INNOVATIONS IN STRENGTH-BASED SOCIAL-EMOTIONAL ASSESSMENT: FACTOR ANALYSIS, PSYCHOMETRIC ANALYSIS, AND CROSS-INFORMANT COMPARISONS WITH THE SEARS-T

by

BRADLEY P. COHN

A DISSERTATION

Presented to the Department of Special Education
and Clinical Sciences
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2011

University of Oregon Graduate School

Confirmation of Approval and Acceptance of Dissertation prepared by:

Bradley Cohn

Title:

"Innovations in Strength-Based Social-Emotional Assessment: Factor Analysis, Psychometric Analysis, and Cross-Informant Comparisons with the SEARS-T"

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Special Education and Clinical Sciences by:

Kenneth Merrell, Chairperson, Special Education and Clinical Sciences
Krista Chronister, Member, Counseling Psychology and Human Services
Deanne Unruh, Member, Special Education and Clinical Sciences
Philip Fisher, Outside Member, Psychology

and Kimberly Andrews Espy, Vice President for Research & Innovation, Dean of the Graduate School for the University of Oregon.

June 15, 2011

Original approval signatures are on file with the Graduate School and the University of Oregon Libraries.

An Abstract of the Dissertation of

Bradley P. Cohn for the degree of Doctor of Philosophy

in the Department of Special Education and Clinical Sciences

to be taken June 2011

Title: INNOVATIONS IN STRENGTH-BASED SOCIAL-EMOTIONAL ASSESSMENT: FACTOR ANALYSIS, PSYCHOMETRIC ANALYSIS, AND CROSS-INFORMANT COMPARISONS WITH THE SEARS-T

Approved: _____
Dr. Kenneth W. Merrell

Many youth under the age of 18 experience high levels of mental health problems, and very few of those youth receive the necessary services to combat those problems. Historically, assessment of behavior and social and emotional functioning and subsequent design of interventions occur using deficit-based measures and tools. Another method of assessing behavior and social and emotional functioning that is receiving more attention over the last decade is strength-based assessment and service delivery. The Social Emotional Assets and Resiliency Scales (SEARS) is a new multi-informant strength-based behavior-rating system currently being developed and researched at the University of Oregon. To assess the factor structure, psychometric properties, and cross-informant correlation of the teacher version of the SEARS, data were gathered from elementary, middle, and high schools throughout the United States. Teachers (n = 1673) were asked

to rate students in their classes in several domains of social and emotional functioning (e.g., problem solving, social skills, empathy, and self-regulation).

Results of the exploratory and confirmatory factor analysis indicate that the SEARS-T is made up of four strong factors—Responsibility, Self-Regulation, Social Competence, and Empathy. Analysis of reliability of total scores reflects very strong internal consistency ($\alpha = .98$) and test-retest reliability ($r = .94$). Reliability of factor scores also reflects strong internal consistency and test-retest reliability. Cross-informant reliability with the SEARS-T indicates relatively weak correlations between teacher reports and child self-reports based on the Pearson-product moment correlation ($r = .37$). Analyses of group differences were carried out for grade, student gender, rater gender, disability status, ethnicity, rater setting, and teacher categorization of academic performance. Results indicate teacher ratings differed based on student gender, disability status, rater setting, and academic performance. Results from this study indicate the SEARS-T is a psychometrically sound measure with a solid factor structure. With an understanding of the need for continued research, the SEARS-T appears to be culturally valid and useful for research and applied purposes.

CURRICULUM VITAE

NAME OF AUTHOR: Bradley Phillip Cohn

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Brigham Young University, Provo

DEGRESS AWARDED:

Doctor of Philosophy, School Psychology, 2011, University of Oregon
Bachelor of Science, Psychology, 2005, Brigham Young University

AREAS OF SPECIAL INTEREST:

Strength-Based Assessment
Social-Emotional Strength and Resilience
Juvenile Delinquency

PROFESSIONAL EXPERIENCE:

Research Assistant, Department of Special Education and Clinical Services, University of Oregon, Eugene 2008-2009

Graduate Teaching Fellow, Department of Special Education and Clinical Services, University of Oregon, Eugene 2007-2008

Research Assistant, Department of Special Education and Clinical Services, University of Oregon, Eugene 2007

Research Assistant, Department of Special Education and Clinical Services, University of Oregon, Eugene 2006-2007

Research Assistant, Department of Psychology, Brigham Young University, Provo, 2003-2005

Therapist, Oregon Social Learning Center Community Programs, Eugene 2008-2010

Treatment Skills Trainer, Oregon Social Learning Center Community Programs, Eugene, 2007-2008

Practicum Student, Child and Family Center, University of Oregon, Eugene, 2008-2009

Practicum Student, School Psychology Program, University of Oregon, Eugene, 2006-2008

GRANTS, AWARDS AND HONORS:

Wes Becker Academic Scholarship, University of Oregon, 2009
Valedictorian, Department of Psychology, Brigham Young University, Provo, 2005
Academic Scholarship, Brigham Young University, Provo, 2003-2005

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

LIST OF TABLES

Table Page

Table Page

CHAPTER I

INTRODUCTION

Mental health needs of youth under the age of 18 in American schools remain at high levels. Current estimates indicate that approximately 20 percent of youth in the United States have mental health problems significant enough to warrant intervention (Greenberg et al., 2003). Thus, about one in five youth experience significant mental health problems. In a typical classroom of 25 students, these issues burden 5 of those students. Of this impacted 20 percent, 75-80 percent *do not* receive appropriate interventions to address their needs (Greenberg et al.). As a result, nearly 15 percent of youth today experience mental health problems without receiving proper support for those problems.

Based on these estimates and population statistics from the US Census Bureau (2006) about 14.7 million youth experience mental health problems. Eleven million of those youth under the age of 18 have been identified with mental health problems but are not receiving proper support. The number of youth suffering from mental health problems without receiving appropriate services is roughly equivalent to the combined population of New York City and Chicago (US Census Bureau, 2005). In addition, of 14-17 year olds, 30 percent engage in multiple high-risk behaviors (e.g., risky sexual behavior, substance abuse, and violence) that could lead to a host of negative outcomes (Greenberg et al., 2003). Based on 2006 US Census Bureau statistics, approximately 5

million youth ages 14-17 are participating in risky behaviors. Such an abundance of youth experimenting with precarious behaviors can have a myriad of deleterious effects now and in the future, both individually and for society in general. Knowing these facts is an important first step toward intervening and compels appropriate actions to create change. One approach to creating change focuses on prevention, which indicates a need to focus efforts at earlier ages.

With respect to the educational and mental health concerns of children and youth, a paradigm shift is occurring in the professional fields that are in place to provide services, particularly in schools. For the purposes of this study, youth will refer to individuals age 12-18 and children will refer to individuals under 12 years old. Practice in some fields (i.e., school psychology) has historically been driven by the idea that assessment is a process of obtaining information to label, categorize, or place an individual in a specific setting. In school psychology, this has commonly been referred to as the "Test-and-Place Model" of service delivery. One of the characteristics of this traditional way of delivering services has been an inordinate focus on problems, pathology, or disorders, while minimizing positive characteristics, assets, or strengths. Given this state of affairs, many researchers and professional leaders are advocating a change in thinking that promotes a model focused on using information to guide intervention planning (e.g., Deno, 2002; Merrell, Ervin & Gimpel, 2006; Tilly, 2002).

Gathering information for the sake of gathering information is not particularly useful and may not fulfill ethical requirements to "do no harm" (APA, 2000; NASP, 2003). Likewise, gathering a limited amount of information simply to make a decision

may also cause more harm than good. Trends in educational and mental-health assessment and service delivery now include a more holistic and ecological assessment of children (Rhee, Furlong, Turner, & Harari, 2001; Stormshak & Dishion, 2002). Using an ecological framework provides useful information for developing and carrying out interventions. Some of that information involves identifying and understanding both the positive and negative factors working in the lives of children.

Assessment of children's social-emotional behavior has advanced considerably in the past two decades. Today many child assessment tools can accurately and reliably identify deficits, pathologies, and problems (Epstein, Harniss, Pearson, & Ryser, 1999; Merrell, 2008). These assessments include rating scales (e.g., Child Behavior Checklist, Achenbach, 2001a; Behavior Assessment Scale for Children-Second Edition, Reynolds & Kamphaus, 2004), functional behavior assessment (Gresham, Watson, & Skinner, 2001), direct behavioral observation, and self-report methods (e.g., Youth Self Report, Achenbach, 2001d; Self Report for Children, Reynolds & Kamphaus, 2004). Research indicates that many of these tools and approaches are reliable, accurate, and valid. Whether these tools are psychometrically sound and fulfill the purpose for which they were created is not being disputed. What researchers, practitioners, and policy makers are disputing is whether the purposes for which they are being used are beneficial to children and youth. That is, is identifying deficits, pathologies, and problems the most beneficial approach for working with children experiencing behavioral, social, or emotional problems? Many contend that a different approach is needed and more useful (Beaver; 2008; Epstein, 1999).

Included in those questioning the use of deficit-based assessment was the U.S. Department of Education: "The U.S. Department of Education's (1994) *National Agenda for Achieving Better Results for Children and Youth with Serious Emotional Disturbance* called for a *strengths-based* approach to assessment" (Epstein, Dakan, Oswald, & Yoe, 2001, p 153). Research shows that programs and practices stemming from a problem-focused paradigm have infused school and community based programs. Research on the outcomes of these programs indicates disappointing and at best mixed results (Leffert et al., 1998).

Work in positive psychology also postulates that even if problem-centered programs worked and ameliorated the problems, *the nonexistence of a problem does not necessarily signify wellness*. According to Jimmerson, Sharkey, Nyborg, and Furlong, 2004, p. 9, "Wellness is more than the absence of disease symptoms." Suldo and Shaffer (2008) purported the idea that there are multiple levels of mental wellbeing. Their findings even indicate that children who report symptoms of mental health problems and score high on measures of subjective wellbeing function better socially and physically than peers rating high mental health symptoms and low subjective wellbeing (Suldo & Shaffer). Knowing that high scores on strength-based measures can be more indicative of children's abilities to function in different areas of life shows a real need for strength-based measurement.

If policy makers, researchers, and practitioners are questioning the utility of deficit-based assessment, and if research indicates these measures are not capturing all the necessary parts of mental wellbeing, other measures are needed to achieve what

deficit-based measures do not. Thus, more measures are needed to assess strengths and positive aspects of children's lives. Strength-based measures can also provide unique information useful for building service plans and carrying out prevention and intervention programs to improve outcomes (Beaver, 2008; Epstein, 1999).

Although the idea of strength-based perspectives has been around for nearly a half-century (Caplan & Grunebaum, 1967), during the last two decades this perspective has received greater attention and research focus (Epstein et al., 2003; Jimmerson et al., 2004). It follows that less is known about strength-based assessment than problem-centered assessment and much more research is needed. However, research on the theory of positive psychology and positive youth development and its implications for strength-based assessment appears solid (e.g., Beaver, 2008; Benson, Scales, Hamilton & Sesma, 2006). What is needed now is much more research supporting the application of the theory (Cox, 2006; Epstein et al., 2003). An integral component of the strength-based approach is proper assessment.

Proper assessment necessitates psychometrically sound and appropriate measurement tools. Only a few truly strength-based measures exist, and of those measures even fewer have enough research to document their psychometric quality and utility (Epstein, 1999; Epstein et al., 2003). At the present time, the majority of the few existing empirically supported strength-based measures are psychometric rating scales. Some of these measures include multiple informant systems (Epstein, 1999). Multi-rater assessment systems require an evaluation of validity and reliability for all forms of the assessment system (i.e., Teacher, student, parent). Multi-rater assessment systems add

another dimension to consider when determining the psychometric properties and utility of the measure. Namely, how do different raters compare in their ratings of an individual?

