the SSCA, Provider Characteristics and Organizational Capacity are used as evaluative terms, which define the SEL implementation barriers construct. Provider Characteristics retained its original structure. It is defined as a person's perceived need, benefits, self-efficacy, and skill proficiency for SEL programming. Organizational Capacity includes training, the prevention delivery system, and the prevention support system. We define it as a person's belief about the school's universal prevention and intervention delivery system, general organizational factors, practices/processes, staffing, and the support system. Figure 4, the construct map, visually represents the final conceptualization of the construct.

Using the construct map as a guide, item-development and outcome space were established as a means to create a concrete manifestation of the instrument. From there, we developed requirements or descriptive components for the items (Table 2). Although there are many components of an instrument, our item development was based on practical constraints, historical precedence, and an arbitrary basis (Wilson, 2005).

Table 2

Descriptive Components: Item Classification Requirements

Must Include the following
Take a contextual or multilevel ecological approach.
Feature a brief description of social emotional learning (SEL) and provide examples of SEL programming/curriculums.
Relate to the literature on SEL and implementation science.
Relate to the process of SEL programming rather than content—with a focus on the needs and readiness of the school.
Relate to the feasibility of implementing programs under real world conditions.
Include demographic information.
Should take not longer than 20 minutes to complete

Construct	Provider Characteristics	Organizational Capacity
Definition	A person's perceived need, benefits, self-efficacy, and skill proficiency for SEL programming.	A person's belief about the school's universal prevention delivery system, general organizational factors, practices/ processes, staffing, and the support system.
Continuum (Direction of Increased Attributes) ↑	Provider believes that universal SEL is important for student development and allocates school time to implement SEL programming. Provider adheres to curriculum components and implements each lesson with quality.	Provider believes systems are in place to support implementation efforts, participates in professional development, colleagues collaborate and integrate universal SEL programming into other school-wide initiatives.
↕	Provider believes it is a school's responsibility but does not believe it is the teacher's role. Implementation should be provided to those who are at risk from a mental health worker. Provider does not feel they have the skill proficiency to implement to all students but may provide ancillary support when the need arises.	Provider believes the school supports them occasionally but may fail to provide enough financial support or the adequate amount of training. The school principal may make SEL programming an objective for all teachers but may fail to schedule regular team meetings and allocate time throughout the day.
↓ (Direction of Increased Attributes)	Provider believes that it is not the school's role to teach students SEL, is concerned parents do not support programming, does not allocate time to implement an SEL curriculum, and may be concerned about their ability/skills to teach SEL.	Provider does not believe the school supports the implementation of universal SEL. Systems are not in place to provide training and ongoing support. There is no support from leadership and no internal advocate to champion SEL efforts.

Figure 4. Schoolwide SEL Capacity Assessment Construct Map

Next, we identified which item typology and outcome space was the best fit for the purpose of the instrument. We concluded that a standardized fixed-response format with polytomous choice was the most appropriate because it is relatively quick and efficient to administer and requires participants to respond in a manageable, finite number of categories, which brings order and consistency to outcome data (Wilson, 2005). In order to calibrate a construct into a quantified continuum along which persons can be positioned, the response categories in the outcome space must be ordered in some way, with differing levels of more and less of the construct. In Likert-style response items, the order is implicit (Wilson, 2005). For those reasons, we employed the following Likert scale: strongly agree, agree, somewhat agree, somewhat disagree, disagree, and strongly disagree.

We concluded by adding a demographics section. We added the following to the item pool: gender; race; years of professional experience; highest level of education completed; grade level at which participants currently work; job position; social economic status of school system; school location; school population size; current use of SEL curricula; who is responsible for implementing universal SEL programming; primary setting SEL programming is being taught; and current practices.

Regarding the latter, we asked teachers to identify their current practices by endorsing one of five stages of implementation. This item was used to answer question 1 of objective 2. The stages were designed to reflect the Stages-of-Change (SOC) Model proposed by Prochaska and DiClemente (1982). There are five stages within the SOC model, ranging from Precontemplation to Maintenance. The following displays how the categories were designed to mirror the five stages:

1. Precontemplation (Not Ready): “I never plan to use an SEL program in this school.”
2. Contemplation (Getting Ready): “I have thought about using an SEL program, but have not taken any steps to start a program.”
3. Preparation (Ready): “I will use an SEL program soon.”
4. Action: “I am using an SEL program, but not on a regular basis.”
5. Maintenance: “I am using an SEL program but only use certain components”; and “I am using an SEL program exactly as described in the program materials/instructions.”

Measurement Model

The measurement model is described by Wilson (2005) as a way to connect the items, response categories, and scored outcomes back to the construct. A measurement model helps identify the relationship between items and the construct, and scored outcomes with the construct. In a broad sense, in order for the SSCA to be useful, it must provide information that allows for valid inferences and decisions. In order to accomplish this, an Exploratory Factor Analysis (EFA) served to investigate the assumption of unidimensionality and the psychometric quality was investigated via the Rasch Rating Scale model (RSM).

Because the RSM is a unidimensional measurement model, it is important to investigate whether or not this assumption is met (de Ayala, 2009). EFA is exploratory: no prior specifications are made in relation to the number of latent factors or the pattern of their relationships (Brown, 2006). An EFA was used in lieu of other variable reduction methods, particularly a principal component analysis (PCA), because it helps to

uncover the underlying structure and increase power. Conducting an EFA allows for the evaluation of this assumption and also provides a sense of the structure of the data—regardless of the unidimensionality assumption.

A principal axis factoring method was used to extract the factors. This technique has the advantage of being free of distributional assumptions and less prone to improper solutions than other methods, such as the maximum likelihood method (Brown, 2006). The disadvantage is that there is no information about model fit. At this point, it is believed that it is better to relax any constraint on the data, that is, not to assume any number of factors.

Once unidimensionality had been established, we proceeded by analyzing the psychometric quality of the SSCA by using the RSM. The RSM has been used extensively in social science research and scale development. As an example, Prieto and Delgado (2007) used it to measure math anxiety, and Oliveira, Fernandes, and Sisto (2014) used it to analyze the School Anxiety Inventory.

Although many different measurement theories exist, Rasch modeling has the potential to offer substantial advantages over analysis based on classical measurement theory for developing instruments, such as confirmatory factor analysis (CFA). Unlike the CFA model, the Rasch Model (a) allows for person and item parameters to be estimated independently of each other and (b) includes no item discrimination parameter (or item loadings) as it is assumed to be equal across all items. As Linacre (2010) illustrates, there are many reasons for using the Rasch model over others such as the CFA model:

1. The person attributes and the item difficulties are additive (linear) measures on the same latent variable, so that the spacing between the measures is meaningful (unlike person raw scores or item *p*-values).
2. Rasch estimates are robust against missing data.
3. The hierarchy of person abilities supports predictive validity: "Do the person measures have the meaning that we intended them to have?"
4. The hierarchy of item difficulties supports construct validity: "Are we measuring what we intended to measure?"
5. The precision of each ability and difficulty is known: "How well does this test classify our person sample?"
6. The expected performance by any person on any item can be inferred from their attributes and item difficulties: "What is next?"
7. The validity of the patterns of responses (item fit or person fit) can be investigated: "Are the responses meaningful?"

In its simplest form, assuming dichotomous scoring, the Rasch model asserts that the probability of a person getting an item correct (or answering in the affirmative) is based only on the person's ability and the difficulty of item. This probability can be expressed as:

$$\Pr(x_{ni} = 1|\theta_n\delta_i) =, \frac{\exp(\theta_n - \delta_i)}{[1 + \exp(\theta_n - \delta_i)]}$$

where θn represents the ability of person *n* and δ_i represents the difficulty of item *i* (Rasch, 1960/1980). The RSM, an extension of the Rasch model, allows for the analysis of ratings in two or more ordered categories, such as the Likert-style response categories used in the SSCA (Wright & Masters, 1982).

To determine where both items and teachers are located on the latent trait continuum, and consequently, if this placement is reasonable, we refer to the logit scale. A logit scale is an interval scale in which the unit intervals between the locations on the variable map have a uniform value or meaning (Bond & Fox, 2001). The logit scale, theoretically, ranges between $-\infty$ to ∞ logits and mirrors the underlying latent construct. Within SEL implementation, $-\infty$ logits represent barriers that are easiest to endorse, and ∞ logits represent barriers that are more difficult endorse. Logit measures can be estimated for any latent construct (e.g., teachers characteristics).

To determine if the items and persons are sufficiently well separated along the logit scale, we refer to the reliability of separation. This coefficient provides a measure of the degree to which the 'elements' within a facet (i.e., the individual constructs, teachers, or items) are separated. This reliability coefficient represents the ratio of the true score variance to the observed score variance (Wright & Masters, 1982) and is defined as follows:

$$R = \frac{SD^2 - MSE}{SD^2}$$

where SD^2 is the observed variance of the element difficulty for a facet and MSE is the mean square calibration error for each element within a facet. Larger differences between the constructs, or elements, within a facet will yield higher reliability of separation coefficients. In addition, the item separation index defined as

$$Gi = \frac{SAi}{SEi}$$

where SAi is the adjusted standard deviation and SEi is the root mean square calibration error.

To address both item fit and person fit (i.e., consistency), we refer to the fit statistics: mean square, outfit, and infit statistics. The mean square is the unstandardized form of the fit statistic and represents the average value of the squared residuals for the characteristic. Outfit mean square statistics (u_i) are the unweighted mean square residual differences between observed values and expected values (Wright & Masters, 1982). Outfit statistics are useful for diagnosing item misfit to the measurement model. The outfit mean square statistic is defined as:

$$ui = \frac{\sum_{n=1}^{N} z_{ni}^2}{N}$$

Outfit mean square statistics greater than 1.2 may indicate inconsistent responses by teachers or items which, for various reasons, illicit inconsistent responses. Outfit mean square statistics greater than 2.0 indicate a great deal of unexplained variance providing more misinformation than information. A disadvantage of the u_i statistic is that outliers, such as only one or two teachers making surprising responses to one or two items, can heavily influence it. Infit mean square (v_i) statistics are similar to outfit mean square statistics differing only in that they are weighted and less influenced by outliers. The infit mean square is calculated as follows:

$$v1 = \frac{\sum_{n=1}^{N} W_{ni} Z_{ni}^2}{\sum_{n=1}^{N} W_{ni}}$$

Data Analytic Plan

Data Collection Materials

We administered the School SEL Capacity Assessment (SSCA) using the online survey tool, surveymonkey.com. Survey Monkey is a web-based interface for creating and publishing custom web surveys, and then viewing the results graphically. The web-

based tool allows researchers to design surveys according to various input questions, Likert-style response categories with various scaling methods and response formats, and open-ended questions. After inputting items and specifications, the survey was made available to participants via a web link.

Data Analysis for Objective 1

We analyzed the data in two stages. Within the first stage, using the Statistical Package for the Social Sciences 19 (SPSS), a predictive analytics software program, an exploratory factor analysis (EFA) of the data were conducted to establish the unidimensionality of the SSCA. Within the second stage, FACETS 3.62 (Linacre, 2007) was used to estimate all parameters and fit statistics using the RSM.

Unidimensionality

The unidimensionality assumption states that the observations of the items are the manifestation of a single variable. In our case, unidimensionality implies a single dominant component that explains the variation and co-variation of the items on the SSCA. Variation refers to the variance of the responses to a particular item. Covariation refers to the covariance between the responses of different items.

Initially, all the items were treated as part of the same construct. After applying the EFA to the data, we determined that there were two dominant factors: Organizational Capacity and Provider Characteristics. Each factor was treated as an individual unidimensional scale.

Psychometric Quality Using the Rasch Rating Scale Model

Conceptually, measurement begins with a continuum on which people can be placed with respect to some latent trait or construct (Wright & Masters, 1982). The

measurement process allows people to be measured by the items that define the latent trait underlying the continuum. According to Wright and Masters (1982), the requirements for accurate measurement are:

- The reduction of experiences to a one-dimensional abstraction.
- More or less comparisons among persons and items.
- The idea of linear magnitude inherent in positioning objects along a line.
- A unit determined by a process, which can be repeated without modification over the range of the variable. (p. 3)

As proposed by Wright and Masters (1982, pp. 90-91), in order to evaluate the five requirements for accurate measurement, we sought to address the following:

1. Have we succeeded in defining a discernible line of increasing intensity?
2. Is item placement along this line reasonable?
3. Do the items work together to define a single variable (i.e., consistency)?
4. Have we succeeded in separating persons along the line defined by the items?
5. How valid is each person's measure?

The first three questions are used to help evaluate the instrument's items ability to work together to define a meaningful variable. The last two questions address the extent to which teachers are separated along the same line and the validity of their individual measures. To determine where both items and teachers are located on the latent trait continuum, and consequently, if this placement is reasonable, we refer to the logit scale.

Rating Scale Utility

We used a standardized fixed-response format characterized by participant self-report with polytomous choice for the SSCA. The SSCA employed six Likert response

categories: strongly agree, agree, somewhat agree, somewhat disagree, disagree, and strongly disagree. Rating scales with polytomous choice are used as a means to gather a breadth of information that increases measurement accuracy and precision beyond simple dichotomous response categories (e.g., agree, disagree).

In order to determine if the six response categories were the most appropriate, we assessed the rating scale utility, and then the categories were optimized to improve effectiveness. The objective was to satisfy all eight guidelines (Table 3) proposed by Linacre (1999) and if necessary, response categories were recategorized to insure adherence. For analytical purposes, the best categorization cannot be observed in raw data and therefore, the guidelines are suggested to improve the overall measurement quality and ensure step calibrations are accurately calculated (Linacre, 1999).

Table 3

Rating Scale Utility Guidelines

Guideline	
1	At least ten observations of each Likert response category
2	Regular observation distribution
3	Average measures advance monotonically with category
4	Outfit-mean squares less than 2.0
5	Step calibrations advance
6	Ratings imply measures and measures imply ratings
7	Step difficulties advance by at least 1.0 logits
8	Step difficulties advance by less than 5.0 logits

The guidelines address category frequency, ordering, the ratings inferential value, and the quality of the scale from a statistical perspective (Linacre, 2002). Guideline 1, at least ten observations of each Likert response category, insures that each response category is stable and precisely estimated. When less than 10 observations are observed in each response category, the estimated scale structure (i.e., step calibrations) can be

imprecise. For example, if participants used the category, strongly agree, less frequently than 10 times, one remedy would be to combine it with the adjacent category, agree, to obtain more stable step calibrations.

Guideline 2, a regular observation distribution, highlights the need for an expected distribution of observations across response categories. When irregularities exist, it may suggest that respondents are using the response categories irregularly and may have departed from an accepted standard. For example, if participants used the category, somewhat agree, infrequently, but used the adjacent category, agree, more frequently, it suggest that the higher rating may not reflect more of that attribute.

Guideline 3, average measures advance monotonically with category, reflects the need for the response categories to represent a sequence and increase consistently. Response categories must represent its intended meaning and higher ratings should represent more of that particular attribute. The categories should be ordinal and not merely nominal.

Guideline 4, an outfit-mean squares less than 2.0, puts a spotlight on the stochastic nature of the Rasch model. An essential conceptual feature of rating scale design is that increasing amounts of the underlying variable in a respondent correspond to increasing probabilities of the respondent being observed in higher categories of the rating scale. It ensures acceptable levels of fit. A reasonable uniform level of randomness is to be expected in the distribution but excessive randomness creates more noise and less predictability. In simple terms, mean-square fit statistics greater than 2.0 suggest unexplained randomness and unexpected use of that category.

Guideline 5, step calibrations advance, focuses on the scale's inferential value and conclusions that may be reached. It states that the step calibrations should advance monotonically with the response categories. Step calibrations refer to the scaling of the items and people; the smaller proportion of high affirmatives, the higher the difficulty of an item and hence the higher the item's scale location. When the step calibrations do not advance monotonically with the response categories, it indicates that the category represents a narrow segment of the latent variable or respondents have a difficult time defining the category. For example, it is essential that participants with more SEL implementation attributes choose higher ratings, such as agree; and participants with less attributes choose lower ratings, such as disagree.

Guideline 6, ratings imply measures and measures imply ratings, depicts the relationship between items and average expected ratings. The way a participant responds to an item should give an indication of what to expect on the remaining items. Consequently, the response should guide our inferences and allow predictions to made regarding where the person is positioned on the latent variable. For example, a participant's rating on one item should, generally, allow us to infer how that participant will respond to other items on the SSCA. Consequently, a participant that has more of the attributes should respond to the items with higher ratings, and vice versa.

Guideline 7 and 8, step difficulties advance by at least 1.0 logits but less than 5.0 logits, expresses the need to communicate location on the rating scale. For practical reasons, a certain location implies below or above a location already reached or not yet reached. By having step difficulties advance by at least 1.0 logits, it provides a predictable degree of space in-between response categories and ensures they are

conceptually different. Step difficulties that advance by less than 5.0 logits reduces the range of performance, so that its category boundaries are not too far apart, creating more precision.

For visual and comparison sake, Figure 5 displays probability curves for a well-behaved rating scale and Figure 6 displays an irregular rating scale. In sum, in order to make effective inferences, the conclusions must be based on properly functioning data. The eight guidelines are a means to improve the functioning of rating scale categories, resulting in better measurement accuracy and precision.

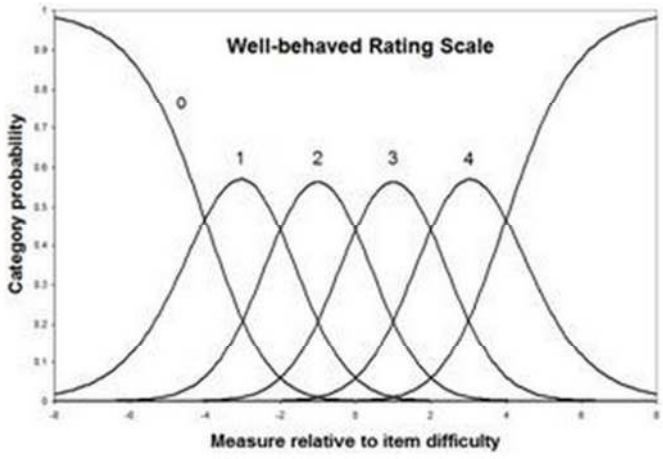

Figure 5. Probability Curves for a Well-behaved Rating Scale

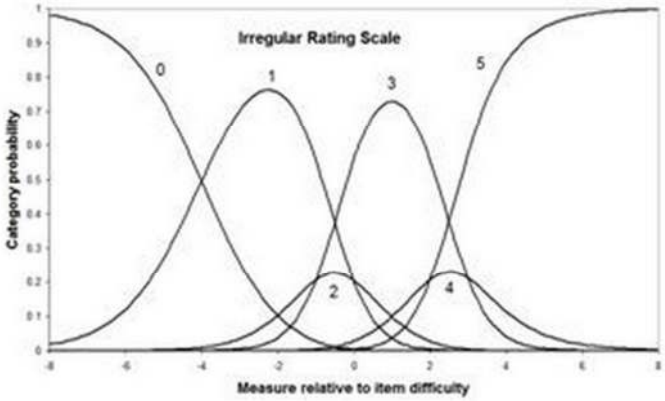

Figure 6. Probability Curves for an Irregular Rating Scale

Data Analysis for Objective 2

Findings from the Rasch analysis and the SEL implementation barriers construct employed for this study guided data analysis for the second objective. The research questions, which involve making preliminary inferences regarding the current state of schools, are presented with an explanation of the procedure used. For analytical purposes, survey response hierarchy differences can also be observed in raw data. To demonstrate that the differences are statistically significant, the reliability of separation, strata, and the fixed-all same chi-square test was used to determine if there are differences in identified subgroups (Wright & Masters, 1982).