A strong research base informs about the degree to which parents and teachers agree on ratings of children's problems (Achenbach, McConaughy, & Howell, 1987; Friedman, Leone, & Friedman, 1999). In their meta-analysis of multi-rater assessment systems, Achenbach, McConaughy, and Howell (1987) examined studies reporting correlations between different raters on a number of deficit-based behavior rating scales. Overall, they found weak correlations between raters. In contrast to the established research base on psychopathology in this area, there is still a paucity of research informing the degree to which different raters agree on children's strengths (Friedman, Leone, & Friedman). In other words, the time is ripe for new research that applies the same basic methods used in Achenbach et al's influential meta-analysis of cross-informant convergence with problem behaviors, applied to positive child assets.

Only two studies have been conducted to date comparing ratings by differing informants on a strength-based measure (Friedman, Leone, & Friedman, 1999; Synhorst, Buckley, Reid, Epstein & Ryser, 2005). Results from these studies indicate that cross-informant correlations of strength-based measures may be higher than for deficit-based measures (Synhorst et al.). These results indicate that standards and findings from cross-informant studies on deficit-based measures may not apply to a high degree to strength-based measures. Because findings between deficit- and strength-based measures of

cross-informant correlations may differ, it is necessary to conduct more studies using cross-informant correlations with strength-based measures.

The current study will contribute to the field by adding to the dearth of research on strength-based measures and being one of the first studies to look at cross informant correlations for strength-based measures using teacher and self report. This study will also break new ground by conducting the first detailed psychometric analysis of the teacher version of Social-Emotional Assets and Resiliency Scales (SEARS-T; Merrell, 2008), part of a new strength-based, cross-informant social-emotional assessment system for children that is being developed by researchers at the University of Oregon. Focus will be given to examine the factor structure, reliability, and validity of the SEARS-T. Methods of evaluating the validity and reliability of a new measurement instrument take several forms and require multiple studies. This study will be a first in series of studies evaluating the psychometrics of the SEARS. Specific methods for this study will include three main areas: (a) evaluating content validity through factor analysis; (b) assessing the temporal, internal, and cross rater stability; and (c) assessing construct validity through evaluation of group differences.

In this study the following four questions were addressed:

1. Using exploratory and confirmatory factor analyses techniques with a national sample, what is the likely underlying factor structure of the teacher component of the SEARS?
2. Using Cronbach's alpha, what is the internal consistency reliability of the teacher version of the SEARS?

3. Using Pearson product-moment correlations, what is the short-term (2 week) temporal stability of teacher-rated SEARS scores?
4. What is the degree of similarity across different informants (teachers, student) who rate the social and emotional strengths and assets of a given student, using the SEARS?
5. For what types of decisions is the teacher version of the SEARS valid based on group differences according to student gender, rater gender, disability status, ethnicity, rater-setting and teacher categorization of academic performance?

CHAPTER II

LITERATURE REVIEW

This brief literature review focuses on research and book chapters published between 1982 and 2008. References were gathered from the areas of positive psychology, positive youth development, school psychology, developmental psychopathology, and clinical child psychology. Key topics addressed in this literature review include social-emotional assessment, strength-based assessment, psychometric issues in applied assessment, an expanded view of validity, behavior rating-scales, the Social-Emotional Assets and Resiliency Scale, cross-informant correlations of deficit-based measures, and cross-informant correlations of strength-based measures.

Social-Emotional Assessment

Many strategies exist to carry out social-emotional assessment of children. Primary objective methods include interviews, observations, sociometrics, and rating scales (Merrell, 2008). Each method has advantages as well as some disadvantages. Best practice in social-emotional assessment includes using a *multi-method, multi-source, multi-setting assessment* (Merrell). Such a comprehensive assessment includes using more than one information gathering technique (i.e., observations and rating scales), obtaining information from more than one person, and including more than one setting in the assessment. Information gathered in this way provides a more comprehensive view

of an individual, and may reduce error variance in comparison to relying solely on one assessment method (Merrell; Stormshak & Dishion, 2002).

Building a strong multi-method, multi-source, multi-setting assessment includes the use of multi-informant rating scales. Multi-informant rating scales form a more ecologically sound assessment than what is attainable through single informant rating scales alone (Merrell, 2008). In addition to obtaining differing points of view from different people, multi-rater assessments help to acquire information from multiple settings (e.g., school, home). Obtaining information from multiple settings and raters can increase assessment validity by acquiring information in a more holistic and ecological manner (Beaver, 2008). Ecological assessment better informs decision-making about individuals and interventions because information is attained concerning an individual's functioning in multiple areas from multiple sources (Stormshak & Dishion, 2002). Currently, numerous multi-informant rating scales exist to assess social-emotional functioning.

An overwhelming majority of social-emotional assessment tools are designed to identify deficits or problems. Although many of these tools have proven effective for that purpose, research indicates that deficit-based assessment is not the sole or best option (Suldo & Shaffer, 2008). Recent research findings support the use of strength-based assessment of social-emotional functioning. Despite the current research supporting strength-based assessment, when assessing the field, Rhee et al., (2001) still identified a dearth of empirical strength-based measures. Since Rhee et al.'s review, other strength-based assessment tools have been developed but these measures are still limited, hard to

access, and require further empirical investigation (Epstein et al., 2003; Epstein et al., 1999; Merrell, 2008). Examples of existing tools include the Behavioral and Emotional Rating Scales (BERS; Epstein & Sharma, 1998), Students' Life Satisfaction Scale (SLSS; Huebner, 1991), and the Strengths & Difficulties Questionnaire (SDQ; Goodman, 2001). Adding to this small supply of strength-based measures requires both the development of new tools and further exploration of existing tools.

It should be noted that there are several multi-rater social skills or social competence assessment systems currently available that have both strong psychometric properties and have been proven useful for a variety of educational and clinical purposes. The Social Skills Improvement Instruction System (Elliot & Gresham, 1998), and the Social Behavior Scales system (Merrell, 2002) are examples of multi-rater social competence or social skills measures that are widely used and are known to be valid and to have strong technical characteristics. However, such measures focus primarily on only one domain of strength-based assessment in positive child development—i.e., social skills—and do not address the broad variety of competencies and characteristics that comprise the broader construct as it is defined for the current investigation. Therefore, social skill assessment tools and systems are not considered to be of primary interest for this study.

Strength-Based Assessment

Strength-based theory is sometimes used synonymously with positive psychology or positive youth development (Beaver, 2008; Jimerson et al., 2004). Strength-based assessment is defined as:

> The measurement of those emotional and behavioral skills, competencies, and characteristics that create a sense of personal accomplishment; contribute to satisfying relationships with family members, peers and adults; enhance one's ability to deal with adversity and stress; and promote one's personal, social, and academic development. (Epstein & Sharma, 1998, p. 3)

The four main tenets of strength-based assessment are: (1) All children and families have strengths; (2) focusing on the positive can be motivating for children and create positive changes; (3) deficiencies are an opportunity to learn skills; and (4) using strength-based plans increases client involvement (Epstein et al., 2001). These four tenets drive the process and outcomes of strength-based assessment.

Strength-based assessment does not preclude any and all discussion or assessment of problems, but provides a context in which problems are viewed (Beaver, 2008). Proponents of positive psychology highlight the usefulness and benefits of assessing both positive and negative aspects of an individual's functioning (Snyder et al., 2003). According to these authors, all individuals have degrees of positive and negative functioning and focusing strictly on the negative paints a skewed and incomplete picture of an individual (Snyder et al.). Recent research in social-emotional assessment (e.g., Suldo & Shaffer, 2008) does not suggest the discontinuance of assessing problems, rather it suggests that assessing problems is not sufficient to understand child functioning.

Work in the field of developmental psychopathology focuses on the roles that risk and protective factors, processes, and mechanisms play in the developmental trajectories of children. One of the defining characteristics of protective factors is that they either

protect against or help remediate negative outcomes. In that sense they serve as buffers against risk factors (Cicchetti, 2006; Cicchetti & Rogosch, 2002). Research on resilience also shows children can achieve positive outcomes despite facing difficult circumstances (Garmezy, 1993; Luthar & Cicchetti, 2000; Rhee et al., 2001). Understanding the risk factors at work in the lives of children is necessary to understand how these children are succeeding despite difficult circumstances. Rhee et al. also stated, "…resilience or protective factors predict outcomes better than deficit or risk factors alone" (p. 10).

Results from a recent study on subjective wellbeing indicate that conceptualizing mental health in terms of a single dimension may not appropriately capture child functioning. In order to more completely explain mental health, contemporary researchers in the positive psychology movement have proposed a dual-factor model of mental health: "In a dual-factor model of mental health, assessments of positive indicators of wellness (i.e., subjective well-being—SWB) are coupled with traditional negative indicators of illness (i.e., psychopathology) to comprehensively measure mental health" (Suldo & Shaffer, 2008, p. 52). With this model more possibilities of mental health status arise beside pathological or not pathological. Specifically, researchers (e.g., Greenspoon & Saklofske, 2001; Suldo & Shaffer) have identified four categories of mental wellbeing: High positive-low negative; high positive-high negative; low positive-low negative; and low positive-high negative. Early research on this model indicates that making these four distinctions has implications for outcomes.

Suldo and Shaffer (2008) conducted a study on the differences in outcomes of youth falling into the four different categories (High positive-low negative; high positive-

high negative; low positive-low negative; low positive-high negative). These researchers assessed differences in mean scores for the four mental health groups on self-perceptions of physical health, academics, and social functioning. Not surprisingly, results indicated that youth reporting high levels of subjective well-being and low levels of problem symptoms perceived their physical, social, and academic functioning highest. More importantly, youth reporting high levels of problem symptoms and high levels of subjective well-being performed better on measures of physical health and social adjustment than those who reported low levels of subjective well-being (Suldo & Shaffer), even when they had lower rates of problem symptoms. Based on these results, subjective wellbeing appears to affect perceived functioning, perhaps in a modulating or mediating manner.

Moreover, focusing on strengths in addition to problems provides necessary information to understand and predict behavior (Rhee et al., 2001). Concerning children with emotional and behavioral disorders, strengths constitute an important role in designing interventions (Rhee et al.). In their summary of positive youth development, Benson et al. (2006) emphasize the point that changing a child's environment can change that child's outcomes: "This is the belief that assets are enhanced when contexts and settings are configured and organized in specific ways. Context matters and contexts can be changed" (p. 909-10). A measure of a child's strengths and assets can inform practitioners as to what contextual changes—interventions—could improve that child's outcomes in one or multiple domains. Interventions are then developed for the purpose of bolstering strengths rather than or in addition to decreasing problem behavior.

Additionally, researchers in the field of positive youth development are calling for evaluations of intervention programs aimed at promoting positive youth and child development (Bernat & Resnick, 2006). In order to complete such evaluations, a strength-based measure would allow for an evaluation of the change in children's strengths and assets in addition to changes in problems and deficits. Conceptually or theoretically, this idea of using strengths to build and evaluate interventions resonates with researchers and clinicians. However, standard practice in social-emotional assessment is rooted in problem-centered assessment (Merrell, 2008). In order to move from problem-centered practice and merely a conceptual understanding of strength-based assessment to a strength-based practice, more research is necessary in strength-based social-emotional assessment.

Psychometric Issues in Applied Assessment

According to Salvia and Ysseldyke (2004) assessment is a "process of collecting data for the purpose of making decisions about individuals and groups" (p. 4). Given that decisions about children can have an important impact on their development and outcomes (e.g., school placement, diagnostic labels, type of medication), the information gathered should be accurate, complete, and useful. Obtaining accurate information from an assessment tool (e.g., rating scale, structured interview, norm-referenced test) depends in a large part on the psychometric properties of the tool. Psychometric properties include measures of reliability and validity (Anastasi, 1988). Calculations of reliability measure the *consistency of scores*—consistency across time, items, and raters (Salvia & Ysseldyke).

In the development of psychometric measures, it is common practice to assess several facets of reliability. By evaluating the consistency of a test across time, researchers and consumers can anticipate changes in scores in order to interpret those changes. Evaluating the consistency within and across items—internal consistency reliability—provides evidence that items are rated in a consistent manner. Raters will rate items intending to measure the same construct in similar ways (Salvia & Ysseldyke, 2004). Both of these measures of reliability provide strength to a measure that the scores obtained are accurate for the purpose of the measure.

For rating scales with multiple informants, reliability also includes looking at the correlations between raters. That is, to what degree do different raters of the same event or individual agree? This question addresses the importance, or lack thereof, of having multiple informants rate a single child (Achenbach, McConaughy & Howell, 1987). Another measure of psychometric strength includes estimates of validity, which measures the degree to which an assessment tool measures what it is intended to measure (Salvia & Ysseldyke, 2004). One way of identifying what exactly a rating scale measures is to look at its factor structure (Floyd & Widaman, 1995). A factor structure explains the theoretical constructs a given measurement tool measures (Floyd & Widaman; Stevens, 2002). In addition to factor structure, validity analyses include evaluation of construct validity through group comparisons.

Assessing the comparisons between different groups (i.e., student gender, grade, ethnicity) provides evidence for whether an assessment tool measures what it purports to measure or measures factors unrelated to the constructs intended:

> "When unintended consequences result from test use, an attempt should be made to investigate whether such consequences arise from the test's sensitivity to characteristics other than those it is intended to assess or to the test's failure fully to represent the intended construct" (AERA, APA & NCME, 1999, p. 23).

That is to say, when a test does not produce results for which it was intended a test is measuring other variables than those intended. Differences between groups do not mean a test is invalid, but those differences should be examined.

Part of validating an instrument requires an exploration and explanation of group differences. Some differences between groups are expected and intended. Due to limited research on strength-based measures, an example from research on measures of depression will be used to explain expected group differences. Research shows that adolescent girls have higher rates of depression than males. So differences between male and female scores on a scale intending to measure depression would be assumed. This does not mean the instrument is biased or invalid, because the results are consistent with research and theory: "Although group differences, in and of themselves, do not call into question the validity of a proposed interpretation, they may increase the salience of plausible rival hypotheses that should be investigated as part of the validation effort" (AERA, APA & NCME, p. 24). This specific study is not designed to address all aspects of validity, however, factor structure and the concept of the "expanded view of validity" will be discussed.

An expanded view of validity. In his seminal article increasing the scope of validity, Messick (1988) presented what he referred to as an expanded view of validity.

Messick asserted that validity includes but is not limited to correlations and reliability. According to Messick, validity consists of four different domains: *Construct validity, relevance/utility, value implications*, and *social consequences*. Though psychometric qualities of measurement tools (e.g., construct validity) are necessary, they are not sufficient prerequisites for good practice (Good & Jefferson, 1998; Messick). Messick contended that validity includes the usefulness of information obtained from a tool (e.g., utility/relevance), how information measured by a tool is used, and the consequences of using information obtained from a measure (e.g., social consequences, values implications). In other words, valid assessment begins, but does not end, with psychometrically valid and reliable tools. In order to identify and effectively use measures of strengths and assets, it is imperative that strength-based assessment tools follow the pretext discussed by Messick and others.

Recent research shows that strength-based assessment can address the expanded view of validity posed by Messick (see Beaver, 2008; Cox, 2006; Epstein et al., 1999; Rhee et al., 2001). Strength-based approaches claim social validity because they produce desirable effects on child trajectories and outcomes (Cox; Epstein et al.; Rhee et al.). Also, strength-based tools can inform intervention and service delivery per their utility and relevance to child development (Beaver; Rhee et al.). For example, strength-based tools are designed to assess culturally and contextually relevant strengths. Thus, the values of individuals are taken into account and interpretation of the results should lend to less pejorative consequences for individuals. Research in the area of positive youth development also speaks to the need and benefits of assessing strengths to inform and

evaluate positive interventions (see Benson et al, 2006; Bernat & Resnick, 2006). Strength-based measures not only have utility for developing interventions, but also for assessing those and other interventions. Knowing a child's strengths in addition to problems allows professionals to describe them in a way that addresses the problems while still emphasizing the child's strengths, thereby improving the social consequences of the data obtained in the assessment. Though some strengths readily present themselves, others do not emerge without appropriate assessment. One way of identifying unseen strengths is by using rating-scales.

Rating-Scales

A widespread tool used in social-emotional assessment is what is commonly referred to as the behavior rating-scale. Advantages of rating scales include being inexpensive, objective, able to assess low-rate behaviors, based on observations in the natural environment over time, and informed from persons who know the individual well (Merrell, 2008). Additionally, behavior rating-scales can fulfill best practice requirements of multi-source and multi-setting. In order to do so, rating scales often have several versions that can be filled out by different raters. This study focused on the relevance and importance of teacher ratings.

Teachers as raters. Using formal and informal assessments, teachers are often utilized to rate performance of children and youth. Many of the major behavior rating scales in practice include a teacher version (e.g., BASC, TRF, SSRS). In terms of assessment, teachers provide a unique perspective given they observe children for a large portion of the day, engaging in activities require sustained attention, and in a setting

where students are forced to navigate a social environment comprised of many other children from different backgrounds. As raters, research shows that teachers provide important and valid information concerning their students:

> As the research base and pool of knowledge regarding the accuracy, reliability, validity, social relevance, and cost effectiveness of teacher judgment of student behavior-performance continue to accumulate, it has become apparent that teachers are one of the best sources of information available regarding student performance in academic, social, and behavioral domains. (Walker & Stieber, 1991, p 4)

Not only do teacher ratings provide unique information, but research also indicates that teacher ratings can be predictive of student outcomes and needs.

Research on teacher ratings of social competence and skills provides information regarding student behaviors, abilities, and outcomes. Sha and Morgan (1996) cited research by Dodge et al. (1985) finding that teacher ratings of social situations discriminate between children in rejected and accepted social categories. In their research, Sha and Morgan found that teacher ratings of social skills were able to predict students with high or low scores on the Children's Depression Inventory (CDI). Likewise, teacher ratings are useful in predicting students at risk for behavior problems and those that are gifted (Ogden, 2003). Teacher ratings also have been found to correlate strongly with parent and self-ratings of social skills, grade point averages, and peer ratings of behaviors and abilities (Ogden, 2003). Using the Walker-McConnell Scale of Competence and School Adjustment, teacher ratings were found to correlate

strongly with school adjustment, antisocial behavior patterns, academic achievement, special education status, and arrest status (Merrell, Sanders & Popinga, 1993; Walker & Steiber, 1991). Some of these findings held true both at the time of the study and the future (Walker & Steiber).

Though a strong base of research exists to suggest that teacher ratings are informative for prediction, identification, and classification (see Gresham, Reschly, & Carey, 1987; Merrell, Sanders, & Popinga; Ogden, 2003; Shah & Morgan, 1996, Stumme, Gresham & Scott, 1982; Walker & Steiber, 1991), the majority of research has been conducted using ratings of social skills, social competence, academic functioning, emotional problems, or other problem behaviors. Drawing on these sources, it is assumed that similar benefits of teacher ratings can be extrapolated to strength-based measures. It is not the purpose of this study to explore the predictive or classification utility of teacher ratings of students' strengths. Information regarding teacher ratings is presented to provide the reader with an understanding of the role teacher ratings can play in assessment and why it is important to evaluate the psychometrics of those rating scales. Despite the strong research supporting the utility of teacher ratings, a single behavior rating scale alone does not comprise an assessment.

Assessment implies a more comprehensive measure of functioning than using a single tool or rater. Merrell (2008) provided three suggestions for how to use behavior rating scales appropriately and effectively: (1) rating scales can be used in screening processes to identify those students who may be in need of additional assessment or support; (2) several rating scales, each assessing a different domain, can be used to obtain

a multiple perspective picture of social-emotional functioning (3) rating scales can be used as a way of assessing progress during and/or after an intervention.

Distinguishing in which of these three capacities strength-based rating scales are appropriate is not the purpose of this study. Nevertheless, research supports the use of strength-based measures as part of a battery of assessment instruments (Beaver, 2008; Stormshak & Dishion, 2002). In fact, some would say strength-based measures are a necessary part of a complete assessment (Beaver; Stormshak & Dishion). This being the case, more rating scales focused specifically on social-emotional assets and strengths need to be developed and researched. One new assessment tool, the Social Emotional Assets and Resiliency Scales (SEARS; Merrell, 2008), can add to the practice and research of strength-based assessment.

Social Emotional Assets and Resiliency Scales (SEARS)

The SEARS is a strength-based assessment system in development. It consists of youth (child—SEARS-C and adolescent—SEARS-A), parent (SEARS-P), and teacher (SEARS-T) reports. Three versions of the SEARS represent a unique feature in strength-based measures; accordingly, strength-based rating scales with three informants needs more research to evaluate their usefulness (Synhorst et al., 2005). For the purposes of this study, the teacher and self-report versions were used. Focus is given to validation of the SEARS-T while the SEARS-C will be used to explore the cross-informant reliability of the SEARS assessment system. Unless otherwise specified, the term SEARS will be used in reference to the assessment system in general and the specific version being discussed will be identified. The SEARS focuses on identifying protective factors in

children's ecology that can be used to identify social and emotional needs and provide information to develop interventions addressing those needs.

The SEARS is intended to measure levels of strength in areas relating to social-emotional functioning (e.g., problem solving, social support, emotional competence, friendship, cognitive strategies, self-regulation, and social-emotional resilience). Items on the SEARS were produced from other social-emotional assessment tools and the literature on social-emotional competence. Items were then sorted into clusters, condensed to reduce repetition, and validated by six psychologists. Adaptations were made to each item in order to make them appropriate for the respective versions (teacher, student). The final item tryout version of the SEARS-T consists of 54 items and the SEARS-C 52 items. As previously stated, this study will focus specifically on evaluating validity and reliability of the SEARS-T. Currently, all other versions of the SEARS are being normed, including investigation of factor structure, using a nationally representative sample.

Preliminary research indicates strong internal consistency reliability for the SEARS-T and possible group differences based on gender, grade, and teacher-perceived levels of academic performance (Endrulat, Tom, Ravitch, Wesley, & Merrell, 2010; Felver-Gant & Merrell, 2009). Preliminary research conducted on the self-report and parent-report version of the SEARS also indicates strong internal consistency reliability and possible grade differences (Cohn, Merrell, Felver-Grant, Tom, & Endrulat, 2009; Tom, Merrell, Endrulat, Cohn, & Felver-Gant, 2009). As a valid and useful strength-based measure, the SEARS may play a key role in moving toward a strength-based

approach to social-emotional prevention and intervention. Although it appears to show initial promise, the SEARS needs to display psychometric soundness prior to use in formal assessments.

Demonstrating the psychometric qualities of a new instrument requires appropriate research (Gregory, 2004). For the SEARS, two important elements of research that have not yet been conducted include specific reliability and validity research on the teacher version (SEARS-T) and research on cross-informant comparisons of scores within the same child sample. Reliability and validity were discussed previously. To date, other studies of cross-informant correlations or convergence have used deficit-based measures, have not included self-reports in the analysis, or have been limited in sample size (Achenbach, McConaughy, and Howell, 1987; Epstein et al., 1999; Synhorst et al., 2005). By identifying the unique information gained using multiple informants, cross-informant research also can demonstrate the utility of different rating forms (Achenbach, McConaughy, & Howell).