Empirically, separation refers to the different levels or clusters (i.e., subgroups) in the observed distribution among facets. It shows how many subgroups within a facet are spread out across the construct. The reliability of separation is most equivalent to KR-20 and Cronbach's alpha in classical measurement. There is no set criterion for reliability coefficients in that reliability depends upon the population and instrument; however, an acceptable standard for items is ≥ 0.9 and persons may range from 0.6 to 0.8 depending on sample size (Linacre, 2014). The reliability of item separation coefficient is calculated with the following formula:

$$R = \frac{SD^2 - MSE}{SD^2}$$

where SD^2 is the observed variance of the element difficulty for a facet and MSE is the mean square calibration error for each element within a facet.

The fixed-all same chi-square test is the equivalent of an omnibus F-test. It is a test that analyzes whether or not every element within a facet is the same (e.g., all items are of equivalent difficulty). The test should be significant in order to reject the null that

the elements do not differ. The fixed-all same chi-square test coefficient is calculated with the following formula:

$$G^2 = 2\,\Sigma x_{ij}\,(\ln x_{ij} - \ln m_{ij})$$

where ln x_{ij} is the natural log of the observed frequency and ln m_{ij} is the natural log of the model or expected frequency.

In order to investigate any statistically significant group differences for each facet, the standardized difference test is calculated with the following formula:

$$D = \frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\left[se(\hat{\theta}_1)\right]^2 + \left[se(\hat{\theta}_2)\right]^2}}$$

The standardized difference test coefficient is a z-distribution. The z-distribution is a probability density function of a normal distribution that have a mean equal to zero and a standard deviation equal to one. In sum, the standardized difference test was used to test the difference of endorsability in groups when the standard deviations are estimated from the sample data.

Validity

In order to gather evidence for the validity of the SSCA, we collected qualitative and quantitative data. Validity evidence is framed according to the sources of validity outlined in the Standards for Psychological and Educational Testing (AERA/APA/NCME, 2014). The process of creating validity arguments was founded on various aspects of the construct-modeling process, an expert evaluation, and think aloud protocols; and the RSM was used to provide psychometric support for the instrument. To further explore validity arguments for SSCA, objective 2—illuminating teacher

perceptions of SEL implementation barriers by evaluating differences—was analyzed to ensure that the data outcomes coincided with our expectations.

Test Content

Test content refers to "the themes, wording, and format of items, tasks, or questions on a test, as well as the guidelines for procedures regarding administration and scoring" (AERA/APA/NCME, 2014, p.11). We collected evidence of test content from previous studies through the construct-modeling process (development of the construct map) and an expert evaluation. We examined the evidence related to the appropriateness of the SSCA content in relation to SEL implementation barriers. The SSCA content refers to the relevance of the construct components (i.e., Provider Characteristics and Organizational Capacity), the relevance of the items related to the construct, how well the items fit the construct factors, and the wording and format of the items.

Efforts were made to receive feedback from experts in the field of social and emotional learning and implementation science; participants were asked to complete the Content Validation: Schoolwide SEL Capacity Assessment. An email was sent out to prospective evaluators explaining the aim of the project and what would be expected of them. There were a total of 10 evaluators; their demographics are displayed in Table 4. As evaluators proceeded, they were informed to log into a URL address associated with a surveymonkey.com account.

Table 4

Demographics of Evaluators

Variables		n	%
Participants		10	100%
Gender			
	Female	4	40%
	Male	3	30%
	Unknown	3	30%
Race/Ethnicity			
	White	6	60%
	Unknown	4	40%
Years of Professional Experience	5-7	1	10%
	8-10	1	10%
	10-15	3	30%
	>15	2	20%
	Unknown	4	40%
Position			
	School Psychologist	3	30%
	Administrator	1	10%
	University Researcher	2	20%
	Unknown	4	40%
Social Economic Status of Schools			
	High	2	20%
	Middle	3	30%
	Low	1	10%
	Unknown	4	40%
Territory			
	City	1	10%
	Suburb	3	30%
	Rural	1	10%
	Unknown	5	50%
School Size	Small	1	10%
	Average	4	40%
	Large	1	10%
	Unknown	4	40%

Evaluators were chosen because of their experience and expertise with SEL and implementation science. Evaluators included university researchers with publications

related to the advancements of SEL and implementation science; leaders of the Collaborative for Academic, Social, and Emotional Learning; and practicing school psychologists with expertise in implementing SEL programs in schools. Evaluators were asked to use their own opinion as the basis for their answer in deciding whether the proposed items for the two construct components measure what they were intended to measure.

The evaluators were familiarized with each component, as well as their definitions, and presented example questions to demonstrate what was expected of them. The evaluation tasks consisted of several steps including a rating and comments section. In the rating section, evaluators were asked to indicate the factor that best fits each statement, how strongly they felt about their placement, and how relevant they believed each item to be for that factor. Lastly, they were asked to provide their feedback on item revision and scale improvement.

Response Process

Response process refers to participants "performance strategies or responses to particular items," and "can contribute to questions about differences in meaning or interpretation of test scores across relevant subgroups" (AERA/APA/NCME, 2014, p.12). Evidence of response process was gathered from previous studies through think aloud protocols (TAP). Evidence related to the response process was based on the analysis of individual responses and documentation of participants' performance strategies to items. The relevancy of the evidence is gauged according to the extent to which participant responses are aligned with the intended interpretation of scores. The purpose of this procedure was to observe participant performance strategies and responses such as how

they approached, analyzed, thought through, and felt about particular items. This enabled investigators to gather data around what evidence is present for response process validity and improved understanding of the construct (Smith & Smith, 2007; Wilson, 2005).

Participants included a total of 6 teachers with at least 3 years of experience working within Massachusetts' public school districts. As indicated by Someren, Barnard, and Sandberg (1994), a small number of participants, ranging from only 4-5 subjects are sufficient for determining about 80% of usability problems. In order to gather data supporting response process validity, the TAP included standardized instructions, a practice phase, a transcription of the protocol, and a thorough analysis and review of the combined data. Instructions briefly explained that the current study was interested in how people interpret items within the SSCA and that they were expected to read the items out loud and say whatever came to mind as they attempted to answer the items.

The practice phase included two parts. First, the administrator modeled an example question and answered as if they were participating in the TAP. Secondly, the participant proceeded by practicing with an example item and the administrator evaluated their attempt and if needed, they provided suggestions. For example, if participants failed to think aloud as they read through the item, we reminded them to do so and provided another model of what our expectations were for the process. The investigator recorded the session and thereafter, transcribed, analyzed, and reviewed the data. From there, items were adjusted if the data illustrated that the participants' interpretation of the items are not consistent with intended interpretation.

Internal structure

Internal structure indicates "the degree to which the relationship among test items and test components conform to the construct on which proposed test items are based" (AERA/APA/NCME, 2014, p.13). Because the conceptual framework of the SSCA posited unidimensionality, evidence of internal structure refers to "the extent to which item interrelationships bear out the presumptions of the framework" (AERA/APA/NCME, 2014, p.13). In this case, the specific type of analysis used was an EFA and the RSM.

The objective of the EFA is to test the assumption of unidimensionality. The fundamental aim is to explain the variation and covariation of the answers to the items using fewer variables—or factors—that are interpreted as latent common characteristics (Brown, 2006; Hardle & Simar, 2003). EFA helps to identify underlying latent factors that explain the responses to the items.

Because the RSM is a unidimensional latent trait model, it is based on the principle that data follow a hierarchy on a single continuum of interest (Rasch, 1960/1980). Rasch measurement provides a model in which the performance of participants and item difficulty are compared in terms of fitting a unidimensional continuum (Wright & Masters, 1982). The objective is to transform ordinal raw scores into interval-scaled measures for the purpose of comparing equal units of distances between points on the continuum.

Relations to Other Variables

Relations to Other Variables include "measures of some criteria that the test is expected to predict, as well as relationships to other tests hypothesized to measure the same construct, and tests measuring related or different constructs" (AERA/APA/NCME, 2014, p.13). Although this study is not gathering evidence based on relationships to other variables, future studies could address convergent and discriminative evidence, test-criterion relationships, and validity generalizability. For example, we could explore convergent evidence between scores on the SSCA and the Teacher SEL Beliefs Scale (Brackett et al., 2012).

Consequence of Testing

Evidence regarding the consequences of testing addresses "the intended and unintended consequences of test use" (AERA/APA/NCME, 2014, p.16). At this point, evidence is not being gathered because the SSCA is still undergoing revisions based on the current validity study. However, future studies could explore sources of invalidity like underrepresentation or construct-irrelevant components. For example, if the items in the SSCA addressed barriers that were outside the control of schools (such as federal policy) and district administrators were holding schools accountable for improving their readiness, then the inferences based off of the results would be inappropriate.

CHAPTER 4

RESULTS

Introduction

This is an exploratory study using quantitative research methods, with two objectives: (1) to extend the SEL implementation assessment literature by developing a new rating scale designed to measure SEL implementation barriers (*School SEL Capacity Assessment* [SSCA]); and (2) make preliminary inferences regarding the current state of SEL implementation. In order to satisfy the first objective, we evaluated the psychometric quality of the SSCA using the Rating Scale model (Wright & Masters, 1982). In order to satisfy objective 2, we illuminated teacher perceptions of SEL implementation barriers by answering four questions:

1. Are there differences in SEL implementation across participants' "stage of change" (Transtheoretical Model; Prochaska & DiClemente, 1982)?
2. Are there differences in SEL implementation across participants with different years of professional experience?
3. Are there differences in SEL implementation across grade level?
4. Are there differences in SEL implementation across social economic status?

As demonstrated later in the unidimensionality section, the SSCA makes use of two scales: (a) Provider Characteristics (PC), and (b) Organizational Capacity (OC). Each of these scales covers a number of sub-domains via 16 and 24 items respectively (40 items altogether). Initially, the components were treated as part of the same construct. Because validation is an ongoing process, the results herein presented are

preliminary—they are intended to inform future use and revisions of this instrument (Messick, 1989).

Demographics Results

The SSCA included a demographics section that captured participants' knowledge, perceptions, and current practices. Table 5 displays the results. In general, 75% of participants indicated that their school uses SEL programs and the majority of their training occurred during a workshop (60%). Further, 46% of teachers are currently implementing and 46% believe they should be responsible for implementation. The classroom is the primary setting SEL is being taught (77%) but only 69% believed it should be.

Table 5

Participants Knowledge, Perceptions, and Practices

Questions		n	%
Participants		87	100%
School Uses SEL			
	Yes	65	75%
	No	12	14%
	Don't Know	8	9%
	Unknown	1	2%
Training			
	½ Day School In-service	14	16%
	Full Day School In-service	27	31%
	Workshop	52	60%
	Read Relevant Books	30	34%
	Watched a Video	11	13%
	Prior Work Experience	19	22%
	Graduate Training	14	16%
	Onsite Coaching	35	40%
	Not Listed	3	3%
Currently Most Responsible for Implementing SEL			
	Regular Education Teacher	40	46%
	Special Education Teacher	5	6%
	School Counselor	8	9%
	Psychologist	6	7%
	Behavioral Specialist	2	2%
	School Social Worker	8	9%
	Administrator	3	3%
	Nobody	3	3%
	Don't Know	8	9%
	Other	4	6%
Primary Setting SEL is Being Taught	Classroom	60	77%
	Small Group	9	12%
	One-on-one	2	3%

Objective 1 Results

Dimensionality

Using the Statistical Package for the Social Sciences 19 (SPSS), a principal axis factoring method was used to extract the factors. The eigenvalues and variance explained for these factors are displayed in Table 6. Looking at the scree plot (Figure 7), it could be interpreted that this could be a three- or two-factor solution, with two very dominant factors. Our choice of factors was not pre-determined, but after examining the factor loadings and careful examination, we decided that a two-factor solution was the most interpretable. The items loaded on the Organizational Capacity (OC) and Provider Characteristics (PC) factors. Given all the information above and taking into consideration that ultimately, factor selection is not only a matter of statistical judgment but that interpretability play a large role as well, we split the scale into two: (a) one scale for Organizational Capacity (OC) and (b) one scale for Provider Characteristics (PC).

Table 6

Total Variance Explained- First 10 Eigenvalues

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	11.00	27.52	27.52	10.72	26.80	26.80
2	5.38	13.47	40.99	5.08	12.70	39.50
3	3.02	7.56	48.56	2.75	6.87	46.38
4	1.99	4.97	53.54	1.62	4.05	50.43
5	1.92	4.80	58.34	1.58	3.97	54.40
6	1.69	4.22	62.56	1.39	3.49	57.89
7	1.64	4.10	66.67	1.25	3.13	61.02
8	1.28	3.20	69.87	.937	2.34	63.37
9	1.18	2.96	72.83	.833	2.08	65.45
10	1.02	2.56	75.40	.722	1.80	67.25

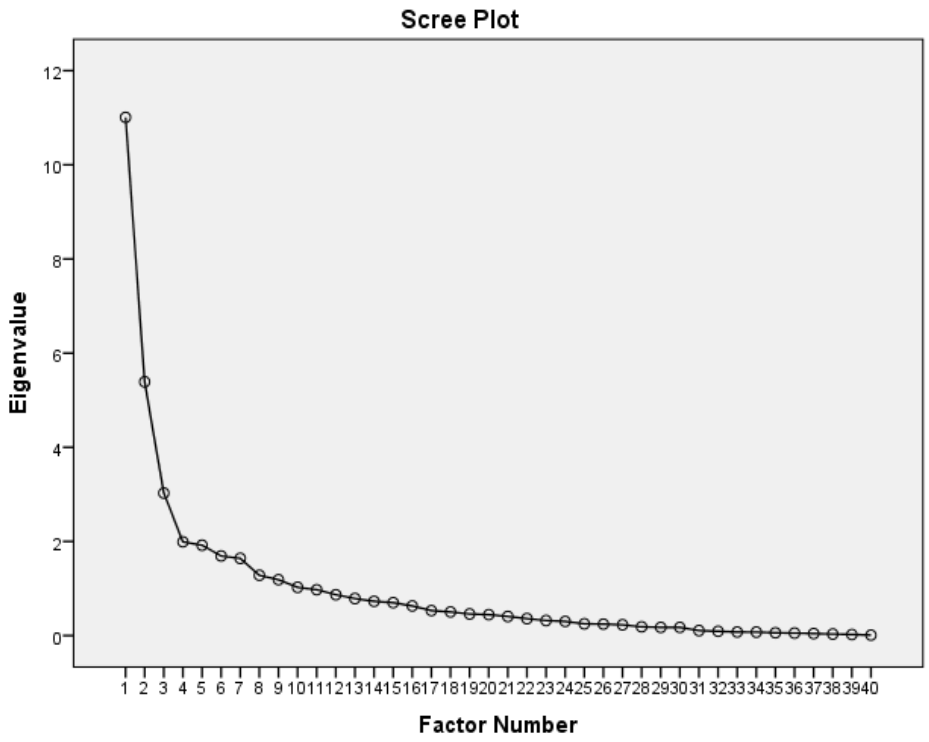

Figure 7. Scree Plot EFA for the SSCA– 40 Items

Provided in Table 7 are the results from these two scales, which proved to be statistically better. By examining the factor loading by item, we are able to represent how well an item correlates with a specific factor. Statistically, factor loading by item is the correlation between the manifest (observed) score and the latent (inferred) score. In terms of the SSCA, it provides an indication as to which of the items appear in the OC or PC factors. In our case, factor loading by item in the OC factor (factor 1) ranged from .410 to .841, and factor loading by item in the PC factor (factor 2) ranged from .293 to .743.

Table 7.

Factor loading by item

Item	Item	Domain	Factor 1	Factor 2
My principal helps communicate a shared-vision for school-wide SEL programming.	48	OC	.841	-.108
My principal plays a key role in implementing school-wide SEL programs.	47	OC	.812	-.091
My school strongly values the use of school-wide SEL programming.	33	OC	.785	.103
There is excellent leadership regarding setting priorities for school-wide SEL programs.	45	OC	.784	-.108
The principal of my school(s) is an advocate of school-wide SEL programming.	43	OC	.744	-.123
My school allots time on a regular basis to implement SEL lessons.	31	OC	.726	.033
There is an internal advocate for school-wide SEL programming.	46	OC	.713	.143
My school allots sufficient time to implement SEL lessons.	32	OC	.678	-.014
I strongly value the use of school-wide SEL programming.	34	PC	.650	.293
New programs are integrated into existing practices and routines.	55	OC	.590	-.338
My school has funding to purchase SEL curriculum/materials for teachers/staff.	39	OC	.585	.027
My district supports school-wide SEL initiatives.	44	OC	.584	-.163
Among staff at my school, there is a shared sense of responsibility to foster students' SEL skill development.	41	OC	.583	-.091
My school plans strategically for new programming.	56	OC	.563	-.405
School staff is united regarding the value and purpose of school-wide SEL programming.	42	OC	.556	-.074
In my school, ongoing coaching or consultation would be provided as new practices are being implemented.	57	OC	.549	-.401
My school allots regular prep time for SEL program planning.	30	OC	.530	-.050

The number of students in my classroom make it easy to do a “non-academic” activity.	50	OC	.517	-.011
Staff at my school tend to work together to resolve disagreements.	51	OC	.497	-.329
New staff are trained in on-going initiatives.	58	OC	.496	-.329
-	38	PC	.494	.314
My school encourages risk-taking in regard to new programming.	54	OC	.485	-.355
Staff feel encouraged to communicate openly with each other.	52	OC	.480	-.384
My school has ENOUGH funding to effectively support SEL programming (e.g., training, ongoing support).	40	OC	.471	-.122
I am confident I have the skills to deliver a SEL program.	35	PC	.468	.235
My school is open to change in regard to new programming.	53	OC	.423	-.284
I am confident that I can build students' emotion awareness.	37	PC	.411	.339
I have received an adequate amount of training about SEL.	49	OC	.410	.243
I believe it is part of the teacher’s job to teach children how to manage distress and other upsetting feelings (e.g., anger, anxiety, sadness, and shame).	25	PC	.350	.743
School-wide SEL programming is relevant to my schools needs.	28	PC	.295	.480
I believe it is part of the teacher’s job to teach children to UNDERSTAND the emotions of others.	20	PC	.293	.707
I believe it is part of the teacher’s job to teach children about optimism.	24	PC	.259	.656
I believe using a SEL program improves the likelihood of students learning how to appropriately manage emotions IN THEIR OWN WAY.	22	PC	.195	.667
It is important for teachers to help children MANAGE their own emotions.	21	PC	.120	.620
It is important for teachers to help children learn how to IDENTIFY their own emotions.	19	PC	.065	.470

I believe it is part of the teacher's job to teach children to CONSIDER the emotions of others.	23	PC	.053	.673
I believe parents place responsibility on teachers to teach children how to manage their emotions.	26	PC	-.012	.248
Only a small group of students in my school(s) need SEL programming.	29	PC	-.038	-.401
I would prefer to use my own resources (no curriculum) to teach emotion awareness.	36	PC	-.182	-.259
I worry about how parents will react to school-wide SEL programming in schools.	27	PC	-.433	-.128

Most of the items that load high (>0.4) on Factor 1, measure the construct of OC, except for items 34, 35, and 38, which measure the PC factor. Also note that the items that load high on Factor two are all measuring PC. Last, it is noted that there are only three items that do not load high on either factor: items 27, 36, and 39. The decision to keep or remove these items will also take into account the fit statistics that result from the rating scale analysis. So in the meantime, we flag these items but do not remove them.

Psychometric Quality of the SSCA

In Wright and Masters (1982) book, Rating Scale Analysis, they establish five guidelines to evaluate the quality of an instrument (i.e., accurate measurement). In the investigation to evaluate the quality of the scales, information was gathered to answer these questions:

1. Have we succeeded in defining a discernible line of increasing intensity?
2. Is item placement along this line reasonable?
3. Do the items work together to define a single variable? (i.e., consistency)
4. Have we succeeded in separating persons along the line defined by the items?
5. How valid is each person's measure?

To answer these questions, we applied the Rating Scale model (RSM) to the data

for each of the scales. We examined item separation statistics (i.e., separation index and reliability), chi-square, and fit statistics for items and persons. Item separation informs us how well items spread out along the measure (i.e., defining a meaningful variable) and reliability informs us about internal consistency (Linacre, 2014). Chi-square (i.e., statistical significance) allows us to see whether the items are defining a discernible line of increasing intensity.

Fit statistics are used diagnostically to measure the fit between observed values and expected values, and they show the size of randomness in the data. Both infit and outfit indexes are reported. Infit is more sensitive to unexpected observations on items in the middle and outfit is more sensitive outlier items. The tables below also display the measure and model standard error for each item in both scales. Briefly, the measure index provides intervals or spacing (i.e., logits) of item endorsability and indicates the placement of an item (i.e., more or less). The standard error (i.e., Model S.E.) quantifies the precision and uncertainty of a measure index.

For visual analysis, refer to the variable item-person map (Figure 8). The variable map shows the distribution of the items and persons, which is laid out vertically on the same continuum, with persons who endorse more items and the most difficult to endorse items at the top. The left-hand column shows the measure in logits, the middle shows the items, and right-hand columns shows teachers (represented by an *) for the OC and PC scales. Although there is no ideal distribution, it should match expectations, which can vary depending on the purpose of the instrument and the target population.

Measure	OC Scale		PC Scale	
	-Items	+Teachers	-Items	+Teachers
(More) 4				
			27	
3				
		*		
		*		
	30	*		
2				
		**		
		**		
		***		*
			38	*
1		****	37	**
	32	***	36	*********
	40, 31	**	26	********
	39, 56	****	35	*********
	57, 45	******	22, 34, 24	********
	58, 42	******	28	*******
(Mean) 0		*********		***
	50, 54, 49	***	25	**
	46, 48	******		****
	53, 41, 47	*******	23	******
	44, 55	********	21	********
	33, 51	***		****
	52	**	29	****
-1		**		*****
				****
	43	****		
		**		
-2				
		*		
			20	
-3				
		*		
		*		
(Less) -4			19	
Measure	-Items	+Teachers	-Items	+Teachers
	OC Scale		PC Scale	

Figure 8. Item-Person Variable Map for the OC and PC Scale

A summary of item statistics for both scales is presented in Table 8, and the main statistical results for the OC and PC items are presented in Table 9 and 10. Also, table 11 presents a summary of person statistics for the OC and PC scales. Our expectations on the item statistics are presented in the foregoing sections.

Table 8

Summary of the Item Statistics

Measures	OC	PC
INFIT		
Mean	1.00	1.01
SD	0.27	0.46
OUTFIT		
Mean	1.01	1.19
SD	0.29	0.94
Separation	3.31	5.18
Strata	4.75	7.24
Reliability	.92	.96
Chi-Square Statistic (fixed)	289.1	546.9
Degrees of Freedom	23	15

Table 9

Results for the OC Scale (arranged by measure)

Measure	Model S.E.	Infit		Outfit		ID	Items
		MnSq	ZStd	MnSq	ZStd		
2.15	0.19	0.82	-1.2	0.77	-1.2	30	My school allots regular prep time for SEL program planning.
0.96	0.19	0.95	-0.3	0.97	-0.1	32	My school allots sufficient time to implement SEL lessons.
0.77	0.19	1.51	2.9	1.62	3.2	40	My school has enough funding to effectively support SEL programming.
0.66	0.19	0.97	-0.1	0.99	0	31	My school allots time on a regular basis to implement SEL lessons.
0.42	0.2	1.55	3	1.61	3.2	39	My school has funding to purchase SEL curriculum/materials for teachers/staff.
0.41	0.2	0.81	-1.1	0.81	-1.1	56	My school plans strategically for new programming.
0.35	0.21	0.75	-1.5	0.76	-1.4	57	In my school ongoing coaching or consultation would be provided as new practices are being implemented
0.33	0.2	0.58	-3	0.58	-2.9	45	There is excellent leadership regarding setting priorities for school-wide SEL programs.
0.23	0.2	0.99	0	1	0	58	New staff are trained in on-going initiatives.
0.22	0.2	1.05	0.3	1.02	0.1	42	School staff is united regarding the value and purpose of school-wide SEL programming.
-0.11	0.21	1.38	2	1.38	2	50	The number of students in my classroom makes it easy to do a “non-academic” activity.
-0.12	0.22	0.71	-1.7	0.67	-1.9	54	My school encourages risk-taking in regard to new programming.
-0.22	0.21	1.67	3.3	1.73	3.5	49	I have received an adequate amount of training about SEL.

-0.27	0.21	1.01	0.1	1.13	0.7	46	There is an internal advocate for school-wide SEL programming.
-0.28	0.21	0.82	-1	0.82	-1	48	My principal helps communicate a shared-vision for school-wide SEL programming.
-0.38	0.21	0.79	-1.2	0.78	-1.2	53	My school is open to change in regard to new programming.
-0.39	0.21	1.06	0.4	1.05	0.3	41	Among staff at my school there is a shared sense of responsibility to foster students' SEL skill development.
-0.4	0.21	0.97	-0.1	0.97	-0.1	47	My principal plays a key role in implementing school-wide SEL programs.
-0.46	0.21	0.91	-0.5	1	0	44	My district supports school-wide SEL initiatives.
-0.5	0.22	0.92	-0.4	0.86	-0.7	55	New programs are integrated into existing practices and routines.
-0.62	0.21	0.81	-1.1	0.77	-1.3	33	My school strongly values the use of school-wide SEL programming.
-0.62	0.21	1.02	0.1	1	0	51	Staff at my school tend to work together to resolve disagreements.
-0.67	0.21	0.98	0	0.96	-0.1	52	Staff feel encouraged to communicate openly with each other.
-1.47	0.22	1.05	0.3	1.02	0.1	43	The principal of my school(s) is an advocate of school-wide SEL programming.
0	0.21	1	0	1.01	0	0	Mean (count: 24)
0.7	0.01	0.27	1.5	0.29	1.6	0.7	S.D. population
0.71	0.01	0.27	1.6	0.3	1.6	0.7	S.D. sample

Model, Sample: RMSE .21 Adj (True) S.D. .68 Separation 3.31 Strata 4.75 Reliability .92
Model, Fixed (all same) chi-square: 289.1 d.f.: 23 significance (probability): .00

Table 10

Results for the PC Scale (arranged by measure)

Measure	Model S.E.	Infit		Outfit		ID	Items
		MnSq	ZStd	MnSq	ZStd		
3.89	0.17	1.57	3.4	1.64	3.7	27	I worry about how parents will react to school-wide SEL programming in schools.
1.08	0.2	0.7	-2	0.72	-2	38	I would prefer helping children to learn how to manage emotions with the aide of a structured SEL curriculum.
0.97	0.2	0.73	-1.8	0.71	-2	37	I am confident that I can build students' emotion awareness.
0.93	0.2	2.02	5.1	1.97	5	36	I would prefer to use my own resources (no curriculum) to teach emotion awareness (reversed)
0.72	0.2	1.04	0.2	1.07	0.4	26	I believe parents place responsibility on teachers to teach children how to manage their emotions.
0.56	0.21	1.19	1.2	1.13	0.8	35	I am confident I have the skills to deliver a SEL program.
0.41	0.21	0.55	-3.4	0.56	-3.3	22	I believe using a SEL program improves the likelihood of students learning how to appropriately manage emotions in their own way.
0.38	0.21	1.16	1	1.05	0.3	34	I strongly value the use of school-wide SEL programming.
0.31	0.21	0.56	-3.3	0.58	-3.2	24	I believe it is part of the teacher’s job to teach children about optimism.
0.18	0.21	0.68	-2.3	0.66	-2.3	28	School-wide SEL programming is relevant to my school’s needs.
-0.22	0.22	0.67	-2.3	0.71	-1.7	25	I believe it is part of the teacher’s job to teach children how to manage distress and other upsetting feelings.
-0.54	0.24	0.77	-1.4	1.06	0.3	23	I believe it is part of the teacher’s job to teach children

							to consider the emotions of others.
-0.59	0.24	0.78	-1.4	0.85	-0.7	21	It is important for teachers to help children manage their own emotions.
-0.76	0.25	1.87	4.2	1.65	2.4	29	Only a small group of students in my school(s) need SEL programming.
-2.81	0.47	0.79	-0.4	0.34	-1.1	20	I believe it is part of the teacher's job to teach children to understand the emotions of others.
-4.5	1.01	1.04	0.3	4.41	1.7	19	It is important for teachers to help children learn how to identify their own emotions.
0	0.28	1.01	-0.2	1.19	-0.1		Mean (Count: 16)
1.75	0.2	0.44	2.6	0.94	2.4		S.D. Population
1.8	0.21	0.46	2.7	0.97	2.5		S.D. Sample

Model, Sample: RMSE .34 Adj (True) S.D. 1.77 Separation 5.18 Strata 7.24 Reliability .96
Model, Fixed (all same) chi-square: 546.9 d.f.: 15 significance (probability): .00

Table 11

Summary of the Person Statistics

Measures	OC	PC
INFIT		
Mean	1.02	1.04
SD	0.53	0.54
OUTFIT		
Mean	1.01	1.01
SD	0.53	1.04
Separation	2.54	1.45
Strata	3.72	2.27
Reliability	.87	0.68
Chi-Square Statistic (fixed)	554.8	255.2
Degrees of Freedom	86	86

Have we succeeded in defining a discernible line of increasing intensity?

To answer the first question, we examined the item separation index for each scale as well as their reliability (Table 8). The item separation index for the SSCA is used to verify the item hierarchy and indicates how well items spread along the measure by providing the standard deviation of the scores in calibration error units (i.e., item discrepancies). A good separation index would be greater than 3 and anything less implies that the sample of teachers is not be large enough to verify the item hierarchy (Linacre, 2014). In turn, the reliability of the scale is a Rasch-based measure equivalent to Cronbach's alpha (Linacre, 2014). The reliability means that there is a high reliability that teachers with high measures do have higher measures than teachers with low measures. A good reliability estimate would be greater than 0.90 and anything less implies that the item difficulty range is not be large enough (Linacre, 2014).

In the case of the OC scale, the item separation index is 3.31, and the reliability is .92. In the case of the PC scale, the separation index is 5.18 and the reliability is .96. Facets output also provides a chi-square test to see whether the items are defining a discernible line of increasing intensity. In both cases, the chi-square was significant, indicating that the items are defining a discernible line of increasing intensity (Linacre & Wright, 1994).

Is item placement along this line reasonable?

In relation to the second question, we analyzed the difficulty of each item in relationship to our expectations. The results are presented in Tables 9 and 10. Generally speaking, our expectations about order were met. For the OC scale, items related to the allocation of time and funding were the most difficult items for teachers to endorse. We often find that teachers have limited time to implement and integrate non-academic programs into their daily routine because they may not be top priorities for schools, which spend a great deal of time improving academic achievement, such as reading and mathematics. This is further compounded by the fact that teachers and schools are often evaluated on students' academic proficiency but not social and emotional competency.

Items related to leadership and a shared vision was amongst the easiest items to endorse. Teachers felt that their principal was an advocate for schoolwide SEL programming and that their school and district support SEL initiatives. In many ways, it is easier for administrators to encourage the use of SEL but much more difficult to adequately set up the structures and provide the resources to make it feasible for teachers to implement.

Administrators may provide supports but they are usually centered on professional development and training, which was observed in the data. Items related to training were moderately difficult for teachers to endorse on the OC continuum. The data meet our expectations in that school and district administrators encourage SEL programming, but as demonstrated, they may not always ensure teachers are given ample time and finances to ensure quality implementation. Although the duplicity of administrative support seems contradictory, it is much easier to voice support than it is to take action, bust barriers, and build drivers within a school.

For the PC scale, items related skill proficiency and self-efficacy were amongst the most difficult items for teachers to endorse, but on the converse, items related to perceived need for the SEL programming were easier. Again, our expectations were met. We expected teachers to believe that it is important to support students' competency and believe it is a part of their job. It is however more difficult to feel proficient and effective at implementing SEL programs. One's belief about perceived need deals with external factors, such as student competencies, and doesn't require agency (i.e., teacher action). On the other hand, skill proficiency and self-efficacy is an evaluation of one's actions and abilities—which aren't mere observations of others but take into consideration personal awareness and confidence.

Interestingly enough, although teachers found it difficult to endorse their skill proficiency and self-efficacy, they also found it difficult to endorse items related to both using a structured SEL curriculum and using their own resources (no curriculum). We were expecting teachers to be more comfortable using a structured curriculum because it greatly improves skill proficiency and self-efficacy, but these two questions may be

addressing teachers' potential philosophical belief that teaching is an art versus a science. It might also show that some teachers feel more confident using as scripted curriculum because it provides structure, while others may find it easier to modify their own resources to meet the demands of their students.

Do the items work together to define a single variable (i.e., consistency)?

In relation to the third question about consistency, we examined the fit indexes for items from both scales (Tables 9 and 10). Fit indices reflect to what extent the items fit the model. Fit statistics are reported as mean square statistics (MnSq) and a criterion for misfit is a MNSQ of <2.0 (Linacre, 2014). Consequently, its expected value is 1.0; values greater indicate underfit, suggesting unmodelled noise and a deviation from unidimensionality in the data (not the scales). Further, items with outfit values greater than 1.2 are not necessarily problematic, however, they should be identified.

For the OC scale, the outfit and infit MnSq of 1.01 and 1.0, respectively, indicate little distortion. Under the criterion for misfit, no items were identified as presenting misfit for the OC scale. There were however, several items that have a MNSQ of <1.2, which were items 39, 40, 49, and 50. The first two items addressed funding, the third addressed classroom size, and the fourth addressed training.

For the PC scale, the outfit and infit MnSq is 1.19 and 1.01, respectively. In general terms, the fit analysis indicates that the data correspond with the additive measures. However, there is one item with clear misfit on the PC scale (4.41 MnSq; item 19: It is important for teachers to help children learn how to identify their own emotions), which may justify the item's removal. There were also several items that have a MNSQ of <1.2, which were items 19, 27, 29, and 36. These items were quite different from each

other. Item 19 addresses the importance of teachers helping children identify their emotions, item 27 addresses parent concerns about SEL programming, item 29 addresses the importance of schoolwide versus small group programming, and item 36 addresses teacher preference for using their own resources in lieu of a curriculum.

Visually, our expectations about consistency are good. With the results presented in Figure 8, the items for both the OC and PC are reasonably spread out. Both items and people's attributes are dispersed along the measures. Although, for the OC scale, there are a few outliers on the high and low end of the scale, resulting in gaps between those items closer to the middle. When the item spread is compared to the person spread, there is not a complete match, suggesting that the OC scale may benefit from more items on the high and low end of the measures. As for the PC scale, the item-person map reveals that the distribution of the items targets the teachers relatively well. Teachers were however, more centrally located. There were items that were much higher and lower than the teachers, which could justify their removal and items could be added to the top and bottom that aren't located well beyond the teacher spread.

Have we succeeded in separating persons along the line defined by the items?

In relation to the fourth question, we examined the person separation index and the test reliability statistics for people (Table 11). The OC scale presents a person separation of 2.54 and a reliability of .87. The PC scale presents a person separation of 1.45 and a reliability of .68. These statistics are lower than for the items, but the expectations of these statistics are also lower. Although the values are within acceptable ranges, it would be preferable if the statistics for the PC scale were higher; the person separation and reliability index are somewhat problematic. Nonetheless, both chi-square tests result

were significant, indicating that the examinees are defining a discernible line of increasing intensity (Linacre & Wright, 1994).

How valid is each person's measure?

In relation to the last question, we examined the person fit statistics (Table 11). The interpretation of these statistics is similar to that of items, in the sense that they indicate the extent to which the model works appropriately for them. As previously mentioned, fit statistics are reported as MnSq statistics and a criterion for misfit is a MnSq of <2.0. In this regard, there are five individuals who do not fit to the OC scale, and five different individuals who do not fit to the PC scale. Still, these individuals were kept in the analysis. Subsequent studies would benefit from speaking with misfitting persons to determine the source of the misfit.

Assess the Rating Scale Utility of the SSCA

The OC and PC scales have standardized fixed-response formats characterized by participant self-report with polytomous choice. Rating scales with polytomous choice such as these, are employed as a means to gather a breadth of information that increases measurement accuracy and precision beyond simple dichotomous response categories. Specifically, the following Likert scale was used: strongly agree, agree, somewhat agree, somewhat disagree, disagree, and strongly disagree.

We determined that through the proceeding analysis, the six Likert-style response categories were not the most appropriate; hence, we optimized the response categories to improve effectiveness. Investigating the rating scale diagnostics is useful in that it allowed us to better understand the extent to which respondents utilized each rating and how well each rating fits the rating scale. The objective was to satisfy all eight guidelines

proposed by Linacre (1999) and when warranted, response categories were recategorized to insure adherence. The analyses are based on the output provided by facets, which is displayed in Tables 12 and 13 for the OC and PC scales.