Cross-Informant Correlations of Deficit-Based Measures

In their highly influential meta-analysis of multi-rater assessment systems, Achenbach, McConaughy, and Howell (1987) examined studies reporting correlations between different raters on a number of deficit-based behavior rating scales. Overall, they found weak correlations between raters. Average correlations between different raters were .28; average correlations between self-report and other raters were .22. Based on ecological theory, weak correlations between raters may be expected as problem behaviors often manifest differently in different contexts—in the home context parents

may witness their child acting dissimilar to what the teacher witnesses at school (Achenbach, McConaughy, & Howell). From these findings several important points should be noted.

First, agreement between different raters of problem behavior is generally low. (Achenbach, McConaughy, & Howell, 1987). Second, behavioral theories have articulated that behavior, problem behavior in particular, is often context specific (Kazdin, 1978; Merrell, 2008). For example, school and home are different settings with different expectations, structure, activities, and people. Children are likely to act differently at home than at school, which leads to different ratings of behavior by parents and teachers. (Achenbach, McConaughy, & Howell; Synhorst et al., 2005). Third, given that different raters observe different behaviors, multiple raters should be used to explain and understand behavior (Achenbach, McConaughy, & Howell; Merrell). Though these findings by Achenbach, McConaughy, and Howell discuss the correlations of problem behavior across settings and raters, results from studies of cross-informant correlations using strength-based measures indicate significant differences.

Cross-Informant Correlations of Strength-Based Measures

Cross-informant correlation research using strength based measures appear to differ from problem-based measures. Simply put, correlations between raters of children's social-emotional assets and strengths are higher (Friedman, Leone, & Friedman, 1999; Synhorst et al., 2005). Teacher-parent correlations using a strength-based measure, the Behavioral and Emotional Rating Scale (BERS) were .20-.67 (Friedman, Leone, & Friedman). Synhorst et al. found parent-child correlations from

.50-.63 on the BERS-2. Given these differences between strength-based and deficit-based cross-informant correlations several measurement and theoretical points arise.

First, higher correlations of raters on strength-based measures may indicate that "behavioral and emotional strengths may not necessarily be situationally specific" (Synhorst et al., 2005, p. 8), or perhaps not *as* situationally specific as problem behaviors. Second, from a measurement perspective this finding suggests that when validating a new strength-based measure, cross-informant correlations solely from deficit-based studies (e.g., Achenbach, McConaughy, & Howell, 1987) are not appropriate comparisons. Third, more research on cross-informant correlations of strength-based measures is needed using different samples and other informants (Epstein et al., 1999; Synhorst et al.).

Summary and Conclusions

In the field of social-emotional assessment a call has gone out for more research in the area of strength-based assessment. A few researchers have heeded this call (Epstein, 1999; Goodman, 2001; Rhee et al., 2001), but more research and development work is needed. Research on the SEARS provides not only more support for the scientific foundations and practical applications of strength-based measures, but also unique information given the use of teacher-student correlations. Continuing to use strength-based measures, specifically the SEARS, without establishing validity would not be considered best practice. In addition, having a three-pronged assessment system is only useful if each of the rating forms is considered valid. Thus, it is necessary to evaluate the teacher version of the SEARS, for the benefits of teacher ratings discussed

earlier in this chapter are only true if the instrument teachers use is psychometrically sound (Ogden, 2003). Because few tools exist to assess children's strengths, further study on the SEARS is a logical and needed step to advance the field of social-emotional assessment.

CHAPTER III

METHOD

Data for this dissertation were gathered from three separate but related studies: The SEARS assessment system national norming study, a test-retest study for the SEARS-T, and a teacher-student cross-informant correlation study. The general methods for each study were similar. Differences across the general methods for each study are explained.

Sample

Teacher-based student rating data were collected from 23 elementary, middle, and high schools from Massachusetts, Hawaii, Oregon, Washington, Illinois, Iowa, North Carolina, Georgia, and California. More than 300 teachers rated a total of 1673 students. This sample comprised the expanded national norming sample for the SEARS-T, from which the normative data will ultimately be derived. Demographic information describing participating teachers and students they rated for the national norming sample is presented in Table 1. Raters' years of experience ranged from 0 to 50 with a mean of 15.

Table 1

Demographic Characteristics of Participating Teachers and Rated Students of the National Norming Sample

Variable	*% of Sample*
Student grade	
Kindergarten	7.4
First	6.2
Second	8.7
Third	9.5
Fourth	9.7
Fifth	9.1
Sixth	5.1
Seventh	6.5
Eighth	6.0
Ninth	6.1
Tenth	8.3
Eleventh	9.7
Twelfth	7.7
Student gender	
Female	49.5
Male	48.8
Student ethnicity	
White/Caucasian	49.0

Table 1 (continued)

Hispanic/Latino	18.6
Black/African American	19.1
Asian/Pacific Islander	8.1
American Indian/Native American	.2
Multiracial	3.1
Other	1.2
Special education status	
Special education	17.7
No special education	79.8
Class/setting of rater	
General education classroom	72.1
Special education classroom	7.2
Other teaching setting	11.1
Non-teaching student support	5.3
Other	2.9
Rater perception of student academic performance	
Below average	23.3
Average	42.9
Above average	29.9
Rater gender	
Female	72.7
Male	25.5

Participants in the test-retest sample were comprised of 30 teachers from two elementary schools in Vancouver, Washington who filled out a total of 118 SEARS-T forms on students in their classrooms. Each teacher was asked to rate five students. Table 2 presents demographic characteristics of the teachers and the students that were rated. Teacher years of experience ranged from 1 to 40 with a median of 12 and a mean of 15.

Table 2

Demographic Characteristics of Participating Teachers and Rated Students in the Test-Retest Sample

Variable	*% of Sample*
Student grade	
Kindergarten	15.3
First	12.7
Second	16.9
Third	17.8
Fourth	16.9
Fifth	20.3
Student gender	
Female	45.8
Male	53.4
Student ethnicity	
White/Caucasian	55.9
Hispanic/Latino	11.0

Table 2 (continued)

Black/African American	20.3
Asian/Pacific Islander	4.2
American Indian/Native American	0.0
Multiracial	8.5
Other	0.0
Special education status	
Special education	22.0
No special education	78.0
Class/setting of rater	
General education classroom	70.3
Special education classroom	.8
Other teaching setting	26.3
Non-teaching student support	2.5
Other	0.0
Rater perception of student academic performance	
Below average	40.7
Average	41.5
Above average	16.9
Rater gender	
Female	85.6
Male	14.4

Participants in the cross-informant correlation sample included parents, teachers, and students from an elementary school in Eugene, Oregon. Parent forms were distributed by mail to 430 parents. Only 151 parents completed and returned forms, for a return rate of 35%. Teacher forms were distributed to 12 teachers in kindergarten, first, and second grade, and 8 teachers in fourth, fifth, and sixth grade. Each teacher in kindergarten through second grade was asked to complete five SEARS-T rating forms. Each teacher in 3rd through 5th grade was asked to fill out 15 forms. Four K to 2nd grade teachers returned forms for a total of 20 forms. Three 3rd to 5th grade teachers returned forms for a total of 45 forms. Because the child self-report form of the SEARS (SEARS-C) is appropriate for use beginning with grade 3, teachers in 3rd through 5th grade were given student forms for each of their students for a total of 211 possible student participants. Fifty-nine students in 3rd to 5th grade participated. Due to a lack of participation and a small number (8) of parent, teacher, and student forms that matched, the researcher was only able to correlate student and teacher ratings. The final sample of teachers and students was 31. Table 3 describes the demographics of the teacher and student cross-informant comparison participants. Teacher raters' years of experience ranged from 6 to 28 with a mean of 22.

Table 3

Demographics of Raters and Rated Student Participants of the Cross-Informant Correlation Sample

Variable	*% of Sample*
Student grade	
Third	51.6
Fourth	41.9
Fifth	6.5
Student gender	
Female	58.1
Male	38.7
Student ethnicity	
White/Caucasian	67.7
Hispanic/Latino	0.0
Black/African American	16.1
Asian/Pacific Islander	9.7
American Indian/Native American	3.2
Multiracial	3.2
Other	0.0
Special education status	
Special education	6.5
No special education	93.5
Class/setting of rater	

Table 3 (continued)

General education classroom	87.1
Special education classroom	0.0
Other teaching setting	12.9
Non-teaching student support	0.0
Other	0.0
Rater perception of student academic performance	
Below average	28.6
Average	32.1
Above average	39.3
Rater gender	
Female	61.3
Male	38.7

Instruments

One cross-informant behavior rating scale system was used in this study—The Social-Emotional Assets and Resiliency Scales (SEARS). The SEARS is a recently developed experimental measure, designed to provide a cross-informant strength-based measure of social and emotional functioning. Within the SEARS system focus is placed on a youth's ability to cope with difficulties, be optimistic in-spite of adversity, solve problems, interact socially, and develop and maintain friendships. Informants consist of parent, teacher, and youth. The child version (SEARS-C) is appropriate for students in grades 3-6. An adolescent version (SEARS-A) is also available for use with students in

grades 6-12 but was not used in this study. The parent (SEARS-P) and teacher (SEARS-T) forms can be used for parents and teachers of students at all grade levels kindergarten through twelve. Only the SEARS-T and SEARS-C were used for purposes of this study.

The present study includes both the first national research efforts for the SEARS-T and some of the first research efforts outside of ongoing national normative data collection and standardization projects using the SEARS-T and SEARS-C.

Preliminary analyses of internal consistency reliability of the SEARS-C national norming sample over two thousand cases is very high, .92 for Cronbach's alpha coefficients. A study by Harlacher (2008) showed similar levels of internal consistency reliability with student self-report ratings on the SEARS-C with a smaller sample at two elementary schools in Springfield, Oregon. Also, the study demonstrated that the SEARS-C was highly sensitive to the effects of a social-emotional learning intervention, and that it had strong convergent construct validity with the School Social Behavior Scales, a teacher rating form for assessing social competence of students. Specific internal consistency coefficients of the SEARS-T are included in the results section of this study.

The experimental teacher version—SEARS-T—consists of 54 items asking teachers to rate students on a 4-point Likert-type scale (Never, sometimes, often, always). Examples of items on the SEARS-T include, "Thinks about his/her problems in ways that help," "Feels sorry for others when bad things happen to them," "makes friends easily." The SEARS-T research rating forms require 15-20 minutes to complete, do not ask for personally identifying information of the student or teacher, but do ask for demographic

information (i.e., child's age, grade, gender, ethnicity, special education eligibility status, teacher's gender, years of experience, setting in which student is taught, and teacher's perception of student level of academic achievement). The child research version of the self-report forms—SEARS-C—consists of 52 items asking students to rate themselves on a 4-point Likert-type scale (Never, sometimes, often, always). Examples of items on the SEARS-C include, "I am good at understanding what other people think," Even when things don't go well for me, I am okay," and "I am good at solving problems." These rating forms require 15-20 minutes to complete. They do not include personally identifying information, but do ask for basic demographic information (i.e., age, grade, gender, ethnicity).

Procedure

Materials for all three studies were prepared at the University of Oregon and mailed to each school, where they were then distributed to teachers and parents. A contact person was identified at the school and placed in charge of disseminating forms. Participation was voluntary and no personally identifying information was collected. Parents were sent an informational letter explaining the purpose and procedure of the study and the procedure parents could follow if they did not want their student involved. Each teacher received a packet including a letter explaining the purpose, procedure, and benefits of the study, instructions for administration, four, five, or fifteen SEARS-T forms depending on the site, and instructions for receiving a compensation honorarium. Only for the cross-informant correlational study did teachers of students in grades three through five receive 15 rating forms. Differences in the number of forms provided

teachers were a result of the purpose for each of the sites. Teachers were compensated with either $20 for filling out four forms, $30 for filling out five forms, or $100 for filling out 15 forms. Teachers gave consent to participate in the study though active participation and returning completed forms.