In order to provide the best response categorization, we used the following eight guidelines for investigating rating scale utility (Linacre, 1999):

1. At least ten observations of each category
2. Regular observation distribution
3. Average measures advance monotonically with category
4. Outfit-mean squares less than 2.0
5. Step calibrations advance
6. Ratings imply measures and measures imply ratings
7. Step difficulties advance by at least 1.0 logits
8. Step difficulties advance by less than 5.0 logits

The initial FACETS output for analysis of the rating scales using the original 6 categories is presented in Tables 13 and 14. The rating scales are numbered as follows: 1 = strongly disagree; 2 = disagree; 3 = somewhat disagree; 4 = somewhat agree; 5 = agree; 6 = strongly agree. The left-hand column displays the cardinal value assigned to each rating category. The data column provides the number of observations in each category, the observed percent of used responses and the percent of the used responses in or above the category.

The quality control column displays the average of the measure (i.e., logit); if an * if present, then it implies that the average measure does not increase with each higher category and the higher category doesn't correspond to more of the variable. The column

also displays the expected value of the average measure according to what would be expected if the data fit the Rasch model. If the observed and expected measure is a close match, then it suggests that the category is aligned with the latent variable. Lastly, the column also displays the outfit MnSq; with an expectation of 1.0 and anything much larger (<2.0) indicate unexpected observations in the category.

The Rasch-Andrich threshold column displays the step calibration, or in other words, the intersection where the probability of a teacher assigning a rating in adjacent categories is 50% or equal. For instance, a person with a measure 1.2 logits has the same probability of selecting category 2 as they have of selecting category 3. Put differently, the step calibrations are the boundaries between categories, and indicate the transition from one category to the next. To illustrate this, use the original category probability curves for each scale in Figure 9 and 10.

In the most probable form column, the categories lowest measure at which it is the most probable (i.e., modal) to be observed is indicated. A low indication means the category is the most probable category at the low end of the scale. The last or right-hand column displays the response category name.

Table 12

Original OC Scale Rating Scale Utility Summary Data

	DATA			QUALITY CONTROL			RASCHANDRICH		MOST	Rating
	Category Counts			Avg.	Exp.	OUTFIT	Thresholds		PROBABLE	Category
Score	Total	%	Cum %	Meas.	Meas.	MnSq	Measure	S.E.	from	
1	143	8%	8%	-2.41	-2.38	1.1			low	Strongly disagree
2	508	28%	36%	-0.47	-0.51	1	-2.42	0.11	-2.64	Disagree
3	967	53%	89%	0.78	0.81	1.1	-.49	0.06	-.49	Somewhat disagree
4	194	11%	100%	2.30	2.25	1	3.13	0.09	3.13	Somewhat agree
5	0									Agree
6	0									Strongly agree

Table 13

Original PC Scale Rating Scale Utility Summary Data

	DATA			QUALITY CONTROL			RASCHANDRICH		MOST	Rating
	Category Counts			Avg.	Exp.	OUTFIT	Thresholds		PROBABLE	Category
Score	Total	%	Cum %	Meas.	Meas.	MnSq	Measure	S.E.	from	
1	21	2%	2%	-0.18	-1.37	3.5			low	Strongly disagree
2	39	3%	5%	-.88*	-0.54	0.6	-1.62	0.27	-1.62	Disagree
3	60	5%	9%	0.43	0.49	1.0	-.44	0.18		Somewhat disagree
4	483	38%	47%	1.14	1.2	0.9	-1.22	0.12	-.83	Somewhat agree
5	105	8%	55%	1.77	1.73	1.3	3.00	0.07		Agree
6	579	45%	100%	2.28	2.25	0.9	.28	.07	1.64	Strongly agree

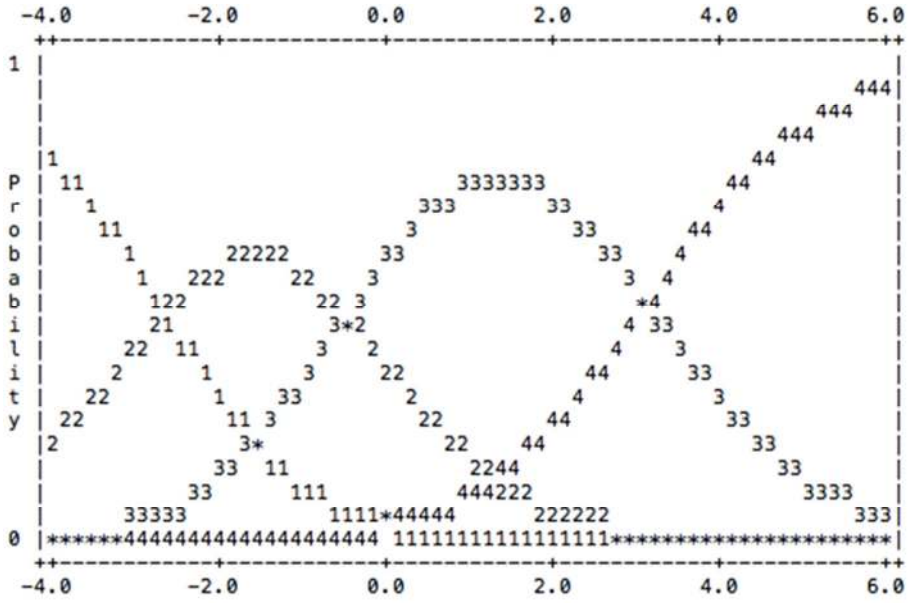

Figure 9. Original Probability Curves for OC scale

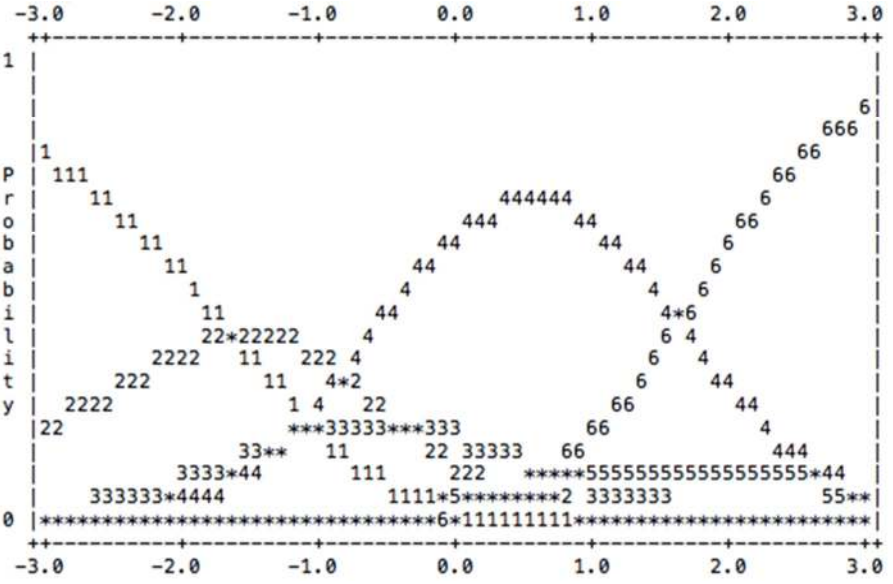

Figure 10. Original Probability Curves for PC scale

Table 14

Original OC and PC Rating Scale Utility Analysis

Guidelines	Organizational Capacity		Provider Characteristics	
	Met/Not	Details	Met/Not	Details
10 obs. per category	Not	Category 5 has no observations	Met	
Regular obs. distribution	Not	There is a high concentration on category 4	Not	Categories 4 and 6 are higher than the others.
Avg. measures advance monotonically with category	Met		Not	Category 2 has a lower average measure than Category 1
Outfit MnSq <2.0	Met		Not	Not met for Category 1
Step Calibrations Advance	Not	Category 3 is never modal	Not	Category 3 and 5 is never modal
Ratings imply measures and measures imply ratings	Not	Besides category 5, coherence is above the 40% benchmark	Not	Coherence is 0% between measures and ratings
Step difficulties advance by ≥1 logit	Not	The difference in step calibrations for category 2 (-.51) and category 3 (-.85) is <1 logit.	Not	There is a problem with categories 3 and 5
Step difficulties advance by <5 logits	Met		Met	

*For both the OC and PC scales: 1 = strongly agree; 2 = agree; 3 = somewhat agree; 4 = somewhat disagree; 5 = disagree; 6 = strongly disagree

The initial analysis of the rating scales using the original 6 categories is presented in Table 14. Each of the Met/Not columns indicates whether the guideline was met or not met for that particular scale. As observed, most of the guidelines were not met. Therefore, changes were made and an iterative analysis was performed to determine if the changes improved the quality of measurement.

For the OC Scale, category 5 has no observations, and therefore should be merged. Given the distribution of observations per category (concentrated in category 4), we merged category 5 with category 6. Another issue is that category 3 is never modal, and so, should be merged with an adjacent category. For similar reasons as category 5, we merged category 3 with category 2. The same solution was undertaken with the PC scale. The new categories for both the OC and PC scales are: 1 = strongly disagree; 2 = disagree; 3 = agree; 6 = strongly agree. The analysis of this new rating scale utility is presented in Table 17. As observed, most of the guidelines are now met for both scales. One issue is that the distribution of observations for the PC scale is skewed. Given that the item separation was large, this may suggest that the phrasing of the questions were such that the probability of answering one category was higher than answering the other categories. However, given our recruitment process, it may be more likely a product of the sample population, implying that the participants may have more experience implementing SEL programs.

Table 15

Revised OC Scale Rating Scale Utility Summary Data

	DATA			QUALITY CONTROL			RASCHANDRICH		MOST	Response
	Category Counts			Avg.	Exp.	OUTFIT	Thresholds		PROBABLE	Category
Score	Total	%	Cum%	Meas.	Meas.	MnSq	Measure	S.E.	from	Name
1	143	8%	8%	-2.41	-2.38	1.1			low	Strongly disagree
2	508	28%	36%	-0.47	-0.51	1	-2.64	.11	-2.64	Disagree
3	967	53%	89%	0.78	0.81	1.1	-.49	.06	-.49	Agree
4	194	11%	100%	2.3	2.25	1	3.13	.09	3.13	Strongly agree

Table 16

Revised PC Scale Rating Scale Utility Summary Data

	DATA			QUALITY CONTROL			RASCHANDRICH		MOST	Response
	Category Counts			Avg.	Exp.	OUTFIT	Thresholds		PROBABLE	Category
Score	Total	%	Cum %	Meas.	Meas.	MnSq	Measure	S.E.	from	Name
1	21	2%	2%	-.03	-1.53	3.0			low	Strongly disagree
2	39	8%	9%	.24	.39	1.0	-2.19	0.27	-2.47	Disagree
3	60	38%	47%	1.90	2.04	1.3	-.30	0.12	-.15	Agree
4	483	53%	100%	4.00	3.93	.9	-2.49	0.07	-2.49	Strongly agree

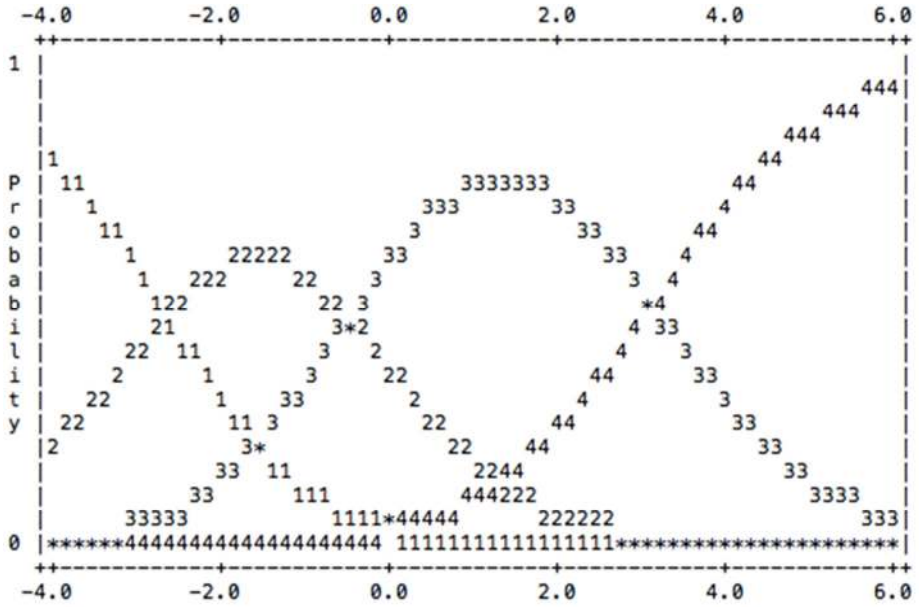

Figure 11. Revised Probability Curves for OC scale

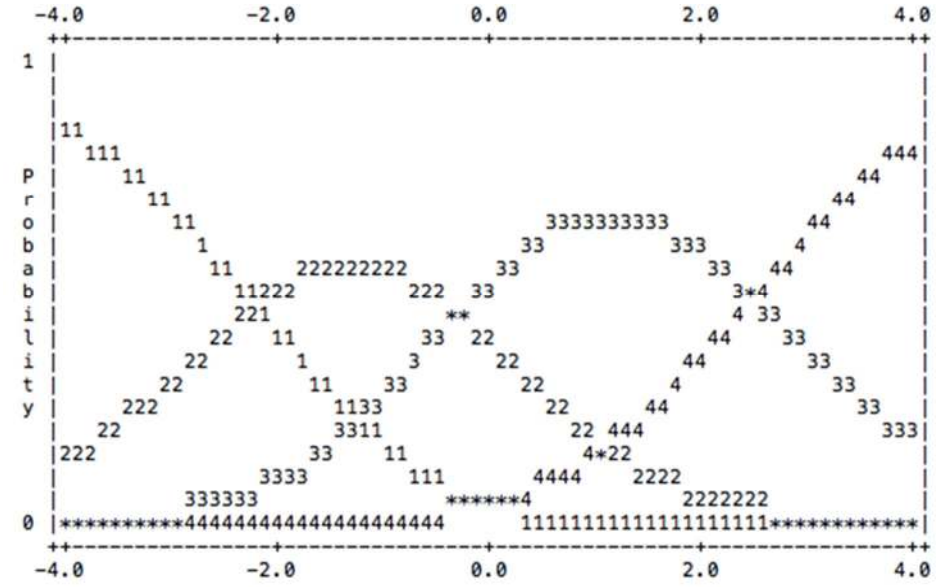

Figure 12. Revised Probability Curves for PC scale

Table 17

Revised OC and PC Rating Scale Utility Analysis

Guidelines	Organizational Capacity		Provider Characteristics	
	Met/Not	Details	Met/Not	Details
10 obs. per category	Met		Met	
Regular obs. distribution	Met		Met	
Avg. measures advance monotonically with category	Met		Met	
Outfit MnSq <2.0	Not	Category 1 presents a high Outfit MnSq.	Met	
Step Calibrations Advance	Not	Category 3 is never modal.	Met	
Ratings imply measures and measures imply ratings	Met		Met	Met for all except for Rating 4.
Step difficulties advance by ≥1 logit	Met		Met	
Step difficulties advance by <5 logits	Met		Met	

*For both the OC and PC scales: 1 = strongly agree; 2 = agree; 3 = somewhat agree; 4 = somewhat disagree; 5 = disagree; 6 = strongly disagree

Objective 2 Results

In order to satisfy the second objective, we answered four research questions: Are there differences in the SEL implementation across participants' (1) Stage-of-Change, (2) years of professional experience, (3) grade levels, and (4) social economic status (SES)? To answer these questions, we examined differences in mean responses (i.e., logit values [thetas]) across subgroups. For example, within SES, there are three subgroups: low, medium, and high. We examined the confidence intervals of the mean rating per subgroup and if two confidence intervals overlapped, then there were not significant differences between the subgroups. Otherwise, the subgroups were significantly different. This was done for the OC and PC scales separately. The results of the analyses are provided below.

For visual analysis, refer to the variable map of the OC and PC scales, in Figure 13 and 14, respectively. The left-hand column shows the measure, and the second and third columns show the items and teachers. Each teacher is indicated by an *. The last four columns show each question's subgroups placed at its mean calibration and locate them along the variable. For statistical analysis, a summary of the facets statistics of both scales is presented in Table 18.

Measures		-Organizational Capacity Items	+Teachers	-Stage	- Yrs.	- Grade	- SES
(More)	3		**				
		Regular prep time.	*				
	2		***				
			****				
	1	Sufficient time. Allots time regularly.	****				
		Enough funding. Funding to purchase.	*****	Exactly.			High
		Plans strategically.	*****		2-4.		
		Ongoing coaching. New staff are trained.	******		>15.		
		School staff is united. Excellent leadership.	******		5-7.	Elem.	
(Mean)	0		******	Components.	10-15.		Mid.
		Risk-taking. Number of students. Adequate	*****	Not reg.			Low
		training. Principal's vision. Internal advocate.	******	Soon.	0-1.	Mid.	
		Shared responsibility. District support.	*********		8-10.	DWP.	
		Principal's role. Open to change.					
		Programs are integrated. Strongly values.					
		Work together.	***	Thought about.			
		Communication.	**	Never.		High.	
	-1		****				
		Principal advocates.	****				
	-2						
			*				
	-3						
(Less)	-4		**				
Measures		-Organizational Capacity Items	+Teachers	-Stage	- Yrs.	- Grade	- SES

Figure 13. Variable Map of the OC Scale for Objective 2.

Measures	-Provider Characteristics Items	+Teachers	-Stage	- Yrs.	- Grade	- SES
(More) 4						
	Parents react.					
3						
2						
		****	Soon.	0-7. 8-15.		
1	Confidence. Structured curriculum.	*********		>15.		
	Own resources. Parental beliefs.	********	Not reg.		Elem.	High.
	Delivery.	*********	Exactly.		High.	Mid.
	Optimism.	********	Components.		DSW.	Low.
	Manage emotions. Values SEL.	*******	Thought about.			
0	Relevant.	*****	Never.		Mid.	
	Distress.	****				
		**********				
	Teach children to consider.	**********				
	Children manage.	**********				
-1	Small group of students.					
		*********				
-2						
	Children to understand.					
-3						
-4						
	How to identify.					
(Less) -5						
Measures	-Provider Characteristics Items	+Teachers	-Stage	- Yrs.	- Grade	- SES

Figure 14. Variable Map of the PC Scale for Objective 2.

Table 18.

Measurement Summary Report

	Stage of Change		Professional Experience		Grade Level		SES	
Measures	OC	PC	OC	PC	OC	PC	OC	PC
INFIT								
Mean	.98	.94	1.04	.96	1.10	1.04	1.01	1.00
SD	.28	.27	.14	.14	.16	.22	.05	.10
OUTFIT								
Mean	.98	1.28	1.05	.97	1.14	1.02	1.02	1.34
SD	.29	.93	.13	.30	.20	.23	.07	.62
Separation	2.56	.78	.00	.00	1.80	.42	3.51	.00
Reliability	.87	.38	.00	.00	.76	.15	.92	.00
Chi-Square Statistic (fixed)	122*	10.3*	2.4	1.0	35.3*	11.7*	37.3*	2.2
Degrees of Freedom	5	5	5	5	3	3	2	2

*Significant

Participants' Stage-of-Change

As depicted in the tables below, there are six subgroups according to the Stage-of-Change variable. For the OC scale, the ordered subgroups, from higher to lower measure, and the results of the pairwise comparisons, are presented in Table 19. To interpret this and the following ordered subgroups tables, the left-hand columns show the cardinal number for each subgroup, the second columns provides a description of each, the third columns displays their measure (i.e., mean calibration), and the right-hand columns presents the results for the pairwise comparison. Again, pairwise comparisons were used to examine whether the subgroups were significantly different from each other. If they were significantly different, it indicated that the subgroup was significantly higher than the remaining subgroups.