Teachers in the national norming sample and in the cross-informant study were asked to randomly select four, five, or fifteen students from their class roster to rate one time each. Teachers from the elementary schools for the test-retest study were asked to rate the same five students they rated at time one two weeks later, at time two. For those teachers in the retest study, student initials were used as a code to rate the same students for time one and two. After filling out the forms, teachers returned the completed rating forms to their school office. Office personnel then returned the forms to the researcher. Once received by the researchers, graduate students in the University of Oregon School Psychology program coded data using SPSS 16.0 (SPSS, 2007).

Procedures for the cross-informant comparison correlational study are explained here in more detail to explain differences from the general procedures. Parents received a letter explaining the study and the consent process for their children participating, a letter explaining the SEARS-P and how they could participate, an honorarium form, and the SEARS-P measure. Parent forms were delivered in pre-stamped envelopes on which the contact person at the school placed address labels and mailed them to the parents. In the letter, the process of having their student withdrawn from the study if they so desired, by calling the school contact person, was explained. Parents completing the SEARS-P

returned it to the researcher in a postage paid business reply envelope. Those parents who completed the SEARS-P were compensated with a $10 gift card to a local retailer.

Teacher materials were delivered on the predetermined week of administration. Materials included a letter explaining the study, an honorarium form for filling out the SEARS-T, instructions for both the SEARS-C and SEARS-T, one SEARS-C measure for each student in the classroom, and 15 SEARS-T measures for teachers of grades 3-5 and five measures for teachers of grades K-2. Teachers who completed the SEARS-T rating packets randomly selected and rated 15 or 5 students from their class roster. Teachers who participated in the study were compensated with a $100, or $30, gift card to a local retailer. Upon completion, teachers returned the teacher and student forms to the school contact person who subsequently returned all materials to the researcher.

Teachers administered the student self report form—SEARS-C—one time during class. Teachers handed out the forms to students, read instructions to the students, the students then filled out the forms and returned them to the teacher. The school received a $2 honorarium payment, to be used for supporting student activities, for every participating student along with copies of the Strong Kids social emotional learning curriculum.

To code the forms for correlation, participants were asked to write the first name of the mother of the student on the form. Teacher and student forms were matched by the code names for data analysis. Data were then coded by the researcher using SPSS 16.0 (SPSS, 2007).

CHAPTER IV

RESULTS

This chapter includes the results of the exploratory factor analysis, confirmatory factor analysis, and reliability and validity analyses. Results of these analyses are arranged in sections by type of analysis. First, the factor analysis is presented starting with an explanation of the process and results of the exploratory factor analysis and ending with an explanation of the process and presentation of the results of the confirmatory factor analysis. Next, reliability analyses are presented starting with the internal consistency reliability, then test-retest reliability, followed by inter-rater reliability. In the final portion of this chapter, the explanation and results of the validity analyses are presented. Validity analyses are divided into sections by independent variable in the following order: Grade, rater gender, student gender, academic performance, rater setting, student ethnicity, and special education status. Details of each analysis are included in the subsequent sections.

Missing Data

Data were screened for normality, range restriction, outliers, missing data, and initial communalities. Missing data appeared to be missing at random. Two strategies of handling missing data were employed. For missing answers to items, data were replaced using mode imputation. Though this method can artificially increase or decrease the

sample mean, it is a traditional and widely used method in research (Chen & Astebro, 2003). Listwise deletion was used for cases missing more than five item-values, so those cases were not included in the analysis (Widaman, 2006).

Exploratory Factor Analysis SEARS-T

To explore the validity of the underlying structure of the Social Emotional Assets and Resiliency Scales-Teacher version (SEARS-T), seven steps were used. These analyses were used to answer research question 1: "*Using exploratory and confirmatory factor analyses techniques with a national sample, what is the likely underlying factor structure of the teacher component of the SEARS*?" In all steps Principle Axis Factor (PAF) analysis was used with Oblimin oblique rotation. PAF is theoretically based and serves to explain the constructs accounting for the variance of a measure. PAF was utilized instead of principle components analysis because it is most useful for identifying latent variables rather than simply reducing the number of items (Preacher & MacCallum, 2003). All assumptions for running PAF were met.

Step 1. Using all items, the Kaiser-Guttman criterion (Floyd & Widaman, 1995; Kaiser, 1960) was used to identify the likely number of factors by keeping the items with an eigenvalue equal to or greater than 1.0. All but two communalities were above .50. Communalities indicate the percentage of variance explained by each item (variable). Thus strong communalities provide evidence that an item relates strongly to the latent factors the scale purports to measure (Floyd & Widaman, 1995). Six factors explaining 66 percent of the variance were retained; thirty-three percent of items loaded on more

than one factor. However, the content of the obtained factors was not clinically interpretable.

Step 2. Using all items, a four-factor model was forced based on Kaiser's Rule (Kaiser, 1960). All but five communalities were above .5 and none were below .42; the four factors explained 62 percent of the variance. Twenty-eight percent of items loaded on more than one factor. Only six items loaded on Factor four and five of the six double loaded on another factor. Again, the contents of the obtained factors were not interpretable.

Step 3. Using all items, a three-factor model was forced using Kaiser's Rule (Kaiser, 1960). All but six item communalities were above .5 and none were below .41; the three factors described 59 percent of the variance. Twenty-four percent of items loaded on more than one factor and once again the content of items in each of the factors was not interpretable.

Step 4. Ten items were removed from the analysis based on researcher judgment regarding possible poorer item content or loadings on multiple factors with less clear relation to any of the factors. Using the remaining 44 items, the Kaiser-Guttman criterion of eigenvalue equal to or greater than 1.0 was used (Floyd & Widaman, 1995; Kaiser, 1960). All but 3 item communalities were above .5 with the lowest being .43. Five factors explaining 65 percent of the variance were retained; twenty-three percent of items loaded on more than one factor. Factor five only contained 4 items, all of which loaded on another factor.

Step 5. Using 44 items, a four-factor model was forced based on Kaiser's Rule (Kaiser, 1960). All but 3 item communalities were above .5 with the lowest being .39. Four factors explained 63 percent of the variance. Twenty-three percent of items loaded on more than one factor. The content of the factors was not clinically interpretable, due to the appearance of multiple constructs represented in some of the factors.

Step 6. Three more items were dropped because of potentially poorer item content or loadings on multiple factors with no clear relation to any of the factors. Using 41 items, Kaiser Guttman criterion of eigenvalue equal to or greater than 1.0 was used (Floyd & Widaman, 1995; Kaiser, 1960). All but 4 item communalities were greater than .50 with the lowest being greater than .43. Five factors explaining 65 percent of the variance were retained; twenty-seven percent of items loaded on more than one factor. Factor five contained only 4 items, each of which loaded more strongly with other factors. The obtained factors were somewhat clinically interpretable, but were also lacking in specificity.

Step 7. For initial communalities, mean coefficients were .63 with only four items falling below .50—Item 11, .43, Item 26, .39, Item 40, .48, and Item 43, .48—but these were still considered acceptable. The Scree Test (Cattell, 1966) was used as an initial means of determining the number of factors to retain following step 6. According to Stevens (2002), this procedure was a valid method because the mean communality was greater than .60 and the *n* size was greater than 250. Initially, five factors were retained using the Scree Test. However, after reviewing item correlations, it was determined that only four items correlated on Factor five. SPSS 16.0 (SPSS, 2007) was then used to

force a four-factor model (Kaiser, 1960). Although the fourth factor contained only six items, four of which loaded on another factor, the factor was retained and the items also loading on other factors were placed in the fourth factor based on clinical relevance. The percent variance explained with the final four-factor model is represented in Table 4. In Table 5 it can be seen that the total percent variance explained by the three factors is 63.19, and the first factor explained the majority of the variance, 49.88%, the second factor explained 6.91%, the third factor explained, 3.8%, and the fourth factor explained 2.6% of the total variance. As will be discussed, the contents of the obtained factors were clinically relevant and clear.

Table 4

Percent of Variance Explained by Retained Factors

Factor	*% of Variance Explained*	*Cumulative %*
1	49.88	49.88
2	6.91	56.79
3	3.80	60.59
4	2.60	63.19

An Oblimin oblique rotation was used based on theory that the factors would correlate (Preacher & McCallum, 2003)—the instrument purported to measure social and emotional strengths and assets of students, and these strengths and assets were theorized not to be completely independent. This analysis yielded factor correlations up to .63, so the Oblimin oblique rotation was verified and employed instead of a VARIMAX orthogonal rotation. Presented in Table 5 are the pattern coefficients for each item. This

value was calculated by doubling the critical values for a two tailed test of correlation coefficients at $\alpha = .01$ (Stevens, 2002).

Table 5

Pattern Coefficients for the Four Factors of the Social Emotional Assets and Resilience Scales—Teacher Version with Oblimin Orthogonal Rotation

Factor	*1*	*2*	*3*	*4*
Responsibility				
36. Dependable	**.83**	.04	-.04	.03
47. I trust her/him	**.81**	-.05	.00	.12
1. Likes to do best in school	**.74**	.01	.00	.02
22. Makes good decisions	**.73**	.00	-.16	.02
38. Accepts responsibility	**.71**	-.01	-016	.05
16. Liked by teachers/students	**.69**	.16	.12	.19
31. Thinks before acting	**.69**	-.14	-.21	-.12
6. Works independently	**.65**	.14	-.11	-.20
48. Works well with others	**.64**	.17	-.06	.08
13. Good listener	**.58**	.01	-.14	.23
Social competence				
9. Comfortable talking to others	-.04	**.87**	.08	.09
20. Good at starting conversation	-.04	**.82**	-.02	.01
10. Makes friends easily	.17	**.75**	.11	.16
32. Comfortable in groups	.08	**.75**	.03	-.09
40. Good at telling stories	-.12	**.67**	-.13	-.01

Table 5 (continued)

15. Fun to be with	.18	**.66**	.07	.20
25. Comfortable telling how feels	-.14	**.64**	-.21	.01
43. Stands up for self	-.05	**.62**	-.20	-.16
14. Others ask to hang out	.15	**.61**	-.01	.20
33. Seen as leader	.30	**.60**	-.09	-.13
35. Respected by others	.49	**.47**	-.04	-.01
26. Asks for help	.22	**.39**	-.09	.08
Self Regulation				
45. Identify and change thoughts	-.08	.03	**-.87**	.07
49. Negative thoughts realistic	-.05	.04	**-.81**	.02
51. Errors in thinking	.06	.05	**-.79**	-.04
44. Calm down when stressed	.10	-.03	**-.71**	.05
53. Handle problems	.15	.09	**-.68**	-.02
28. Understands differing feelings	.01	.09	**-.62**	.25
37. Thinks of problems in helpful ways	.19	.10	**-.60**	.06
54. Doesn't let things get him/her	.18	.08	**-.58**	-.02
52. Knows how to set goals	.33	.12	**-.56**	-.16
24. Good at setting disagreements	.02	.23	**-.54**	.16
29. Stays in control when angry	.32	-.16	**-.44**	.20
11. Disagrees without arguing	.16	.03	**-.36**	.26
19. Good at solving problems	.46	.20	**-.34**	-.15

Table 5 (continued)

Empathy				
5. Understands other points of view	.27	.06	-.38	**.30**
2. Feels sorry when bad things happen to others	.24	.05	-.11	**.60**
3. Knows when others are upset	.01	.14	-.26	**.53**
12. Tries to understand how others feel	.07	.12	-.38	**.48**
30. Cares what happens to others	.35	.10	-.12	**.48**
21. Understands how others feel	.08	.24	-.36	**.39**

Note. Bold item correlations denote items that are part of the corresponding factor. Items are listed underneath each factor from largest to smallest item correlation on each factor.

Pattern coefficients in Table 5 indicate to what degree each variable correlates with the four factors. Factor 1, Responsibility, contained 10 items. Factor 2, Social Competence, included 12 items. Factor 3, Self-Regulation, included 13 items. Factor 4, Empathy, contained 6 items. The factor labels were determined based on common content of the items with which each was comprised. Factors one, two, and three showed robust evidence for being individual clusters of variables because they both had at least four variables correlating with them above .60 (Stevens, 2002). Given the smaller number of variables correlating with Factor four and the small percent of unique variance it explains, it is a less strong factor and could simply be explaining variance from specific variables instead of a cluster of variables. It was accepted as a unique factor based on the sample size being greater than 300 (Stevens, 2002) and because the clinical content of the

six items carried together strongly and was highly relevant to assessing students in school settings.

Confirmatory Factor Analysis SEARS-T

Following the Exploratory Factor Analysis, 13 items were removed from the original 54-item SEARS-T, based on results of the steps described in the previous section. To evaluate the proposed factor structure identified through the EFA on the revised 41-item SEARS-T, a confirmatory factor analysis (CFA) was carried out using AMOS 5.0 (Arbuckle, 2003). Data for the CFA were obtained by gathering a random sample of 836 cases of the 1673 cases used in the EFA. Although one could argue that pulling from the same sample used in the EFA as a cross-validation sample in the CFA limits the interpretability and generalization of the results to other samples, the effects of capitalizing on chance are lessened with a sample size greater than 1,200 (MacCallum, Roznowski, & Necowitz, 1992), making this method an appropriate choice. Prior to running the CFA the data were analyzed to evaluate statistical assumptions for performing a CFA (Kline, 2008). Model assumptions were tested for normality of distributions using visual analysis and measures of skewness and kurtosis. No outliers were found in the data and data appeared to meet assumptions of normality, so CFA was determined to be an appropriate statistical approach to analyze these data.

Through the EFA, four factors—Responsibility, Social Competence, Self-Regulation, and Empathy—were retained, with 10, 12, 13, and 6 items loading on each factor. Parameters were estimated using Maximum Likelihood (ML) estimation in AMOS 5.0 (Arbuckle, 2003). The model presented in Figure 1 converged. Standardized

estimates of structure coefficients presented in Figure 1 represent the estimated correlation between each subscale and the latent factor—social emotional assets and strengths. All observed factors appear to correlate highly with the latent variable with correlations ranging from .72 to .91. Descriptive statistics of the factor and total scores are presented in Table 6.

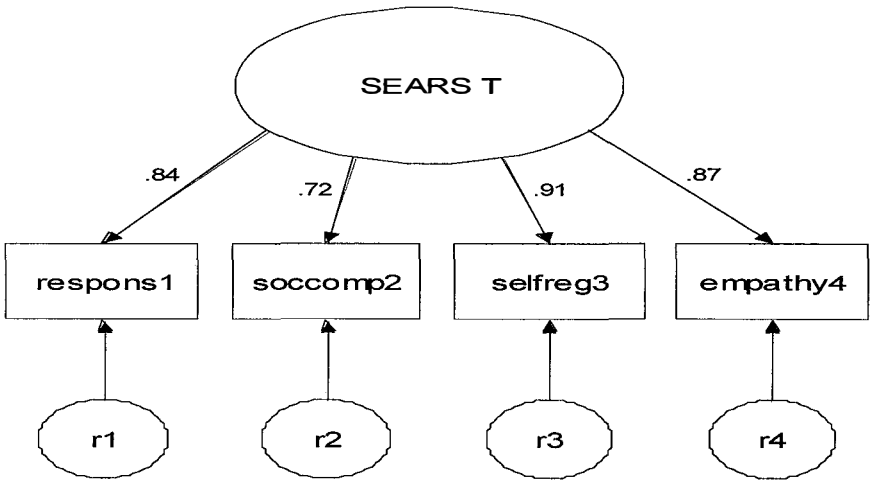

Figure 1. Confirmatory factor analysis of the SEARS-T four-factor structural model with standardized estimates

Table 6

Descriptive Statistics of the SEARS-T National Norming Sample Factor and Total Scores

Variable[a]	*M*	*SD*
Responsibility	19.22	7.34
Social competence	20.18	8.35
Self regulation	19.49	9.14
Empathy	10.07	4.28
Total score	68.97	25.83

[a]n = 1673 for each factor and total score.

Goodness of fit. Model fit was analyzed with Amos 5.0 (Arbuckle, 2003) using four goodness-of-fit indices: the chi-square (χ^2) value, the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). These indices measure how well the model represents the data (Kline, 2005). Chi-square values are a test statistic of the null hypothesis that the model fit represents the population data, so failure to reject the null ($p > .05$), or higher values, indicates good model fit (Kline). CFI values greater than .95 and RMSEA and SRMR values less than .06 indicate good model fit (Hue & Bentler, 1999). Results from the goodness-of-fit indices are presented in Table 7. The obtained GOF indices, χ^2 (2) = 7.765, p = .021, CFI = .997, RMSEA = .059, SRMR = .009 indicate the model fits the data well (Hue & Bentler).

Table 7

Summary Results of Model Fit Indices for the SEARS-T Four-Factor Model

Model	*df*	χ^2	CFI	RMSEA	SRMR
Four factor	2	7.765	.997	.059	.010

Note: Chi-square value is statistically significant, $p < .05$.

Reliability Analysis

Internal consistency reliability. Data from the SEARS-T national norming sample were used to answer research question two—"*Using Cronbach's alpha, what is the internal consistency reliability of the teacher version of the SEARS*?" Internal consistency reliability was computed using Cronbach's alpha procedure in SPSS 16.0 (SPSS, 2007) on each of the four factor scores and the summative or total score. Basic descriptive statistics were also computed, including means, standard deviations, and sample size. Alpha levels of the internal consistency reliability on each factor and total score were uniformly high–above .91. Reliability coefficients are presented in Table 8.

Table 8

Chronbach's Alpha Coefficients of the SEARS-T Factor and Total Scores

Factor score	# of items	α
Responsibility	10	.95
Social competence	12	.94
Self regulation	13	.95
Empathy	6	.92
Total score	41	.98

Test-retest reliability. Data from two elementary schools in Vancouver, Washington were used to answer research question three—"*Using Pearson product-moment correlations, what is the short-term (2 week) temporal stability of teacher-rated SEARS scores*?" Data were analyzed using SPSS 16.0 (SPSS, 2007). The Pearson product-moment correlation procedure was employed to analyze the temporal stability of the SEARS-T factor and total scores over a two-week interval. Results presented in Table 9 indicate strong test-retest reliability, with alpha's ranging from .84 to .94. The lowest reliability of .84 was on the empathy factor, which has the lowest number of items—6. Given the low number of items, lower reliability is not surprising based on greater variability with a lower number of items.

Table 9

Two-Week Test-Retest Reliability Coefficients of the SEARS-T Factor and Total Scores, For a Sample of 118 Teacher Ratings of Elementary-Age Students

	Time 1		Time 2		
Variable[a]	*M*	*SD*	*M*	*SD*	*r*
Responsibility	19.70	6.81	19.47	7.10	.92
Social competence	21.38	6.90	21.22	7.10	.92
Self regulation	18.14	8.41	18.81	8.47	.90
Empathy	10.01	3.73	10.41	3.96	.84
Total score	69.24	21.93	69.91	22.93	.94

[a]$n = 118$ for each score

Cross-informant reliability. Data from the separate sample of 31 matched elementary-age students and their teachers were used to answer research question four—"*What is the degree of similarity across different informants (teachers, student) who rate the social and emotional strengths and assets of a given student, using the SEARS*?" Pearson product-moment correlations were applied to analyze the cross-informant reliability of the total scores on teacher and student forms. Factor score correlations were not calculated because the SEARS-C consists of total score only. Descriptive statistics of the total scores are presented in Table 10. Pearson product-moment correlations for the teacher-student forms were statistically significant at $p < .05$, $r = .37$. This finding indicates that teachers and student ratings correlate weakly to moderately on students' social-emotional assets and strengths, and that there is considerable variance across raters.

Table 10

Descriptive Statistics of the SEARS-T and SEARS-C Cross-Informant Correlation Sample

Variable[a]	*M*	*SD*
Student report	73.29	15.44
Teacher report	92.65	27.26

[a] $n = 31$ for each form.

Validity Analysis: Group Comparisons

Data from the SEARS-T national norming sample were used to answer research question five—"*For what types of decisions is the teacher version of the SEARS valid based on group differences according to student gender, rater gender, disability status,*

ethnicity, rater-setting and teacher categorization of academic performance?" Data were analyzed using SPSS 16.0 (SPSS, 2007) to carry out a series of Analysis of Variance (ANOVA) and *t*-tests. Data were evaluated for statistical assumptions prior to running analyses. All statistical assumptions were met. Results are presented by group category.

Grade. To evaluate grade differences, grades were split into two groups: Kindergarten through Sixth (primary) and Seventh through Twelfth (secondary). Grades were evaluated in two groups instead of twelve because this is a traditional breakdown that is commonly used in developing behavior rating scale norms, where differing norms may be used for primary age children, and for secondary age adolescents. Mean total scores on the SEARS-T for students in primary grades and students in secondary grades were evaluated with an independent observations *t*-Test using Welch's *t'* with Satterthwaite's degrees of freedom. Students in primary grades (M = 68.34, SD = 25.71) did not have a significantly different total score than students in secondary grades (M = 68.34, SD = 25.93), $t'(1671) = 1.19$, $p > .05$. Using confidence intervals, we are 95% confident that the K-6 minus the 7-12 difference in means is between -1.07 and 3.92. Thus, we fail to reject the null hypothesis that $\mu_{Primary} - \mu_{Secondary} = 0$. Cohen's d effect size was also calculated and found to be not meaningful, $d = .05$. Results from the factor score *t*-tests were not statistically significant ($p < .05$) except for the Self Regulation factor, $p < .001$. Results with effect sizes for the factor and total score *t*-tests are presented in Table 11. In sum, the differences in SEARS-T scores based on the traditional elementary grade versus middle and secondary grade breakdowns were not significant or

meaningful overall, although older students did receive significantly higher scores in one area—the Self Regulation subscale.

Table 11

Grade Differences in SEARS-T Scores: Means and Standard Deviation of Primary Grades (n = 932) and Secondary Grades (n = 741) with t-Test Scores and Effect Sizes

SEARS-T	*Grade(n)*	*M(SD)*	*t*(1671)	*ES*
Responsibility	Primary	19.32(7.45)	-.61	.03
	Secondary	19.10(7.19)		
Social concept	Primary	20.30(8.37)	-.70	.03
	Secondary	20.02(8.34)		
Self regulation	Primary	18.61(9.25)	4.48*	.22
	Secondary	20.61(8.89)		
Empathy	Primary	10.10(4.36)	-.36	.02
	Secondary	10.03(4.19)		
Total score	Primary	69.76(25.71)	1.19	.05
	Secondary	68.34(25.93)		

Note. The distribution of total score for students in primary and secondary grades is roughly symmetrical with no severe outliers.

* $p < .05$

Rater gender. The mean total score on the SEARS-T for female and male raters was evaluated with an independent observations *t*-Test using Welch's *t'* with Satterthwaite's degrees of freedom. Female raters did not rate students significantly different (M = 68.95, SD = 26.23) than male raters (M = 69.05, SD = 24.87), $t'(782) = -.07$, $p > .05$. We are 95% confident that the male minus the female difference in means

is between -2.96 and 2.75. Cohen's d effect size was also calculated and found to be not meaningful, $d = .00$. Mean scores between male and female raters on all four factor scores were also evaluated using Welch's *t'* and there were no statistically significant differences between male and female raters ($p > .05$). Thus, we fail to reject the null hypothesis that $\mu_{\text{male raters}} - \mu_{\text{female raters}} = 0$. Effect sizes of factor and total scores are presented in Table 12.