As can be observed, the majority of teachers (i.e., 67%) were in both the Maintenance stages (fidelity [subgroup 1] and adoption [subgroup 2]), implying that they are implementing SEL programs, which is a most likely a product of the sample population. For clarification, the Maintenance-fidelity subgroup (subgroup 1) differ from the Maintenance-adoption subgroup (subgroup 2) in that the former are implementing an SEL program in full accordance with its published details, rather than choosing certain components, which may lead to a greater potential for inconsistencies. Because subgroups 3, 4, and 6 have limited sample sizes, the generalizability of the data may be limited, suggesting there are too few people to make any meaningful inferences.

Based on the ratings obtained by means of the OC scale, the chi-square test provided by FACETS indicates that there is a significant difference between subgroups across participants' Stage-of-Change.

Table 19.

Ordered SOC Subgroups According to the OC Scale

Subgroup	n	Description	Measure	Significantly Higher Than
1	22	I am using an SEL program exactly as described in the program materials	.74	2, 3, 4, 5, 6
2	36	I am using an SEL program but only use certain components	.06	3, 5, 6
3	9	I am using an SEL program, but not on a regular basis	-.20	5, 6
4	2	I will use an SEL program soon	-.17	6
5	11	I have thought about using an SEL program	-.53	-
6	6	I never plan to use an SEL program	-.77	-

In particular, subgroup 1("I am using an SEL program exactly as described in the program materials") has the highest mean rating, and was significantly higher than all of the

other subgroups. Higher implies that it is easier for that particular subgroup to endorse the items in the scale (i.e., more attributes). Several of the other subgroups were also significantly different. As observed there are five groupings: (a) one formed by subgroup 1, (b) a second formed by subgroup 2, (c) a third formed by subgroup 3, (d) a fourth by subgroup 4, and (e) a fifth by subgroups 5 and 6.

In relation to the PC scale, the ordered subgroups, from higher to lower measure, and the results of the pairwise comparisons, are presented in Table 20.

Table 20

Ordered SOC Subgroups According to the PC Scale

Subgroup	n	Description	Measure	Significantly Higher Than
2	36	I am using an SEL program but only use certain components	1.13	3, 4, 6
1	22	I am using an SEL program exactly as described in the program materials	0.85	3, 4, 6
5	11	I have thought about using an SEL program	0.77	4, 6
3	9	I am using an SEL program	0.56	-
4	2	I will use an SEL program soon	0.43	-
6	6	I never plan to use an SEL program	0.36	-

The chi-square test for the PC scale is also significant, meaning that there are differences between groups across participants' Stage-of-Change. In particular, the subgroups with the highest mean rating was subgroups 1 and 2 ("I am using an SEL program exactly as described in the program materials", and, "I am using an SEL program but only use certain components", respectively), which were significantly higher than all of the other subgroups. Several of the other subgroups were also significantly different. As observed there are three groupings: (a) one formed by subgroups 1 and 2, (b) a second formed by subgroups 5, and (c) a third formed by subgroups 3, 4, and 6.

Years of Professional Experience

As depicted in the tables below, there are three subgroups according to the years of professional experience variable. Based on the ratings obtained by means of the OC and PC scale, the chi-square test provided by FACETS indicates that there is not a significant difference, meaning that there are no differences between subgroups across years of experience for either scale. The ordered subgroups, from higher to lower measure, and the results of the pairwise comparisons, are presented in Table 21 and Table 22.

Table 21

Ordered Year of Experience Subgroups According to the OC Scale

Subgroup	n	Description	Measure	Significantly Higher Than
3	36	> 15 years	0.22	-
1	20	0 to 7 years	0.18	-
2	30	8 to 15 years	0.13	-

Table 22

Ordered Year of Experience Subgroups According to the PC Scale

Subgroup	n	Description	Measure	Significantly Higher Than
1	20	0 to 7 years	0.68	-
3	36	> 15 years	0.65	-
2	30	8 to 15 years	0.62	-

Grade Level

As depicted in the tables below, there are four subgroups according to the grade level variable. The majority of participants (i.e., 83%) were elementary school teachers. The remaining sample of teachers was much lower, with the least amount in the high school subgroup (n = 3). This may be a product of the recruitment process or that elementary teachers are more

engaged in the implementation of SEL programs, which would explain why there were so few middle and high school teachers that participated.

Because they're so few teachers in the other grade levels, any comparison would be tenuous and we have opted to not to include the analysis in our results. For transparency, the ordered subgroups, from higher to lower measure, and the results of the pairwise comparisons, are presented in Table 23 and Table 24.

Table 23.

Ordered Grade Level Subgroups According to the OC Scale

Subgroup	n	Description	Measure	Significantly Higher Than
1	71	Elementary school	0.19	3, 4
2	5	Middle school	-0.33	3
4	6	Districtwide	-0.40	-
3	3	High school	-0.76	-

Table 24.

Ordered Grade Level Subgroups according to the PC Scale

Subgroup	n	Description	Measure	Significantly Higher Than
1	71	Elementary school	.90	2, 3, 4
4	6	Districtwide	.48	-
3	3	High	.53	-
2	5	Middle	.31	-

Social Economic Status

As depicted in Table 25, there are three subgroups according to the Social Economic Status (SES) variable. In this case, the number of participants for each subgroup was more evenly spread out. The majority of teachers taught in middle SES schools, with high SES having the second, and low SES having the lowest amount of participants. The argument for

generalizability is much greater for the SES facet because of the greater number of participants in each subgroup.

Based on the ratings obtained by means of the OC scale, the chi-square test provided by FACETS indicates that there is a significant difference between subgroups across SES. The ordered categories according to the mean measure, as well as the results of the pairwise comparisons, are shown in Table 25.

Table 25

Ordered SES Subgroups According to the OC Scale

Subgroup	n	Description	Measure	Significantly Higher Than
1	25	High SES	.50	2, 3
2	42	Middle SES	.01	3
3	17	Low SES	-.17	-

The subgroup with highest mean rating was category 1 (high SES), and was significantly higher than all the other two subgroups. The other two subgroups were also significantly different. As observed there are three groupings: (a) one formed by subgroup 1, (b) a second formed by subgroup 2, and (c) a third formed by subgroup 3.

In relation to the PC scale, the chi-square test was not significant, meaning that there are no differences in means across SES. The ordered categories according to the mean measure, as well as the results of the pairwise comparisons, are shown in Table 26.

Table 26

Ordered SES Groups According to the PC Scale

Subgroup	n	Description	Measure	Significantly Higher Than
1	25	High SES	0.86	-
2	42	Middle SES	0.72	-
3	17	Low SES	0.63	-

CHAPTER 5

DISCUSSION

Summary of the Present Study

The current investigation sought to further develop and validate the School SEL Capacity Assessment (SSCA). Past development efforts of the SSCA involved the use of the construct modeling process, which supported the conceptualization of the construct and provided supporting evidence of test content. The process included the construct map, item design, outcome space, and the measurement model (Wilson, 2005). Validity evidence on test content was also gathered using expert evaluations and evidence of response process was gathered using Think Aloud Protocols. The purpose of the current study was to continue the validation process using the Rasch scale model (RSM), in which validity evidence based on internal structure was gathered and preliminary inferences about the current state of SEL implementation are reported.

Because the RSM is a unidimensional measurement model, dimensionality and quality control procedures were employed to support its internal structure. Initially, the SSCA components were treated as part of the same construct, but exploratory factor analysis established that it was multidimensional. Thus, the SSCA was separated into two unidimensional scales: (a) Organizational Capacity (OC) and (b) Provider Characteristics (PC).

During the next stage, we evaluated the psychometric quality of the SSCA. We proceeded by applying the RSM to the data and used Wright and Masters (1982) guidelines for accurate measurement. The guidelines helped evaluate how well the SSCA's item ability work together to define a meaningful variable. It also addresses the extent to which teachers are separated along the same line and the validity of their responses.

First, the data demonstrates that the items for both the OC and PC scales define a

discernible line of increasing intensity along the SSCA implementation construct continuum, with items that are easier to endorse to progressively more difficult items. Second, the placement of the items is reasonably dispersed along the measure. Although, by analyzing the variable map for the OC scale, there is justification for adding more items on the extreme ends of the measure to better match the teachers that are located on the very low and very high ends of the continuum. As for the PC scale, most teachers were located on the middle of the measure, hence, more items could be added to the middle to better match the teachers' locations. Third, the items define a single variable, with the exception of one clear misfit for the PC scale: item 19, "It is important for teachers to help children learn how to identify their own emotions."

Further, the OC and PC scales succeeded in separating teachers along the measure, which included teachers that were more and less able to endorse difficult items. There were however five teachers for the OC scale and five different teachers for the PC scale who did not fit according to the RSM. In all, the data is encouraging and provides promising validity evidence of internal structure for the OC and PC scales. Within reason, the SSCA met the qualifications for accurate measurement.

Since the SSCA employs a rating scale with polytomous choice (i.e., strongly agree, agree, somewhat agree, somewhat disagree, disagree, and strongly disagree), we assessed the rating scale utility using guidelines proposed by Linacre (1999). It was determined that the response categories were not the most appropriate and we recategorized them to insure adherence. For both the OC and PC scales, we merged the category "strongly agree" with "agree", and category "somewhat disagree" with "disagree." The revised categories for both the OC and PC scales are: strongly agree; agree; disagree; and strongly disagree. Most of the guidelines are now met for both scales.

Findings from the Rasch analysis were also used to answer questions from the second objective. The study analyzed survey response differences among facets within the SSCA implementation construct. Facets included participants' Stage-of-Change (SOC), years of professional experience, grade level, and schools' level of social economic status. Simply put, we asked if there were differences in SEL implementation across participants' responses according to their indicated levels (i.e., subgroups) within the four facets.

For the purpose of this study, the SOC model asserts that the implementation of SEL programming ranges from precontemplation (not planning on implementing) to maintenance (already implementing). For the SOC facet, the chi-square test indicates that there is a significant difference among subgroups for the OC and PC scales. Generally, teachers who are implementing an SEL program found it easier to endorse items on the SSCA, suggesting that they have more OC and PC attributes. The opposite is true for teachers not implementing an SEL program, in that they found it difficult to endorse items, suggesting that they have less OC and PC attributes.

Specifically, the results of the pairwise comparisons for the OC scale show that the maintenance-fidelity subgroup ("I am using an SEL program exactly as described in the program materials") is significantly higher (i.e., mean rating) than all of the other subgroups. In this case, higher implies that it is easier for that particular subgroup to endorse the items in the scales. On the opposite end, teachers in precontemplation ("I never plan to use an SEL program") were the lowest subgroup, and as a consequence, have the least OC attributes. As for the PC scale, both the maintenance-fidelity and maintenance-adaptation ("I am using an SEL program but only use certain components") subgroups had statistically similar mean ratings and were significantly higher than all of the other subgroups. The preparation ("I have thought about using an SEL

program") and precontemplation subgroups had statistically similar mean ratings and were the lowest subgroups.

For the years of professional experience facet, the chi-square test indicates that there is not a significant difference among subgroups for the OC and PC scales. Contrary to what we would expect, years of professional experience is not related to teachers' endorsement of OC and PC attributes. For the grade level facet, they're so few teachers in the middle, high, and districtwide grade levels, hence, we have opted to not to include the analysis in our results. For the Social Economic Status (SES) facet, the chi-square test indicates that there is a significant difference among subgroups for the OC scale. As anticipated, the results of the pairwise comparisons show that the high SES subgroup is significantly higher than the middle and low SES subgroups; indicating that teachers reported level of OC is affected by their schools SES—less resources equal less capacity. The chi-square test is not significant for the PC scale, indicting that teacher reported level of PC are not affected by their schools SES.

The following discussion includes an interpretation of the main findings of this study. This discussion is presented in relation to the proposed research questions. Limitations of the study are reviewed and implications for future research are also discussed.

Objective 1 Conclusions

A principal axis factoring method was used to extract the factors and determine if the SSCA is unidimensional. The analysis indicated that the SSCA consisted of two dominant factors and based off the results, we split the SSCA into two scales, which proved to be statistically better. Factor 1 was identified as Organizational Capacity (OC) and factor 2 was identified as Provider Characteristics (PC). By examining the factor loading by item, we looked at how well each item correlates with the OC and PC factors.

The OC factor loading by item ranged from .410 to .841. There are three items that load high (>0.4) on the OC factor, but were originally conceptualized as attributes of PC: items 34, 35, and 38. A closer look at these items that do not conform to our expectations, reveals possible causes. Item 34, "I strongly value the use of school-wide SEL programming," may be correlated with the OC of schools if teachers interpret the item as referring to the school's social emotional learning (SEL) program in-place and not about using schoolwide SEL programs in general. Item 35, "I am confident I have the skills to deliver a SEL program," may be contingent on the OC of teacher's particular school, although the item attempted to ask their confidence in having the skills in general. For item 38, "I would prefer helping children to learn how to manage emotions with the aid of a structured SEL curriculum," teachers may be concerned that a structured SEL curriculum requires an investment in organizational resources, such as time and finances, and as a result, is relating more to OC factors.

The PC factor loading by item ranged from .293 to .743. There are no items that load high on the PC factor that were originally conceptualized as attributes of OC. As indicated by the principal axis factoring method, all items from the PC scale correlated well with the PC factor. However, it should be noted that there are three items that do not load high on either factor but were in fact conceptualized as attributes of PC: items 27, 29, and 36.

Item 27, "I worry about how parents will react to schoolwide SEL programming in schools," and item 29, "Only a small group of students in my school(s) need SEL programming," appear to not be related to their schools' OC or to the teachers' PC in schoolwide SEL programming. Perhaps there is a third variable related to community factors that aren't being addressed in either of the scales. In Durlak and DuPre's (2008) meta-analysis used for the development of the SSCA, their conceptualization of community level variables focused heavily

on the dissemination of information about prevention science and less on how the community might hinder schools' and teachers' ability to implement SEL. With the exception of funding, we appropriately opted to not include these variables in the SSCA. Future revisions might explore more malleable community factors that could improve the conceptualization of the SSCA.

Item 36, "I would prefer to use my own resources (no curriculum) to teach emotion awareness," is similar in respect to item 38 in that it addresses questions around the use of a curriculum. The notion of using curriculums for SEL programming needs to be explored in more depth. Both items might benefit from being reworded to better reflect the relationship between the provider (i.e., teacher) and their personal characteristics. It could be that the items would benefit from using an adjective rather than a verb, such as replacing "I would prefer" with "I feel more comfortable."

Fit statistics that result from the rating scale analysis should be considered when deciding to keep or remove items from the PC scale (i.e., 27, 29, 36, and 38). Item 27 has an outfit mean square statistic (MnSq) of 1.64, which is under the criterion for misfit (<2.0; Linacre, 2014). The item's measure index (i.e., logit) is 3.89, which means it is dispersed on the very high end of the PC scale. This slight misfit could be due to the fact that no teachers are near it, so the problem may not be with the item but rather with the biased sample. As for item 29, its outfit MnSq is 1.65 but its item measure index is -0.76, which means it is dispersed in the middle of the PC scale. Because we want to retain the items in the middle, we may want to keep item 29. The same logic could be applied to Item 36. It has an outfit MnSq of 1.65, which is also under the criterion for misfit and its measure index is 0.76, which means it is dispersed in the middle of the PC scale. Thus, it may be warranted to keep the item in the SSCA. The MnSq for

item 38 is 0.72 and has a measure index of 1.08. Item 38 has the second highest measure index and matches the teacher spread. Since the highest teacher index is 1.17, it would be beneficial to keep item 38.

Once unidimensionality was established, we evaluated the psychometric quality of the SSCA (for both the OC and PC scales) using Wright and Masters five guidelines (1982) for accurate measurement by applying the Rating Scale model (RSM) to the data. Item separation statistics, chi-square, and fit statistics for items and persons were examined for each scale. First, to determine whether we have succeeded in defining a discernable line of increased intensity, we looked at the separation index and reliability, as well as the chi-square for each scale. The separation index and reliability for the OC scale is 3.3 and 0.92; and the separation index and reliability for the PC scale is 5.2 and 0.96. For both scales, the chi-square is significant. Because the chi-square is significant and the separation indices are greater than 3, we can confidently state that we have developed items that cut across the entire construct of OC and PC, measuring a full range of endorsability.

Second, in order to determine if the item placement along the line is reasonable, we looked at the outfit and infit MnSq, and we analyzed the items visually on the item-person variable map. The outfit and infit MnSq for the OC scale is 1.01 and 1.0, and the outfit and infit MnSq for the PC scale is 1.19 and 1.01. Statistically, the summary fit analysis of item statistics for both the OC and PC scales were close to the expected value of 1.0, suggesting that the unidimensional, additive measures (i.e., logits or parameter estimates) present a clear picture of the data and the items correspond well with the additive measures.

Because the item summary statistics suggest that there is no unmodelled noise, we looked at the variable map to identify areas where there is an item person mismatch. Looking at the OC

scale, the items are reasonable spread out, however, there are teachers located on the high and low end of the measure, indicating that items could be added to both ends of the continuum. The highest item (i.e., item 30: "My school allots regular prep time for SEL program planning") has an item measure of 2.15 and the highest teacher measure is 2.61, demonstrating that there is a slight mismatch. The lowest item (i.e., item 43: "The principal of my school(s) is an advocate of school-wide SEL programming") has an item measure of -1.47 and the lowest teacher measure is -3.63, also demonstrating a mismatch.

Because the highest item has a measure of 2.15 and the second highest item (i.e., item 32: "My school allots sufficient time to implement SEL lessons") has an item measure of 0.96, it would be advisable to add items above and in-between these two items. Interestingly, the fourth most difficult item, (i.e., item 31: "My school allots time on a regular basis to implement SEL lessons") also addresses the allocation of time, which might indicate that items related to time are important variables to expand on in future revisions of the SSCA. The same argument could be made for the third (i.e., 40: "My school has enough funding to effectively support SEL programming) and fifth (i.e., item 39: "My school has funding to purchase SEL curriculum/materials for teachers/staff") most difficult items, which address barriers to funding, and may also be a good candidate for expanding on in the SCCA.

On the low end of the OC scale, there were several teachers who were placed well below the easiest item, indicating that the item spread does not perfectly match the person spread. Items could be placed below item 43, "The principal of my school(s) is an advocate of school-wide SEL programming," which was the easiest item. Nevertheless, the SSCA is designed to identify barriers that teachers commonly face, and items on the difficult end warrant more attention. Simply speaking, it will be the difficult items that prevent successful schoolwide SEL

implementation.