Table 12

Rater-Gender Differences in SEARS-T Scores: Means and Standard Deviation of Females (n = 427) and Males (n = 1216) with t-Test Scores and Effect Sizes

SEARS-T	*Rater gender*	*M(SD)*	*t*(1641)	*ES*
Responsibility	Female	19.30(7.47)	.59	-.03
	Male	19.06(6.97)		
Social concept	Female	21.07(8.24)	.29	-.12
	Male	20.06(8.45)		
Self regulation	Female	19.33(9.30)	-1.29	.07
	Male	20.00(8.73)		
Empathy	Female	10.12(4.32)	.77	-.04
	Male	9.94(4.19)		
Total score	Female	68.95(26.23)	-.07	.00
	Male	69.05(24.87)		

Note. The distribution of total score for students in primary and secondary grades is roughly symmetrical with no severe outliers.

Student gender. The mean total score on the SEARS-T for females and males was evaluated with an independent observations *t*-test using Welch's *t'* with

Satterthwaite's degrees of freedom. On the total score, teachers rated female students significantly higher (M = 73.60, SD = 25.17) than male students (M = 64.41, SD = 25.80), $t'(1643) = 7.31$, $p < .05$, indicating their perceptions that girls as a group had stronger social-emotional strengths than boys. We are 95% confident that the male minus the female difference in means is between 6.72 and 11.65. Cohen's *d* effect size was also calculated and found to be small but meaningful, $d = .36$. Mean scores between males and females on all four subscales were also evaluated using Welch's *t'* and on all four factor scores teachers rated females significantly higher than males at the $p < .001$ level. Thus, we reject the null hypothesis that $\mu_{males} - \mu_{females} = 0$. Effect sizes on factor scores are presented in Table 13. This finding of female students being rated as having higher levels of social-emotional assets, competency, and resilience is consistent with prior theory and research, and will be explored in more detail in the Discussion chapter.

Table 13

Student-Gender Differences in SEARS-T Scores: Means and Standard Deviation of Females (n = 817) and Males (n = 828) with t-Test Scores and Effect Sizes

SEARS-T	*Rater gender*	*M(SD)*	*t*(1643)	*ES*
Responsibility	Female	20.66(7.06)	8.17*	.40(small)
	Male	17.77(7.33)		
Social concept	Female	21.07(8.24)	4.25*	.21(small)
	Male	19.32(8.45)		
Self regulation	Female	20.81(9.11)	5.82*	.29(small)
	Male	18.21(8.98)		

Table 13 (continued)

Empathy	Female	11.06(4.12)	9.46*	.23(small)
	Male	9.11(4.21)		
Total score	Female	73.60(25.17)	7.31*	.36(small)
	Male	64.41(25.80)		

Note. The distribution of total score for students in primary and secondary grades is roughly symmetrical with no severe outliers.

*$p < .001$

Special education status. The mean total score on the SEARS-T for student special education status was evaluated with an independent observations *t*-test using Welch's *t'* with Satterthwaite's degrees of freedom. Results of total scale *t*-test are reported in Table 14. Teachers rated students with Special Education status significantly lower (M = 54.35, SD = 22.97) than students in General Education (M = 72.14, SD = 25.30), $t'(463) = -11.76$, $p < .001$. We are 95% confident that the GENED minus the SPED difference in means is between -20.76 and 14.82. Cohen's *d* effect size was also calculated for all scores and found to be meaningful, with a medium size effect for the total score, $d = .74$. Mean scores between SPED and GENED on all four subscales were also evaluated using Welch's *t'*. Significant differences between teacher ratings of students in special education and general education were found on all of the four factor scores at the $p < .001$ level. Thus, we reject the null hypothesis that $\mu_{SPED}-\mu_{GENED} = 0$. Effect sizes of differences on the factor scores are also presented in Table 14.

Table 14

Special Education Status in SEARS-T Scores: Means and Standard Deviation of SPED (n = 1319) and GENED (n = 293) with t-Test Scores and Effect Sizes

SEARS-T	*SPED status*	*M(SD)*	*t*(1610)	*ES*
Responsibility	SPED	15.53(6.80)	-9.82*	.65(medium)
	GENED	20.04(7.18)		
Social concept	SPED	15.88(7.34)	-10.87*	.67(medium)
	GENED	21.16(8.29)		
Self regulation	SPED	14.61(7.97)	-11.14*	.69(medium)
	GENED	20.50(9.05)		
Empathy	SPED	8.33(4.07)	-7.78*	.51(medium)
	GENED	10.45(4.23)		
Total score	SPED	54.35(22.97)	-11.76*	.74(medium)
	GENED	72.14(25.30)		

Note. The distribution of total score for students in primary and secondary grades is roughly symmetrical with no severe outliers.

*$p < .001$

Disability type. Analysis of the effect of disability type on factor and total scores was not carried out. The data *n* size for each disability were extremely disproportionate; five of the eleven scales had less than 25 cases and two more had less than 50. Only one category had a sound sample size, specific learning disability. Thus to run an ANOVA would not have been appropriate. In addition, given the low *n* for most of the categories, a matched sample would not have been feasible. Without matching the sample, the results would not have been interpretable due to the possible effect of extraneous

confounding variables (e.g., gender, ethnicity, and academic performance) (Keppel & Zedeck, 1989). Statistical information regarding the differences in teacher ratings of student social emotional assets and strengths on disability groups can be found in the sections analyzing differences between students in special education versus general education and differences in the rater's setting. Further explanation of these differences and results will be given in the next chapter.

Academic performance. Data were analyzed with a one-way, between-subjects analysis of variance with the results presented in Table 15. The independent variable was teacher perceived level of student academic performance with three levels: (a) below average (n = 389), (b) average (n = 718), and (c) above average (n = 500). Teachers were asked to make this rating of perceived academic performance in the student information section of the SEARS-T research form. The dependent variable was scores on the SEARS-T factor and total scores. The effect of perceived academic performance on the total score of the SEARS-T was significant, $F(2, 1604) = 368.14$, $p <$ 001. The effect of perceived academic performance on the each of the factor scores was also statistically significant at the $p < .001$ level. As was true in the case of the findings of significant differences in SEARS-T scores based on student gender, this finding is consistent with prior research, and will be explored further in the Discussion chapter.

Table 15

One-Way, Between Subjects Analysis of Variance Summary Table for the Effects of Teacher-Perceived Level of Academic Performance on the Factor and Total Scores of the SEARS-T

Source	*df*	*SS*	*MS*	*F*
Responsibility				
Academic performance	2	2980.96	14840.48	414.13*
Error	1604	57480.31	35.84	
Total	1606	87161.27		
Social competence				
Academic performance	2	23483.33	11741.67	211.64*
Error	1604	88989.88	55.48	
Total	1606	112473.21		
Self regulation				
Academic performance	2	35844.20	17922.10	292.79*
Error	1604	134028.92	61.21	
Total	1606	134028.92		
Empathy				
Academic performance	2	4548.58	2274.29	145.16*
Error	1604	25130.70	16.67	
Total	1606	29679.28		

Table 15 (continued)

Total score				
Academic performance	2	338926.77	169463.39	368.14**
Error	1604	738355.38	460.32	
Total	1606	1077282.15		

* $p < .001$

Post-hoc comparisons of the effect of teacher-perceived academic performance on the factor and total scores were carried out using the Tukey test to evaluate all possible differences between pairs of means. The Tukey test was used to reduce familywise type I error because there were no complex comparisons to analyze, so only pairwise comparisons were evaluated (Keppel & Zedeck, 1989). Results are presented in Table 16. Comparisons between all levels of academic performance were significant on all factor and total scores ($p < .001$), with lower perceived levels of academic performance being associated with lower mean SEARS-T scores, and higher academic performance associated with higher SEARS-T scores. To further evaluate the meaning of the pairwise differences, effect sizes of each of the significant pairwise contrasts are presented in Table 17.

Table 16

Descriptive Statistics for SEARS-T Subscale and Total Scores by Teacher-Perceived Level of Academic Performance

	Below		Average		Above	
Variable[a]	*M*	*SD*	*M*	*SD*	*M*	*SD*
Responsibility	12.78[a]	6.27	19.17[b]	6.35	24.46[c]	5.12
Social competence	14.30[a]	7.13	20.45[b]	7.43	24.78[c]	7.61
Self regulation	12.47[a]	6.72	19.54[b]	7.93	25.41[c]	8.30
Empathy	7.56[a]	4.09	10.15[b]	3.83	12.13[c]	3.97
Total score	47.11[a]	20.39	69.31[b]	21.97	86.78[c]	21.24

Note. Mean values by row with different superscript letters are significantly different, $p < .001$.

[a]$n = 1509$

Table 17

Effect Size of Post Hoc Pairwise Comparisons of the effect of Perceived Academic Performance on Factor and Total Scores of the SEARS-T

Comparison	Below-Average	Average-Above	Below-Average
Responsibility	1.01(large)	.92(large)	2.04(large)
Social competence	.85(large)	.58(medium)	1.42(large)
Self regulation	.96(large)	.72(medium)	1.71(large)
Empathy	.65(medium)	.51(medium)	1.13(large)
Total score	1.05(large)	.81(large)	1.91(large)

Setting of rater. Data were analyzed with a one-way, between-subjects analysis of variance with the results presented in Table 18. The independent variable was setting of rater with five levels: (a) general education classroom (n = 1207), (b) special education classroom (n = 120), (c) other teaching setting (n = 185), (d) non-teaching setting (n = 88), and (e) other (n = 49). The dependent variable was scores on the SEARS-T factor and total scores. The effect of setting of rater on the total score of the SEARS-T was significant, $F(4, 1644) = 6.78, p < 001$. The effect of setting of rater on the each of the four factor scores was also statistically significant at the $p < .01$ level.

Table 18

One-Way, Between Subjects Analysis of Variance Summary Table for the Effects of Rater Setting on the Factor and Total Scores of the SEARS-T

Source	*df*	*SS*	*MS*	*F*
Responsibility				
Rater setting	4	1013.30	253.32	4.77*
Error	1644	87338.26	53.13	
Total	1648	88351.56		
Social competence				
Rater setting	4	1123.21	280.80	4.07*
Error	1644	113573.77	69.08	
Total	1648	114696.98		
Self regulation				
Rater setting	4	3173.69	792.92	9.70**
Error	1644	134438.55	81.78	

Table 18 (continued)

Total	1648	137610.25		
Empathy				
Rater setting	4	253.93	63.48	3.50*
Error	1644	29792.93	18.12	
Total	1648	30046.86		
Total score				
Rater setting	4	17768.63	4442.16	6.78**
Error	1644	1078000.00	655.56	
Total	1648	1096000.00		

* $p < .01$

** $p < .001$

Post hoc comparisons of the effect of rater setting on factor and total scores were carried out use the Tukey test to evaluate all possible differences between pairs of means. Again, the Tukey test was used to reduce familywise type I error because there were no complex comparisons to analyze, so only pairwise comparisons were evaluated (Keppel & Zedeck, 1989). Results are presented in Table 19. Mean comparisons on the responsibility factor indicate significant differences between special education and both general education and non-teaching settings. Mean comparisons on the social competence factor indicate significant differences between special education and three other settings: general education, other teaching setting, and non-teaching setting. Mean comparisons on the self-regulation factor indicate significant differences between general education and both special education and non-teaching setting, and non-teaching setting

and other setting. Mean comparisons on the empathy factor indicate significant differences between special education and both general education and non-teaching setting. Mean comparisons on the total score indicate significant differences between special education and all other settings. Results were significant at the $p < .05$ level. To further evaluate the meaning of the pairwise differences, effect sizes of each of the significant pairwise contrasts are presented in Table 20. The finding that students rated in special education settings had lower SEARS-T scores is consistent with prior research on social-emotional characteristics of students with disabilities, and will be explored further in the next chapter.