Looking at the PC scale, the items are also reasonable spread out, but in this case, there are more teachers located on the middle of the measure and less of a spread. On the high end, the most difficult item, 27 ("I worry about how parents will react to schoolwide SEL programming in schools") has an item measure of 3.08, but the highest teacher measure is 1.17. Since item 27 has an item measure that is well beyond the teacher spread, and as already mentioned, does not load high on the PC factor, it could be a candidate for removal.

With that said, future revisions of the SSCA might benefit from exploring more community level factors (e.g., parental concerns). It could be the case that the community plays a role in schools' OC but not in teachers' PC. In other respects, revisions to item 27 might benefit from being more direct. For example, the item could be stated as such: "I know that parents react negatively (or positively) to schoolwide SEL programming in schools."

On the low end of the PC scale, there are two items that do not match the teachers' locations. Items 19 and 20, "I believe it is part of the teacher's job to teach children to understand the emotions of others" and "It is important for teachers to help children learn how to identify their own emotions" have a measure of -2.81 and -4.5. However, the lowest person measure was -1.14, demonstrating that those items are easy to endorse, providing less information on barriers that are difficult to implement. In order to make the PC scale more concise and efficient to administer, items 19 and 20 could be candidates for removal.

Third, in order to identify whether the items work together to define a single variable, we examined the fit indexes from both scales. Since the criterion for misfit is a MnSq of <2.0, no items were identified as presenting misfit for the OC scale (Linacre, 2014). For the PC scale, item 19 presented a clear misfit (4.41 MnSq); as previously mentioned, the item measure was

low (-2.81), presenting a mismatch from the teacher population, which provides more evidence that the item could be removed.

Fourth, we examined the person separation index and reliability statistics in order to determine if we succeeded in separating persons along the line defined by the items. The person separation and reliability for the OC scale is 2.5 and .87, respectively; and for the PC scale, the values are 1.45 and .68. The values were in within an acceptable range, and because the chi-square test results are significant, we can state that the teachers are defining a discernable line of increased intensity (Linacre & Wright, 1994).

Fifth, we examined the person fit statistics, which indicated that there were five teachers who do not fit to the OC scale, and five different teachers who do not fit to the PC scale. For the OC scale, the teachers Outfit MnSq were <2.0, ranging from 2.06 to 2.80. For the PC scale, the Outfit MnSq ranged from 2.23 to 4.71. Although MnSq values were above the criteria, the teachers' aberrant response patterns do not mean that their data are misleading. After taking the teachers out of the analysis, the summary results did not change, and because of that, we left them in the final analysis.

Once the psychometric quality of the SSCA was confirmed, we assessed the Rating Scale Utility of the SSCA (for both the OC and PC scales) using guidelines proposed by Linacre (1999). For the original OC and PC scales, we used the following Likert scale: strongly agree, agree, somewhat agree, somewhat disagree, disagree, and strongly disagree. Initial analysis indicated that the original response categories were not the most appropriate. We merged several categories to maximize adherence to the proposed guidelines and the revised categories for both the OC and PC scales are: strongly agree; agree; disagree; strongly disagree. In doing so, most of the guidelines are now met for both scales. One issue we still face is that the distribution for

the PC scale is skewed. It could be the case that the phrasing of the questions influenced teacher response, resulting in an increased probability of responding more in one category than another. Also, the sample of teachers chosen for the study could have limited the way they responded to the scales, suggesting that their tendency to choose "agree" and "strongly agree" was a product of the population.

Objective 2 Conclusion

In addressing Objective 2 we answered four research questions (i.e., facets): Are there differences in the SEL implementation across participants' (1) Stage-of-Change, (2) years of professional experience, (3) grade levels, and (4) social economic status (SES)? We first assessed the chi-square test provided by FACETS to identify if there were significant differences between subgroups and then we proceeded by assessing the differences in mean responses by examining the confidence intervals across subgroups within each facet. If confidence intervals did not overlap, we were able to confidently state that the subgroups were statistically different from the other subgroups.

Stages of Implementation

Prochaska and DiClemente (1982) Stages-of-Change (SOC) model was originally conceptualized as a heuristic to better understand behavior change. We revised it to better understand teachers' current schoolwide SEL implementation efforts, and how the stages related to their schools' OC and their individual PC. The SOC as applied to the SSCA proposes that teachers' willingness to implement schoolwide SEL is dependent on higher levels of school OC and individual PC attributes. Teachers indicated their current implementation stage, ranging from maintenance to precontemplation, as depicted in Table 27.

Table 27

SOC Model Applied to Stages of Implementation

Stage	Description
Maintenance-fidelity	I am using an SEL program exactly as described in the program.
Maintenance-adaptation	I am using an SEL program but only use certain components.
Action	I am using an SEL program, but not on a regular basis.
Preparation	I will use an SEL program soon.
Contemplation	I have thought about using an SEL program.
Precontemplation	I never plan to use an SEL program.

Caution is warranted about the generalizability of some of the stages because the sample sizes were small, especially for the Preparation and Precontemplation stages. Most of the participants endorsed one of the two Maintenance stages (i.e., 67%), with just 11% in the Action stage, 2% in the Preparation stage, 13% in the Contemplation stage, and 7% in the Precontemplation stage. This aligns with our expectations in regard to the sample population, which consisted of some teachers that champion SEL programming, and also maintain an active SEL program in their classroom. Nevertheless, generalizability is compromised and in this case, general rather than specific inferences would be more justified.

For the SOC facet, the chi-square test indicates that there are significant differences between the subgroups for both the OC and PC scale. The results suggest that there is a relationship between schools' OC and teachers' implementation of schoolwide SEL programs. The results also suggest the same relationship for teachers' PC. For the ordered subgroups in the OC and PC scale, along with the results of the pairwise comparisons, refer to Table 28 and Table 29.

Table 28

Ordered Subgroups According to the OC Scale for the SOC Facet

Subgroup	Significantly Higher Than
Maintenance-fidelity	All
Maintenance-adaptation	Action, Contemplation, Precontemplation
Preparation	Contemplation, Precontemplation
Action	Precontemplation
Contemplation	-
Precontemplation	-

Table 29

Ordered Subgroups According to the PC Scale for the SOC Facet

Subgroup	Significantly Higher Than
Maintenance-fidelity	Action, Preparation, Precontemplation
Maintenance-adaptation	Action, Preparation, Precontemplation
Contemplation	Preparation, Precontemplation
Action	-
Preparation	-
Precontemplation	-

Generally speaking, teachers who are currently using an SEL program (i.e., maintenance) found it easier to endorse items on the SSCA (i.e., more attributes). On the converse, teachers who are not using an SEL program (i.e., precontemplation) found it more difficult to endorse items. In all, the SOC model was relatively good at predicting teachers' level of OC and PC. Teachers with high levels of OC and PC were implementing SEL programs and teachers with low levels were not implementing SEL programs.

As anticipated for the OC scale, teachers within the maintenance-fidelity stage have the highest measure index (0.74), indicating they have more attributes, and they were statistically higher than all the other subgroups. The results imply that teachers who are using an SEL program with fidelity essentially have schools with more OC construct attributes (i.e., more

funding, leadership championing SEL, time allocated to implement, etc.). A similar observation can be made for teachers in the maintenance-adaption stage, who were the second highest subgroup and statistically higher than all the remaining stages, with the exception of the preparation stage.

For the PC scale, the teachers who are maintaining an SEL program (fidelity or adaption) have more PC construct attributes, such as perceived need, benefits, self-efficacy, and skill proficiency. Specifically, teachers within both maintenance stages (fidelity and adaption) have the highest measure indexes (0.85 and 1.13) and are statistically higher than all the other subgroups; that is, with the exception of teachers in the contemplation stage (measure of 0.77). For teachers in the maintenance-adaptation stage, the data may demonstrate that their high levels of self-efficacy and skill proficiency make them more comfortable modifying the program to better meet students, organizational, and community needs.

It is not surprising that these teachers also have high levels of PC attributes. Teachers with self-efficacy and skill proficiency may feel more comfortable deviating from the program to help ensure compatibility. What the data doesn't show us is whether they are acceptable adaptations. The debate continues between the pros and cons of fidelity and adaptability, but what is more clear, is that teachers who are maintaining implementation of an SEL program feel as though they have more PC.

In spite of the fact that the stages make conceptual sense and provide a means to classify teachers' level of implementation, the data show that the stages are not mutually exclusive. We arbitrarily divide the stages, but statistically, the stages better reflect a continuum of implementation. Hence, it may make clinical sense to communicate stages, but we should be cautious to assume that they are distinctly different—there is overlap.

As schools reduce barriers and increase drivers, it may facilitate teachers' progression through the stages, going from not ready to taking action. Teachers stage-of-implementation may reflect present levels of systems-level factors, which interact to either augment or hinder their progress through the stages. Although the results do not show causation, they do provide valuable information for future studies. The central argument is that teachers' stages of implementation should be viewed as the product of the interplay of schools OC and teachers PC processes. Further research may want to examine this mutual relationship.

Years of Professional Experience

It was expected that years of professional experience would play a role in schools' level of OC and teachers' PC. For example, professional experience could result in schools with more OC because capacity building is a long-term continual process that involves the use of teachers' PC. It could be assumed that teachers with more experience would spend less time on instructional planning and more time on improving their schools' OC. Also, more experience could help teachers create a stronger sense of self-efficacy and skill proficiency through mastery experiences and professional development.

Despite our expectations, the chi-square test was not significant; there were no significance differences between groups for the OC and PC scale. In other words, the study did not find that teachers' years of professional experience affected their schools level of OC or their level of PC. Perhaps we didn't capture the nature of years of professional experience. It might be that the question was phrased improperly. The ordered subgroups in the OC and PC scale, along with the results of the pairwise comparisons, are displayed in Table 30 and Table 31.

Table 30

Ordered Subgroups According to the OC Scale for the Experience Facet

Subgroup	Significantly Higher Than
0 to 7 years	-
8 to 15 years	-
> 15 years	-

Table 31

Ordered Subgroups According to the PC Scale for the Experience Facet

Subgroup	Significantly Higher Than
0 to 7 years	-
8 to 15 years	-
> 15 years	-

Grade Level

One limitation with the results is the skewed distribution of sample sizes in grade levels. The elementary grade level teachers accounted for 83% of all the participants. The sample size for the middle, high, and districtwide teachers were limited. The small sample sizes for the higher grades may be a result of the recruitment process, and although it is believed that it could be a reflection of the current state of SEL implementation among teachers, limitations about generalizability are warranted. Thus, we have opted out of comparing the subgroups and including the analysis in our results.

Social Economic Status

The Social Economic Status (SES) facet was divided into high, middle, and low SES subgroups. The chi-square test indicates that there is a significant difference between subgroups across SES for the OC scale but not for the PC scale. For the ordered subgroups in the OC and PC scale, along with the results of the pairwise comparisons, refer to Table 32 and Table 33.

Table 32

Ordered Subgroups According to the OC Scale for the SES Facet

Subgroup	Significantly Higher Than
High SES	All
Middle SES	Low
Low SES	-

Table 33

Ordered Subgroups According to the PC Scale for the SES Facet

Subgroup	Significantly Higher Than
High SES	-
Middle SES	-
Low SES	-

As expected, the subgroup with the highest mean rating for the OC scale was teachers in high SES schools. The data implies that high SES schools have more funding, leadership that champions SEL, and more time to allocate towards implementation, to name a few. More funding allows schools to invest in programs and personnel. The end result may be leaders that are willing and able to champion evidence-based programming. These personnel not only advocate more often for SEL in their schools, but may also have more time to bust barriers and ensure drivers are in place to improve schools OC.

High SES schools may also face less adversity, resulting in a homogeneous student population with similar needs. The end result may be schools with more resources to invest in a smaller subset of risk factors. On the converse, low SES schools face a plethora of diversity and student needs. In the low SES case, limited resources would be spread thin across various programs, such as health and nutrition, English language learning, tier 3 supports for high risk students, reduced motivation, and issues of truancy, to name just few.

As the data show, high SES schools have improved OC, middle SES schools are second, and low SES schools have the lowest levels of OC. The results meet our expectations. What is interesting is that SES does not affect teachers PC. Teachers in high, middle, and low SES schools are not statistically different in this regard. In some ways, this meets our expectations. We would expect teachers in various SES schools to have similar self-efficacy and skill proficiency. However, it would be expected that teachers in low SES schools would have a higher perceived need for SEL programming. Students in low SES schools face more adversity and have higher risk, thus the need and benefits would be greater. Exploring these relationships would greatly benefit reform efforts and improve the implementation of SEL programming.

Links to Universal Social and Emotional Learning (SEL) Implementation

The present study is relevant to universal SEL delivery in schools, particularly during the planning stage of implementation. In order to minimize barriers and ensure drivers are in place to promote successful outcomes, educators need to attend to the readiness of their schools and ensure adequate supports are provided. The development of the SSCA will contribute to SEL implementation science by providing a tool to help identify context-specific barriers and drivers, and monitor progress over time. Accountability will be focused on system-level variables, essentially providing the means to support successful program impact and sustainability.

Buchanan et al. (2009) survey study indicates that teachers see the value in SEL programming but less than half are currently implementing SEL programs. This may be the result of poor planning; the study points out that teachers do not have enough time to implement and are not receiving proper training. The current study supports similar claims that there are barriers preventing teachers from implementing SEL programs successfully. In particular, the allocation of time and funding, a focus on strategically planning for new programming, and

training were among the hardest for teachers to endorse for the OC scale. For the PC scale, the use of a curriculum, parental beliefs, and confidence in delivering SEL programs were also difficult to endorse, implying that more work needs to be done prior to actual implementation.

To further support the claim that barriers may be preventing successful implementation, the current study shows that there may be a relationship between teachers' stages of implementation, and schools' OC and teachers' PC. The relationship demonstrates that teachers who are implementing SEL programs endorse more OC and PC attributes. In other words, teachers with high levels of OC and PC are implementing SEL programs and teachers with low levels are not implementing SEL programs.

Although more research is needed, teachers, who see the need and benefit, have self-efficacy and skill proficiency, and are provided proper funding, time allocation, and administrative leadership, find it easier to maintain an SEL program in their schools. In order to ensure teachers maintain quality implementation, present levels of context-specific factors need to be assessed and improved, which may interact to either augment or hinder their progress through the stages. The stance taken is that there is a mutual relationship between teachers' stages of implementation and their levels of OC and PC attributes.

The current study also suggests that elementary school may be the ideal place to start a universal SEL program. Elementary teachers endorsed more PC, in that they recognized the need and have more perceived skill proficiency. They also reported that their schools have more OC to support the implementation of SEL programs. As suggested earlier, a focus on strategic planning for new programming is needed, and in order to ensure adequate supports are provided, school districts may want to hone their efforts on a particular subset of their student population, such as elementary grade students.

Although the data show that elementary schools may be an ideal place to initiate reform, extra care is warranted for lower SES schools. On a positive note, teachers' level of PC was not affected by their schools level of SES; however, the data imply that low SES schools have less funding, reduced leadership that champions SEL, and less time to allocate towards implementation. This should not be viewed as evidence that low SES schools should not implement SEL programs. Rather, they should be particularly attentive to OC barriers that may impede success, seeing as though they may have more obstacles to overcome. With that being said, the SSCA was designed to ensure that schools could effectively identify these barriers, which may be especially important for lower SES schools.

Researchers such as Severson, Walker, Hope-Doolittle, Kratochwill, and Gresham (2007) suggest that assessment tools need to align with the needs and priorities of teachers. Although the authors were referring to student screening approaches, their point is applicable to the SSCA. The suggestion indicates that teachers respond best to assessment tools that meet their needs and support their priorities. As schools champion SEL programming and mandate evidence-based practice, teacher priorities will shift to accommodate these demands. However, if their needs aren't met, implementation may be lackluster, and they may feel under supported, and reform efforts may fail.

Payton et al., (2008) assert that a well-designed evaluation of implementation is an important feature of quality SEL programming. It would be quite nefarious for administrators to expect teachers to implement with fidelity while failing to evaluate their own means to ensure teacher success. The importance of supporting teacher implementation cannot be overlooked. The research is clear in that implementation barriers prevent successful outcomes (Durlak et al., 2011). If we fail to attend to implementation barriers, SEL programming will yield significantly

reduced effects in skills, social behavior, emotional stress management, and academic performance for students (Durlak et al., 2011). After all, teachers are increasingly being evaluated and held accountable for students' academic performance. Perhaps it is time to evaluate the barriers and drivers that prevent or augment teacher success.

Often times, implementation quality is referred to as program fidelity, with a specific focus on teacher delivery; however, teachers are only able to implement to the degree afforded to them by their schools' OC and their own PC. It is the responsibility of the school and school district to ensure they provide the supports needed to be successful. This is truly, the systems-level perspective. In all, in order to implement evidence-based SEL programs that result in academic success, attending to the implementation science and systems-level perspective is critical.

Limitations

The results from this study should be considered in light of the limitations, which place restrictions on the conclusions. Cook and Campbell (1979) proposed major threats to the validity of experimental studies, and although the current study is exploratory, their framework still serves as a useful reference. Several proposed threats to validity are of concern in the present study. The primary limitations include internal validity, issues of measurement, and external validity.

A major limitation of the study has to do with the internal validity, which includes recruitment, sample population, and sample size (Cook & Campbell, 1979). The recruitment process sought to recruit diverse school districts from the Midwest and Northeast United States. The objective was to recruit a representative sample of teachers (Snyder & Dillow, 2012).

Recruitment is often a difficult process for any study, and although we tried to curb

expected obstacles by carrying out the study in multiple geographic locations, the initial recruitment process resulted in fewer participants than expected. In order to have enough power and meet our minimum sample size, we made revisions to our recruitment methodology. In doing so, our selection of participants changed from our original conceptualization.

Our sampling technique relied less on the explicit criteria prior to initiation of the study. Our second attempt was based more on convenience. We used word-of-mouth and email recruiting as an alternative method. Our sample population was secured from the SEL Alliance for Massachusetts (SAM) and Michigan's Integrated Behavior and Learning Support Initiative (MiBLSi). We sent an email to members from both groups, which included an array of professionals. We were however explicit on our criteria for inclusion, and were able to recruit enough teachers to meet our minimum sample size needed for the study. Nevertheless, our sample size was the minimum and could have benefitted from more participants and greater diversity (Linacre, 1994).

Some of participants originated from an organization that champions the use of SEL programming in the state of Massachusetts (i.e., SAM). Hence, a further limitation of our study was our sample population (Heppner, Kivlighan, & Wampold, 1992). Some of teachers in our study may be different than teachers throughout the country. These teachers may believe there is a greater need for SEL and may also have more experience implementing SEL programs.

Another limitation includes issues of measurement, such as the nature of self-reporting (Cook & Campbell, 1979). One limitation of measurement had to do with volunteer bias (Barker, Pistrang, & Elliott, 2002). When teachers volunteer to participate in a study, they may not be representative of the larger population of teachers. Our sample may overrepresent teachers who have strong opinions about SEL. Rater bias may also be affected by social

desirability. Teachers participating in the study could have changed their responses, resulting in positive results and expectation bias (Barker et al., 2002). The tendency for participants to respond in socially desirable ways is a concern for psychological research. In general, teachers could have responded in a way that makes them look good, thus under- or over-reporting certain behaviors. For example, teachers could have rated their principal's leadership in a more positive light, resulting in a leniency effect. Also, teachers could have rated certain items based off of investigator expectations. We did however take several steps to minimize bias by ensuring confidentiality and providing anonymity. We sought to conceal the identities of participants and we limited the right of access to the data to strengthen the validity of teachers' responses.

Finally, another potential concern has to do with external validity (Shadish, Cook, & Campbell, 2002). The best way to be certain that the results from the study are representative of all teachers is to randomly select from the population of teachers. We sought to have a representative sample, but after an unsuccessful attempt, we chose practicality over representativeness. Thus, the validity of our generalized inferences may be diminished and we may not be able to confidently generalize our results to other teachers and situations.

Because part of our sample population had unique characteristics, the extent to which we can generalize from the teachers who participated to teachers in the general population (i.e., generalizability across people) may be limited. We cannot be fully confident that our sample population is a good representation of the population mean. For example, after the rating scale was revised to adhere to the guidelines proposed by Linacre (1999), the distribution of PC scale was still skewed. Perhaps the sample population limited the way they responded to the items, suggesting that they were more likely to choose "agree" and "strongly agree" than their similar counterparts.

The survey was taken online and the extent to which we can generalize from the teachers who participated online to teachers who may participate in a school setting (i.e., generalizability across situation) may also be limited. All situational specifics (e.g., location of survey administration, confidentiality, and anonymity) of the study potentially limit generalizability.

In a real-life situation, school staff would evaluate summary results, and teachers might be more concerned about whether their endorsements would be viewed negatively by administrators. For example, the tendency for teachers to rate their principals' leadership in a positive way might be affected by the size of the school, the expectations of their fellow teachers, and the concern that they may be reprimanded. In general, systems would need to be in place to ensure that teachers don't under- or over-reporting certain behaviors because of fear of reprisal and other concerns.

Implications of Future Research

Taking into consideration the results and limitations, the current study provides direction for future studies. Although the current study provided validity evidence for the SSCA and identified interesting information that is valuable for educators, the study was preliminary. Future studies could broaden the selection of participants, extend the validation process, and focus on ensuring the items are representative of the OC and PC constructs.

After items are revised, eliminated, and added to the SSCA based off of the studies' findings, future studies may want to repeat the validation process. A follow-up study could be investigated in a similar manner as the current study. It would also be beneficial to conduct a study that includes a larger and more diverse sample of teachers. Ensuring that teachers with limited SEL experience are included would help eliminate current limitations and improve the generalizability of the results.

A follow up study would also benefit from assessing other education professionals, such as school psychologists, school social workers, and administrative staff. Often times, support and administrative staff play a key role in implementing SEL programs. Whether support and administrative staff are co-teaching, evaluating student progress and teacher fidelity, providing professional development and technical assistance, or building implementation drivers and busting barriers, they will play a key role in ensuring teacher success with SEL.

Similar to the discussion on limitations, future studies may benefit from administering and evaluating the results in the school setting. It would not only improve the generalizability across situations, but it would also lend well to better understanding the consequence of testing—providing more validity evidence as outlined in the Standards for Psychological and Educational Testing (AERA/APA/NCME, 2014). Evidence regarding the unintended consequences of using the SSCA could help shape the intended use of the SSCA within schools. For example, we might find that administrators have plans to use the data for teacher evaluations, but because there are some items outside the control of teachers, the unintended use would be harmful. We would be better prepared to communicate our intended uses of the SSCA and could take a number of steps to ensure educators adhere to those standards. As one example, the SSCA administration protocol could be edited to better reflect the intended use and we would be better poised to address specific and common unintended uses.

In addition to validity evidence related to the consequence of testing, future studies may want to explore evidence related to relations to other variables (AERA/APA/NCME, 2014). This is akin to the traditional classification of predictive and concurrent validity. First, we could demonstrate a predictive relationship between the SSCA and the implementation of SEL programs. Second, we could demonstrate a relationship between the SSCA and student

outcomes. Third, we could demonstrate a concurrent relationship between other measures of OC and PC with the SSCA.

Since quality implementation is needed to effectively transmit SEL programs and practices to schools, it is important to understand whether the SSCA does in fact capture important barriers and drivers. We would predict that high levels of OC and PC would equate to better implementation. A measure of teacher SEL implementation fidelity would help ensure that varying levels of OC and PC do in fact affect implementation. For example, a criterion measure such as teacher adherence to SEL components, the number of lessons taught, and the quality of instruction (e.g., direct observation, fidelity checklists) could be correlated with teachers' level of PC and their schools OC.

Evaluating the predictive relationship between the SSCA and student outcomes is another fruitful endeavor. It would be predicted that schools implementing SEL programs would observe better student outcomes if they have higher scores on the SSCA. A measure of student outcome variables would help ensure that varying levels of OC and PC do in fact affect student outcomes. For example, a criterion measure such as students SEL competency (e.g., Social Emotional Assets and Resilience Scales), behavior (e.g., office discipline referrals) and academic proficiency (e.g., grades) could be assessed.

Ultimately, it would be expected that when schools evaluate and improve their schools' OC and teachers' PC, teachers would implement SEL programs with more fidelity, and student SEL competencies would improve, behavior problems would decrease, and students' academic performance would increase. By demonstrating a relationship between the SSCA, and teacher implementation and student outcomes, we could show the importance of a need assessment and quality improvement. Further, we may also be able to show which items in the SSCA account

for more variance in quality SEL implementation and student outcomes.

In addition to the aforementioned, a concurrent relationship between other measures of OC and PC would help demonstrate that the SSCA correlates well with measures that have previously been validated. For example, we could explore convergent evidence between scores on the SSCA and the Teacher SEL Beliefs Scale (Brackett et al., 2012) or the Intermediate Unit Capacity Assessment (St. Martin, Goodman, Harms, & LeVesseur, 2013). It would be predicted that the SSCA would correlate well with other tools hypothesized to measure similar constructs.

APPENDIX A

CONSENT AND ASSENT FORMS

Parent Consent Form for Participation in a Research Study
University of Massachusetts Amherst

Principal Investigator:	Cheyne LeVesseur
Faculty Sponsor:	Sara Whitcomb
Study Title:	Implementing Universal SEL Programs: Validation and Findings from the Schoolwide SEL Capacity Assessment

---

This consent form will give you the information you will need to understand why this study is being done and why you are being invited to participate. It will also describe what you will need to do to participate and any known risks, inconveniences or discomforts that you may have while participating. We encourage you to take some time to think this over and ask questions now and at any other time. If you decide to participate, you will be asked to sign this form and you will be given a copy for your records.

Primary participants recruited for this study will include approximately 100 teachers, representative of the general kindergarten through twelfth grade population. The Social and Emotional Learning Likeliness Survey (SLS) will be administered using the online survey tool, Survey Monkey, or if preferred, a paper-pencil version of the survey will be made available. Teachers will have the option to choose where and when they take the survey and should expect to take 15 minutes or less to complete. Investigators will encourage the participants to choose a setting that is comfortable and quiet, free from distractions and with adequately Internet connection. Participants will not be contacted in the future.

The purpose of the proposed project is to extend the social and emotional learning (SEL) implementation assessment literature by developing a new rating scale designed to measure implementation barriers and make inferences regarding the current state of SEL implementation across a wide range of communities and populations. The study findings will be used for the purpose of refinement of the draft version of the SLS instrument.

If you agree to take part in this study, you will be required to provide consent online, prior to completing the survey. After opening the designated URL, this consent form will be presented, and you will have the option to agree to consent; if your agree, you will be asked to read the directions, complete the demographics section, and rate each item according to the provided scale. The purpose of this survey is to identify strengths and concerns of a school community that can be used in response to student needs for improving the implementation quality of SEL programming. The type of items that will be asked of you will range from your views and perceptions of SEL to the general organizational capacity of your school. You may skip any items you feel uncomfortable answering.

A direct benefit for participation in the research study will include an improved understanding of social emotional learning implementation barriers. As a result of the study procedures, a risk for participation in this study may include the allocation and possible inconvenience of time it takes to complete the survey.

The following procedures will be used to protect the confidentiality of your study records. The researchers will keep all study records (including any codes to your data) in a secure locking file cabinet. Research records will be labeled with a code. A master key that links names and codes will be maintained in a separate and secure location. All electronic files (e.g., database, spreadsheet, etc.) containing identifiable information will be password protected. Any computer hosting such files will also have password protection to prevent access by unauthorized users. Only the members of the research staff will have access to the passwords. At the conclusion of this study, the researchers may publish their findings. Information will be presented in summary format and you will not be identified in any publications or presentations. Confidentiality will be maintained unless some law has or will be broken such as reporting child abuse and neglect.

Take as long as you like before you make a decision. We will be happy to answer any question you have about this study. If you have further questions about this project or if you have a research-related problem, you may contact the principal investigators Cheyne LeVesseur (clevess@educ.umass.edu or at 248-425-8931), or the faculty sponsor Dr. Sara Whitcomb (swhitcomb@educ.umass.edu or at 413-545-6904). If you have any questions concerning your rights as a research subject, you may contact the Linda Griffin, Associate Dean for Academic Affairs, 123 Furcolo, 413-545-6985 or lgriffin@educ.umass.edu. You do not have to be in this study if you do not want to. If you agree to be in the study, but later change your mind, you may drop out at any time. There are no penalties or consequences of any kind if you decide that you do not want to participate.

I have read this form and decided that I will participate in the project described above. The general purposes and particulars of the study as well as possible hazards and inconveniences have been explained to my satisfaction. I understand that I can withdraw at any time.

_____	_____	_____
Participant Signature:	Print Name:	Date:

By signing below I indicate that the participant has read and, to the best of my knowledge, understands the details contained in this document and has been given a copy.

_____	_____	_____
Signature of Person Obtaining Consent	Print Name:	Date:

APPENDIX B

Schoolwide SEL Capacity Assessment (SSCA)

Introduction

Social and Emotional Learning (SEL) refers to the process by which individuals acquire knowledge and skills to help navigate through life's challenges. Knowledge and skills learned include self-awareness, social awareness, recognition and self-regulation of emotions, relationship skills, empathy, and responsible decision-making. Social emotional competence has been shown to support mental health, academic performance and learning, substance abuse, antisocial behavior, and school nonattendance. Implementation quality is defined as the extent to which essential intervention components are delivered and received, producing a degree of proportionate quantity and quality in a comprehensive and consistent manner by an interventionist trained to deliver the intervention.

Overview

The SSCA is part of a planning process for schoolwide/classwide (i.e., universal) prevention practices. Within this context, universal programming is used in reference to a multilevel prevention system consisting of three levels: universal, targeted, and individualized supports. Within the universal prevention framework, all students receive prevention programming regardless of risk. Universal prevention is defined as being sequenced, active, focused, and explicit.

Purpose

The survey data collection process involves collecting and examining information about school-wide issues and then utilizing that data to determine priority goals, to develop a plan, and to allocate funds and resources for universal programming. The purpose of this survey is to identify strengths and weaknesses of a school community that can be used in response to student needs for improving the implementation quality of universal SEL programming.

Schools starting their schoolwide SEL implementation efforts and developing their capacity to support school staff may use the SSCA to gather baseline data, assess their current status, and action planning. Schools already engaged in schoolwide SEL implementation programming may use the SSCA to support capacity building, set priorities and action plans, and monitor progress and evaluate goals.

SSCA Components

Table 1

Construct Domains and Definitions

Construct Domains	**Provider Characteristics**	**Organizational Capacity**
Definition	A person's perceived need, benefits, self-efficacy, and skill proficiency for SEL programming.	A person's belief about the schools universal prevention delivery system, general organizational factors, practices/ processes, staffing, and the support system.

Table 2

Construct Components, Subordinate Elements, and Factors

Construct Domains	Factors	Subfactors
Provider Characteristics	Perceived Need for Innovation	
	Perceived Benefits of Innovation	
	Self-efficacy	
	Skill Proficiency	
Organizational Capacity	Community Level Factors	Funding
	Prevention Delivery System	General Organizational Factors
		Specific Practices and Processes
		Specific Staffing Considerations
	Prevention Support System	Training
		Technical Assistance

Intended Participants

The SSCA is completed by the majority of school staff, including teachers, support staff, and administration.

Administration of the SSCA

Scheduling

The SSCA should be completed in the spring and an action plan should be drafted within 30 days of analyzing the results. The SSCA can be completed again in the winter to assess progress and modify the action plan accordingly. It is acceptable to complete the SSCA annually, bi-annually, or quarterly, depending on need. Scheduling should be determined in advance.

Preparation

Prior to completing the SSCA, school staff should make sure the following are in place:

1. A staff member has been chosen as the liaison to lead SSCA efforts.
2. Staff have been introduced and trained on the SSCA.
3. Staff should agree on scheduling – assessment dates, data analysis and action planning timeframe.
4. Staff have direct access to the SSCA liaison for support and to clarify items and answer questions.
5. Previous data analysis and action plans are available to staff.
6. The SSCA is accessible (paper or electronic) to all staff.

The SSCA should take approximately 15 minutes to complete. However, time should be allocated to the SSCA liaison for preparation, administration, and data analysis. Further, time will need to be allocated for action planning purposes.

School SEL Capacity Assessment

Please complete the following information. Do not write your name or other identifying information on this survey. Information provided on this survey will not be associated to you. Thank you for your time.

1. Consent for Participation

☐I agree to consent

2. Please indicate the name of your school (if preferred, you can leave this item blank)

_____

3. Gender

☐Female ☐Male

4. Race/ Ethnicity

☐White ☐Black or African American ☐Asian ☐Latino/Hispanic

☐Native Hawaiian or other Pacific Islander American Indian or Alaska Native
☐Multiracial (please specify) _____

5. Years of professional school experience
☐0-1 years ☐2-4 years ☐5-7 years ☐8-10 years ☐10-15 years ☐More than 15 years

6. What is the highest level of education you have completed?
☐Haven't graduated from high school ☐GED▪ ☐High School Graduate ☐Associates
☐Bachelors ☐Masters▪ ☐Masters with Specialist Degree ☐Doctorate

7. Grade Level at which you currently work
☐Pre K ☐K▪ ☐1st ☐2nd ☐3rd ☐4th ☐5th ☐6th ☐7th ☐8th ☐9th ☐10th ☐11th ☐ 12th
☐District wide position ☐Other (please specify) _____

8. Please indicate which one best applies
☐Regular Education Teacher ☐Special Education Teacher ☐Specialist (art teacher, gym teacher, Librarian, etc.) ☐Educational Assistant/Paraprofessional ☐Other (please specify)

9. Mark only one that applies to your school
☐Low SES (i.e., <$20,000 Family Annual Income)▪
☐Mid SES (i.e., $20,000-$60,000 Family Annual Income)
☐High SES (i.e., >$60,000)

10. Mark only one that applies to your school(s)
☐City ☐Suburb ☐Town ☐Rural

11. Select one that provides the best representation of the size/population of your school (s): Please provide your best guess and if working in multiple schools, provide an overall average
☐Small (≈ 450) ☐Average (≈ 600) ☐Large (≈ 875)

Please read the following description below and answer the proceeding questions. Please provide a response to all questions.
Many universal programming efforts such as Positive Behavior Interventions and Supports may encourage the development of SEL competence; but within this context, SEL practices refer to programs and curricula designed specifically to support students acquisition of "the skills to recognize and manage emotions, develop caring and concern for others, make responsible decisions, establish positive relationships, and handle challenging situations▪ effectively" (see www.CASEL.org). Examples include but are not limited to:

- Open Circle▪
- PATHS: Promoting Alternative Thinking Strategies▪
- Positive Discipline▪

- Project ACHIEVE□
- Responsive Classroom□
- Ruler Approach□
- Second Step: Skills for Social and Academic Success□
- Strong Kids: A Social Emotional Learning Curriculum□
- 4Rs Program

12. Does your school(s) use SEL curricula (i.e. examples above or others)?:
☐Yes▪ ☐No▪ ☐Don't Know

13. Throughout your professional experience in schools, please indicate the kind of training you have had to implement SEL programs (choose as many that apply):
☐None▪ ☐1/2 day school in-service▪ ☐Full day school in-service ▪ ☐Workshop▪
☐Read relevant books▪ ☐Watched a video/TV program ☐Prior work experience▪
☐Graduate training▪ ☐On-site coaching▪ ☐Not Listed

14. Who is CURRENTLY MOST responsible for implementing a SEL curriculum in your school?
☐Regular Education Teacher▪ ☐Special Education Teacher▪ ☐Specialist (e.g., art teacher, gym teacher, Librarian, etc.) ☐School Counselor▪ ☐School Psychologist▪
☐Behavior Specialist▪ ☐Psychologist▪ ☐School Social Worker▪ ☐Educational Assistant▪
☐Administrator▪ ☐Outside Agency▪ ☐University Researcher▪ ☐Nobody▪ ☐Don't Know
☐Other (please specify) _____

15. Who SHOULD be most responsible for implementing a SEL curriculum in your school?
☐Regular Education Teacher▪ ☐Special Education Teacher▪ ☐Specialist (e.g., art teacher, gym teacher, Librarian, etc.) ☐School Counselor▪ ☐School Psychologist▪
☐Behavior Specialist▪ ☐Psychologist▪ ☐School Social Worker▪ ☐Educational Assistant▪
☐Administrator▪ ☐Outside Agency▪ ☐University Researcher▪ ☐Nobody▪ ☐Don't Know
☐Other (please specify) _____

16. Please indicate the primary setting these skills are being taught:
☐Classroom▪ ☐Small Group▪ ☐One-to-one ☐Don't Know ☐Other (please specify) _____

17. Please indicate the primary setting you BELIEVE these skills should be taught:
☐Classroom▪ ☐Small Group▪ ☐One-to-one ☐Don't Know ☐Other (please specify) _____

18. Please indicate which one best reflects your current practices:

☐I never plan to use an SEL programs in this school.▪

☐I have thought about using an SEL program, but have not taken any steps to start a program.

☐I will use an SEL program soon.▪

☐I am using an SEL program, but not on a regular basis.▪

☐I am using an SEL program but only use certain components.▪

☐I am using an SEL program exactly as described in the program materials/instructions.

Using the multiple choices below, identify the answer that best represents your on-the-spot belief about each statement or question. For each statement, rate how strongly you agree or disagree by placing a check mark in the appropriate box. Place each statement within your school context. Please provide a response to all questions.

Instructions for questions 19-29: These statements express some beliefs about children's emotional development. Because children's abilities may develop over time, please consider the age level that you teach and respond to these statements for children of that age.

19. It is important for teachers to help children learn how to IDENTIFY their own emotions.

☐Strongly Agree▪ ☐Agree▪ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree▪

20. I believe it is part of the teacher's job to teach children to UNDERSTAND the emotions of others.

☐Strongly Agree▪ ☐Agree▪ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree▪

21. It is important for teachers to help children MANAGE their own emotions.

☐Strongly Agree▪ ☐Agree▪ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree▪

22. I believe using a SEL program improves the likelihood of students learning how to appropriately manage emotions IN THEIR OWN WAY.

☐Strongly Agree▪ ☐Agree▪ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree▪

23. I believe it is part of the teacher's job to teach children to CONSIDER the emotions of others.

☐Strongly Agree▪ ☐Agree▪ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree▪

24. I believe it is part of the teacher's job to teach children about optimism.
☐Strongly Agree ☐Agree ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree

25. I believe it is part of the teacher's job to teach children how to manage distress and other upsetting feelings (e.g., anger, anxiety, sadness, and shame).
☐Strongly Agree ☐Agree ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree

26. I believe parents place responsibility on teachers to teach children how to manage their emotions.
☐Strongly Agree ☐Agree ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree

27. I worry about how parents will react to school-wide SEL programming in schools.
☐Strongly Agree ☐Agree ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree

28. School-wide SEL programming is relevant to my schools needs.
☐Strongly Agree ☐Agree ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree

29. Only a small group of students in my school(s) need SEL programming.
☐Strongly Agree ☐Agree ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree

Please identify the answer that best represents your on-the-spot belief about each statement regarding school-wide SEL programming (e.g. classroom teachers are asked to take some responsibility for SEL instruction). Place each statement within your school context. Please provide one response to each question.

30. My school allots regular prep time for SEL program planning.
☐Strongly Agree ☐Agree ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree

31. My school allots time on a regular basis to implement SEL lessons.
☐Strongly Agree ☐Agree ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree

32. My school allots sufficient time to implement SEL lessons.
☐Strongly Agree ☐Agree ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree

33. My school strongly values the use of school-wide SEL programming.
☐Strongly Agree・ ☐Agree・ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree・

34. I strongly value the use of school-wide SEL programming.
☐Strongly Agree・ ☐Agree・ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree・

35. I am confident I have the skills to deliver a SEL program.
☐Strongly Agree・ ☐Agree・ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree・

36. I would prefer to use my own resources (no curriculum) to teach emotion awareness.
☐Strongly Agree・ ☐Agree・ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree・
・
37. I am confident that I can build students' emotion awareness.
☐Strongly Agree・ ☐Agree・ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree・

38. I would prefer helping children to learn how to manage emotions with the aide of a structured SEL curriculum.
☐Strongly Agree・ ☐Agree・ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree・

39. My school has funding to purchase SEL curriculum/materials for teachers/staff.
☐Strongly Agree・ ☐Agree・ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree・

40. My school has ENOUGH funding to effectively support SEL programming (e.g., training, ongoing support).
☐Strongly Agree・ ☐Agree・ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree・

41. Among staff at my school, there is a shared sense of responsibility to foster students' SEL skill development.
☐Strongly Agree・ ☐Agree・ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree・

42. School staff is united regarding the value and purpose of school-wide SEL programming.
☐Strongly Agree・ ☐Agree・ ☐Somewhat Agree

☐Somewhat Disagree ☐Disagree ☐Strongly Disagree▪

43. The principal of my school(s) is an advocate of school-wide SEL programming.
☐Strongly Agree▪ ☐Agree▪ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree▪

44. My district supports school-wide SEL initiatives.
☐Strongly Agree▪ ☐Agree▪ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree▪

45. There is excellent leadership regarding setting priorities for school-wide SEL programs.
☐Strongly Agree▪ ☐Agree▪ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree▪

46. There is an internal advocate for school-wide SEL programming.
☐Strongly Agree▪ ☐Agree▪ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree▪

47. My principal plays a key role in implementing school-wide SEL programs.
☐Strongly Agree▪ ☐Agree▪ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree▪

48. My principal helps communicate a shared vision for schoolwide SEL programming.
☐Strongly Agree▪ ☐Agree▪ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree▪

49. I have received an adequate amount of training about SEL.
☐Strongly Agree▪ ☐Agree▪ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree▪

The following are general statements regarding your school and do not have to be related to SEL programming. Specifically, new programming refers to any program, whether they are SEL related or not.

50. The number of students in my classroom make it easy to do a "non-academic" activity.
☐Strongly Agree▪ ☐Agree▪ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree▪

51. Staff at my school tend to work together to resolve disagreements.
☐Strongly Agree▪ ☐Agree▪ ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree▪

52. Staff feel encouraged to communicate openly with each other.

☐Strongly Agree· ☐Agree· ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree·

53. My school is open to change in regard to new programming.
☐Strongly Agree· ☐Agree· ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree·

54. My school encourages risk-taking in regard to new programming.
☐Strongly Agree· ☐Agree· ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree·

55. New programs are integrated into existing practices and routines.
☐Strongly Agree· ☐Agree· ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree·

56. My school plans strategically for new programming.
☐Strongly Agree· ☐Agree· ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree·

57. In my school, ongoing coaching or consultation would be provided as new practices are being implemented.
☐Strongly Agree· ☐Agree· ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree·

58. New staff is trained in on-going initiatives.
☐Strongly Agree· ☐Agree· ☐Somewhat Agree
☐Somewhat Disagree ☐Disagree ☐Strongly Disagree·

APPENDIX C

LIST OF ACRONYMS

CASEL	Collaborative for Academic, Social, and Emotional Learning
CFA	Confirmatory Factor Analysis
EFA	Exploratory Factor Analysis
MDE	Major Depressive Episode
MiBLSi	Michigan's Integrated Behavior and Learning Support Initiative
MnSq	Mean Square Statistics
MTSS	Multi-tiered System of Support
OC	Organizational Capacity
PCA	Principal Components Analysis
PC	Provider Characteristics
PBIS	Positive Behavioral Interventions and Supports
RTI	Response to Intervention
RSM	Rating Scale Model
SAFE	Sequenced, Active, Focused, and Explicit
SAM	SEL Alliance for Massachusetts
SEARS	Social-Emotional Assets and Resiliency Scales
SEL	Social-Emotional Learning
SES	Social Economic Status
SOC	Stages-of-Change
SSCA	Schoolwide SEL Capacity Assessment
TAP	Think Aloud Protocols

REFERENCES

American Educational Research Association, American Psychological Association, National Council of Measurement in Education [AERA/APA/NCME]. (2014). *Standards for Educational and Psychological Testing.* Washington, DC: American Psychological Association.

Barker, C., Pistrang, N., & Elliott, R. (2002). *Research Methods in Clinical Psychology: An Introduction for Students and Practitioners* (2nd ed.). England: Wiley.

Bjorklund, D. F., & Pelligrini, A. D. (2002). *The origins of human nature*. Washington, DC: The American Psychological Association.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences.* Mahwah NJ: Lawrence Erlbaum Association.

Brackett, M., Reyes, M. R., Rivers, S. E., Elbertson, N. A. & Salovey, P. (2012). Assessing Teachers' Beliefs About Social and Emotional Learning. *Journal of Psychoeducational Assessment, 30*, 219–236.

Bronfenbrenner, U. (1979). *The Ecology of Human Development: Experiments by Nature and Design*. Cambridge, MA: Harvard University Press.

Brown, T. A. (2006). *The common factor model and exploratory factor analysis. Confirmatory factor analysis for applied research*. NY: The Guilford Press.

Buchanan, R., Gueldner, B. A., Tran, O. K., & Merrell, K. W. (2009). Social and emotional learning in classrooms: A survey of teachers' knowledge, perceptions, and practices. *Journal of Applied School Psychology, 25*, 187-203.

Carter, A. S., & Briggs-Gowan, M. J. (2004). Assessment of young children's socio-emotional development and psychopathology: recent advancements and recommendations for practice. *Journal of Psychology and Psychiatry*, 45, 109-134.

Coie, J., Watt, N., West, S., Hawkins, J., Asarnow, J., Markman, H., Ramey, S., Shure, M., & Long, B. (1993). The science of prevention: A conceptual framework and some directions for a national research program. *American Psychologist, 48*, 1013–1022.

Collaborative for Academic, Social, and Emotional Learning. (2003). Safe and sound: An educational leader's guide to evidence-based social and emotional learning programs. Retrieved from http://casel.org/wp-content/uploads/1A_Safe__Sound-rev-21.pdf

Collaborative for Academic, Social, and Emotional Learning. (2007). *Background on Social and Emotional Learning*. Retrieved March, 2013, from http://www.eric.ed.gov/PDFS/ED505362.pdf

Collaborative for Academic, Social, and Emotional Learning. (2011). Leading an SEL school: Steps to implement social and emotional learning for all students. Retrieved March, 2013, from http://casel.org/wp-content/uploads/Leading-an-SEL-School-EDC1.pdf

Coie, J., Watt, N., West, S., Hawkins, J., Asarnow, J., Markman, H., & Long, B. (1993). The science of prevention: A conceptual framework and some directions for a national research program. *American Psychologist*, *48*, 1013–1022.

Coll, C. G. (2004). The interpenetration of culture and biology in human development. *Research in Human Development, 1*, 145-159.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin Company.

de Ayala, R. J. (2009). The one-parameter model. In R. J. de Ayala, *The theory and practice of item response theory* (pp. 11-38). New York, NY: The Guilford Press.

Denham, S.A., Ji, P., & Hamre, B. (2010). Compendium of preschool through elementary school social-emotional learning and associated assessment measures. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning and Social and Emotional Learning Research Group, University of Illinois at Chicago.

Doll, B., & Lyon, M. (1998). Risk and resilience: Implications for the delivery of educational and mental health services in schools. *School Psychology Review, 27*, 348–363.

Domitrovich, C., & 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*, 193-221.

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

Durlak, J., Weissberg, R., Dymnicki, A., Taylor, R., & Schellinger, R. (2011). The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions. *Child Development, 82*, 405–432.

Elder, J. P., Ayala, G. X., & Harris, S. (1999). Theories and intervention approaches to health-behavior change in primary care. *American Journal of Preventive Medicine 17*, 275-284.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines* (3rd ed.). Boston: Pearson.

Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). Implementation research: A synthesis of the literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Fuchs, D., Mock, D., Morgan, P., & Young, C. (2003). Responsiveness-to-Intervention: Definitions, evidence, and implications for the learning disabilities construct *Learning Disabilities Research & Practice, 18*, 157–171.

Greenberg, M. T., Weissberg, R. P., Utne O'Brient, M. T., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist, 58.* 466–474..

Haggerty, K., Elgin, J., & Woolley, A. (2011). Social-Emotional Learning Assessment Measures for Middle School Youth. Seattle: Social Development Research Group, University of Washington Commissioned by the Raikes Foundation.

Hall, G. L., & Loucks, S. A. (1977). A developmental model for determining whether the treatment is actually implemented. *American Educational Research Journal, 14*, 263-276.

Hardle, W., & Simar, L. (2003). *Applied multivariate analysis*. Berlin Heidelberg: Springer-Verlag.

Heppner, P. P, Kivlighan, D. M., & Wampold, B. E. (1992). Major research designs. In C. Verduin (Ed*.), Research design in counseling* (pp. 115-165). Pacific Grove, CA: Brooks-Cole.

Kauffman, J. M. (1999). How we prevent the prevention of emotional and behavioral disorders. *Exceptional Children, 65*, 448-468

Kazdin, A., Kraemer, H., Kessler, R., Kupfer, D., & Offord, D. (1997). Contributions of Risk-Factor Research to Developmental Psychopathology. *Clinical Psychology Review*. 17, 375-406.

Lichstein, K. L., Riedel, B. W., & Greive, R. (1994). Fair tests of clinical trials: A treatment implementation model. *Advances in Behaviour Research and Therapy 16*, 1-29.

Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement, 3*, 103-122.

Linacre, J. M. (2002). Optimizing Rating Scale Category Effectiveness. *Journal of Applied Measurement,* 3, 85-106.

Linacre, J. M. (1999). Rasch advantage over other conventional methods. Retrieved March, 2013, from http://www.winsteps.com/cgi-local/forum/Blah.pl?b-cc/m-1281500582/

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transaction, 7*, 328.

Linacre, J. M. (2014). Winsteps Rasch Measurement Computer Program User's Guide. Beaverton, Oregon: Winsteps.com

Linacre, J. M., & Wright, B. D. (1994). Chi-square fit statistics. *Rasch Measurement Transactions*, 8, 350.

Merikangas, K., He, J., Brody, D., Fisher, P., Bourdon, K., & Koretz, D. (2010). Prevalence and Treatment of Mental Disorders Among US Children in the 2001 2004 NHANES. *Journal of the American Academy of Pediatrics*, *125*, 75-81

Merrell, K. (2000). Linking prevention science and social and emotional learning: The Oregon resiliency project. *Psychology in the Schools, 47*, 55-70.

Merrell, K., Parisi, D., & Whitcomb, S. (2007). *Strong Start: A Social and Emotional Learning Curriculum*. Baltimore, MD: Paul H. Brooks Publishing Co.

Messick, S. (1989). Validity. In R. L. Linn, *Educational Measurement* (pp. 13-103). New York: American Council of Education and Macmillan.

Miller, T., & Hendrie, D. (2008). *Substance Abuse Prevention Dollars and Cents: A Cost-Benefit Analysis,* DHHS Pub. No. (SMA) 07-4298. Rockville, MD: Center for Substance Abuse Prevention, Substance Abuse and Mental Health Services Administration.

Morrison, G., Brown, M., D'incau, B., O'Farrell, S., & Furlong, M. (2006). Understanding resilience in educational trajectories: Implications for protective possibilities. *Psychology in the Schools*, 43, 19-31.doi:10.1002/pits.20126.

Mrazek, P. J., & Haggerty, R. J. (1994). *Reducing risks for mental disorders: Frontiers for prevention research* Washington, DC: National Academy Press.

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

Oliveira, S., Fernandes, D., & Sisto, F. (2014). Analysis of the School Anxiety Inventory in Brazil using the Rasch Rating Scale model. *Psychological Reports*,*115*, 165-178.

Payton, J., Weissberg, R. P., Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Schellinger, K. B., & Pachan, M. (2008). *The positive impact of social and emotional learning for kindergarten through eighth-grade students: Findings from three scientific reviews.* Chicago, IL: Collaborative for Academic, Social, and Emotional Learning. Retrieved from http://www.casel.org.

Pianta, R. C. (1998). Applying the construct of resilience in schools: Caution from a developmental systems perspective. *School Psychology Review, 27*, 407-417.

Prieto, G., & Delgado, A. R. (2007). Measuring math anxiety with the Rasch rating scale model. *Journal of Applied Measurement*, *8*, 149-160.

Prochaska, J. O. (2006). Moving beyond the transtheoretical model. *Addiction, 101*, 768-774.

Prochaska, J. O., & DiClemente, C. C. (1982). Transtheoretical therapy: Toward a more integrative model of change. *Psychotherapy: theory, research and practice, 19*, 276-288.

Prochaska, J. O. (2000). With Science and Service We can Survive and Thrive. In S. Stoldz & L. McCullough (Eds.), *Reconciling Empirical Knowledge and Clinical Experience* (pp. 241–252). Washington, DC; American Psychological Association.

Rasch, G. (*1960/1980*). *Probabilistic models for some intelligence and attainment tests* Chicago, IL: The University of Chicago Press.

Rones, M. & Hoagwood, K. (2000). School-based mental health services: A research review. *Clinical Child and Family Psychology Review, 3*, 223-241

Sanetti, L. M., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: introduction to the special series. *School Psychology Review, 38*, 445-459.

Schulte, A. C., Easton, J.E., & Parker, J. (2009). Advances in treatment integrity research: Multidisciplinary perspective on the conceptualization, measurement, and enhancement of treatment integrity. *School Psychology Review, 38*, 460-475.

Seligman, M., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist, 55*, 5-14.

Severson, H. H., Walker, H. M., Hope-Doolittle, J., Kratochwill, T. R., & Gresham, F. M. (2007). Proactive, early screening to detect behaviorally at-risk students: Issues, approaches, emerging innovations, and professional practices. *Journal of School Psychology*, *45*, 193-223.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Company.

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

Snyder, T. D., & Dillow, S.A. (2012). Digest of Education Statistics Retrieved March, 2013, from http://nces.ed.gov/pubs2012/2012001.pdf

Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). The think aloud method: A practical guide to modeling cognitive processes. Retrieved March, 2013, from ftp://akmc.biz/ShareSpace/ResMeth-IS-Spring2012/Zhora_el_Gauche/Reading %20Materials/Someren_et_al-The_Think_Aloud_Method.pdf

St. Martin, K., Goodman, S., Harms, A., & LeVesseur, C. A. (2013). Intermediate Unit Capacity Assessment. Unpublished instrument, Michigan Department of Education, Lansing, MI.

Strein, W., Hoagwood, K., & Cohn, A. (2002). School psychology: A public health perspective. *Journal of School Psychology*, *41*, 23-38.

Substance Abuse and Mental Health Services Administration. (2006). Results from the 2005 National Survey on Drug Use and Health: National Findings (Office of Applied Studies, NSDUH Series H-30, DHHS Publication No. SMA 06-4194). Rockville, MD.

Substance Abuse and Mental Health Services Administration. (2009). Results from the 2008 National Survey on Drug Use and Health: National Findings (Office of Applied Studies, NSDUH Series H-36, HHS Publication No. SMA 09-4434). Rockville, MD.

Tolan, P. H., Guerra, N. G., & Kendall, P. C. (1995). A developmental-ecological perspective on antisocial behavior in children and adolescents: Toward a unified risk and intervention framework. *Journal of Consulting and Clinical Psychology, 63,* 579–584.

U.S. Government Accountability Office (2012). Children's mental health concerns remain about appropriate services for children in Medicaid and foster care. Retrieved from http://www.gao.gov/assets/660/650716.pdf

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1997). Fostering educational resilience in inner-city schools. *Children and Youth, 7,* 119–140.

Wang, R. H., Hsu, H., Lin, S., Cheng, C., & Lee, S. (2010). Risk Behaviors Among Early Adolescents: Risk and Protective Factors. *Journal of Advanced Nursing*, *66*, 313-323.

Wilson, M. (2005). *Constructing Measures: An Item Modeling Approach*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

World Health Organization (2003). Organization Investing in Mental Health. Geneva, Switzerland. Retrieved from http://www.who.int/mental_health/media/investing_mnh.pdf

Wright, B. D., & Masters, G. N. (1982). Veryfying variables and supervising measures. In Wright, B. D., & Masters, G. N., *Rating scale analysis* (pp. 90-117). Chicago, IL: MESA Press.

Yell, M. L. (2011). *The law and special education* (3rd ed.). Upper Saddle River: NJ: Pearson.

Zins, J., & Elias, M. (2006). Social and Emotional Learning: Promoting the Development of All Students. *Journal of Educational and Psychological Consultation*, *12* (2&3), 233-255.