Table 19

Descriptive Statistics for SEARS-T Subscale and Total Scores by Rater Setting

	General Education (n = 1207)		Special Education (n = 120)		Other Teaching Setting (n = 185)		Non-Teaching Setting (n = 88)		Other (n = 49)	
Variable[a]	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Responsibility	19.68[abd]	7.33	16.71[abd]	7.26	18.82[c]	7.23	20.11[abd]	6.92	18.69[e]	7.59
Social competence	20.64[ab]	8.45	17.25[abed]	8.17	20.34[bc]	7.88	21.19[bd]	7.81	20.64[e]	8.08
Self regulation	20.01[abd]	9.20	15.76[abde]	9.28	18.65[c]	8.17	22.86[abd]	8.05	21.60[be]	8.75
Empathy	10.29[abd]	4.30	8.81[ab]	4.37	10.14[c]	4.07	10.66[ad]	4.19	10.18[e]	4.00
Total score	70.62[ab]	26.00	58.53[abcde]	26.07	67.95[bc]	23.95	74.83[bd]	24.09	71.11[be]	25.72

Note. Mean values by row sharing superscript letters are significantly different at the $p < .05$.

[a]n = 1649 for all scale scores.

Table 20

Effect Size of Post Hoc Pairwise Comparisons of the effect of Rater Setting on Factor and Total Scores of the SEARS-T

Comparison	*Responsibility*	*Social Competence*	*Self Regulation*	*Empathy*	*Total Score*
General ed-sped	.41(small)	.41(small)	.46(small)	.34(small)	.46(small)
General ed-non teaching	.06		.33(small)	.09	
Sped-other teaching		.38(small)			.38(small)
Sped-non teaching	.48(small)	.43(small)	.82(large)		.65(medium)
Sped-other	.65(medium)				.49(small)

Ethnicity. To evaluate the impact of student ethnicity on SEARS-T scores, data were analyzed with a one-way, between-subjects analysis of variance with the results presented in Table 21. The independent variable was student ethnicity with four levels: (a) White/Caucasian (n =819), (b) Hispanic/Latino (n = 312), (c) African American (n = 319), and (d) Asian/Pacific Islander (n = 135). Three categories (Native American (n = 3), multi racial (n =52), and other (n = 20) were not included in the ANOVA. Native American cases were not included because there were only three cases. Multi racial and the other categories were not included because they were composed of heterogeneous samples that are not interpretable in the analysis. The dependent variable was scores on the SEARS-T factor and total scores. The effect of student ethnicity on the total score of the SEARS-T was significant, $F(3, 1584) = 5.76$, $p < 01$. The effect of student ethnicity was significant on the responsibility, self-regulation, and empathy factor scores at the $p <$ .05 level. However, the effect of student ethnicity on the social competence factor score was not statistically significant at the $p < .05$ level.

Table 21

One-Way, Between Subjects Analysis of Variance Summary Table for the Effects of Student Ethnicity on the Factor and Total Scores of the SEARS-T

Source	*df*	*SS*	*MS*	*F*
Responsibility				
Ethnicity	3	572.78	190.93	3.55*
Error	1581	85020.60	53.78	
Total	1584	85593.73		
Social competence				
Ethnicity	3	453.82	151.27	2.17
Error	1581	110286.59	69.76	
Total	1584	110740.41		
Self regulation				
Ethnicity	3	2548.96	849.65	10.37**
Error	1581	129588.64	69.76	
Total	1584	132137.60		
Empathy				
Ethnicity	3	268.15	89.38	4.85*
Error	1581	29159.60	18.44	
Total	1584	29427.74		

Table 21 (continued)

Total score				
Ethnicity	3	11456.43	3818.81	5.76*
Error	1581	1047000.00	662.49	
Total	1584	1659000.00		

* $p < .05$

** $p < .01$

As with the other post hoc comparisons, post hoc comparisons of the effect ethnicity on factor and total scores were carried out using the Tukey test to evaluate all possible differences between pairs of means. The Tukey test was used to reduce familywise type I error because there were no complex comparisons to analyze, so only pairwise comparisons were evaluated (Keppel & Zedeck, 1989). Results are presented in Table 22. Mean comparisons on the responsibility factor indicate significant differences between Caucasian and African American groups. Mean comparisons on the social competence factor were not significantly different for any of the groups. Mean comparisons on the self-regulation factor indicate significant differences between Caucasian and Latino, and Asian and both Latino and African American groups. Mean comparisons on the empathy factor indicate significant differences between Caucasian and Latino groups. Mean comparisons on the total score indicate significant differences between Caucasian and Latino and Latino and Asian groups. Results were significant at the $p < .05$ level. To further evaluate the meaning of the pairwise differences, effect sizes

of each of the significant pairwise contrasts are presented in Table 23. All effect sizes were small, but meaningful.

Table 22

Descriptive Statistics for SEARS-T Subscale and Total Scores by Student Ethnicity

	Caucasian (n = 819)		Latino (*n* = 312)		African American (*n* = 319)		Asian/Pacific Islander (*n* = 135)	
Variable[a]	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Responsibility	19.72[ac]	7.27	18.80[b]	7.29	18.34[ac]	7.66	19.94[d]	7.02
Social competence	20.66[a]	8.30	19.28[b]	8.64	20.00[c]	8.34	20.36[d]	8.05
Self regulation	20.38[ab]	9.09	17.53[abd]	9.00	18.76[cd]	9.21	21.69[bed]	8.57
Empathy	10.43[ab]	4.26	9.50[ab]	4.40	9.77[c]	4.29	10.61[d]	4.30
Total score	71.19[ab]	25.71	65.11[abd]	25.73	67.07[c]	26.24	72.60[bd]	24.73

Note. Mean values by row sharing superscript letters are significantly different at the $p < .05$.

[a]$n = 1584$ for the total scale

Table 23

Effect Size of Post Hoc Pairwise Comparisons of the effect of Ethnicity on Factor and Total Scores of the SEARS-T

Comparison	*Responsibility*	*Social Competence*	*Self Regulation*	*Empathy*	*Total Score*
Caucasian-African American	.18				
Caucasian-Latino			.32(small)	.21(small)	.24(small)
Latino-African American			.14		
Latino-Asian			.47(small)		.30(small)
African American-Asian			.33(small)		

To parse out extraneous variance from the analysis of difference on total scores based on ethnic groups, three two-way ANOVAs were carried out. Further analyses were carried out on the total score and not the factor scores because results on the total score are more reliable, and represent the broader construct better than factor scores. In addition, results of these follow up analyses on the total score would be similar for factor scores, given that the total score is a composite of the factor scores. Given that statistically significant differences in mean scores on the SEARS-T were found for three variables—gender, special education status, and perceived academic performance—these variables could be confounding the differences found between mean scores by ethnic groups. By carrying out a two-way ANOVA with one of the three possible confounding variables, the effect of ethnicity on the total score is analyzed without the effect of that second variable (e.g., gender, special education status, and perceived academic performance). Three two-way ANOVAs were carried out rather than a four-way ANOVA including all four variables to control for alpha slipage, and because the researcher was not interested in the interaction among the three possible confounding variables. Results of the two-way ANOVAs are presented separately by the second independent variable.

Ethnicity by student gender. Data were analyzed as a two-way, between-subjects analysis of variance. The independent variables were ethnic group with four levels (White/Caucasian, Hispanic/Latino, African American, and Asian/Pacific Islander) and student gender with two levels (male and female). The dependent variable was the SEARS-T total score. Sample size for each cell was adequate, $n > 25$. Results are

reported in Table 24. Results indicate that mean differences based on ethnicity without the effect of student gender were significantly different. This result indicates that gender does not confound the effect of ethnicity on SEARS-T total scores. Results also indicate that gender and ethnicity do not interact, that is, that the effect of one of these variables is not dependent on the other.

Table 24

Ethnicity by Student Gender Analysis of Variance Summary Table

Source	*df*	Sums of Squares	Mean Square	*F*
Ethnicity	3	10986.89	3662.30	5.67*
Gender	3	23709.36	7903.12	12.24*
Ethnicity by gender	3	361.20	120.40	.19
Error	1555	1004000.00	645.67	
Total	1565	8579000.00		

*$p < .01$.

Ethnicity by special education status. Data were analyzed as a two-way, between-subjects analysis of variance. The independent variables were ethnic group with four levels (White/Caucasian, Hispanic/Latino, African American, and Asian/Pacific Islander) and student special education status with two levels (special education and general education). The dependent variable was the SEARS-T total score. Sample size for all but one cell was adequate—$n > 25$. One group, Asian and special education, only had a sample size of 8, which is much less than the desired 25 or greater. Thus, results of this analysis could be limited based on small sample size. Results are reported in Table 25.

Results indicate that mean differences on the SEARS-T total score, based on the effect of ethnicity without the effect of student special education status, were not significantly different. This result indicates that special education status likely confounds the effect of ethnicity on SEARS-T total scores, because without the effect of special education status, total scores on the SEARS-T did not differ based on ethnicity. Results also indicate that special education status and ethnicity do not interact, that is, that the effect of one of these variables is not dependent on the other.

Table 25

Ethnicity by Special Education Status Analysis of Variance Summary Table

Source	*df*	Sums of Squares	Mean Square	*F*
Ethnicity	3	169.89	56.63	.09
Sped	3	29838.31	9946.10	16.20*
Ethnicity by sped	6	803.60	133.93	.22
Error	1558	956584.00	613.98	
Total	1571	8610000.00		

*$p < .001$.

Ethnicity by academic performance. Data were analyzed as a two-way, between-subjects analysis of variance. The independent variables were ethnic group with four levels (White/Caucasian, Hispanic/Latino, African American, and Asian/Pacific Islander) and teacher perceived level of academic performance with three levels (below average, average, and above average). The dependent variable was the SEARS-T total score. Sample size for all but one cell was adequate—-$n > 25$. One group, Asian and below

average, only had a sample size of 17, which is less than the desired 25 or greater. Thus, results of this analysis could be limited based on small sample size. Results are reported in Table 26. Results indicate that mean differences on the SEARS-T total score, based on ethnicity without the effect of academic performance, were not significantly different. This result indicates that academic performance likely confounds the effect of ethnicity on SEARS-T total scores. Results also indicate that academic performance and ethnicity do not interact, that is, the effect of one of these variables is not dependent on the other.

Table 26

Ethnicity by Academic Performance Analysis of Variance Summary Table

Source	*df*	Sums of Squares	Mean Square	*F*
Ethnicity	3	22417.92	805.97	1.77
Academic performance	2	182640.93	91320.47	199.98*
Ethnicity by academic	6	1484.22	247.37	.54
Error	1514	691351.05	456.64	
Total	1526	8354000.00		

*$p < .001$.

CHAPTER V

DISCUSSION

The purpose of this study was to evaluate the factor structure and psychometric properties of a new strength-based assessment tool, the Social and Emotional Assets and Resiliency Scales—Teacher version. Five research questions were investigated in this study:

- *Using exploratory and confirmatory factor analyses techniques with a national sample, what is the likely underlying factor structure of the teacher component of the SEARS?*
- *Using Cronbach's alpha, what is the internal consistency reliability of the teacher version of the SEARS?*
- *Using Pearson product-moment correlations, what is the short-term (2 week) temporal stability of teacher-rated SEARS scores?*
- *What is the degree of similarity across different informants (teachers, student) who rate the social and emotional strengths and assets of a given student, using the SEARS?*
- *For what types of decisions is the teacher version of the SEARS valid based on group differences according to student gender, rater gender, disability status, ethnicity, rater-setting and teacher categorization of academic performance?*

Results of the analysis produced a solid four-factor structure for the SEARS-T that was replicated. Reliability analyses indicated strong internal consistency and test re-test reliability for all subscales and the total scale scores; cross-informant correlations between the teacher and child version of the SEARS were relatively modest, as expected.

Results indicate there are no meaningful differences on SEARS-T scores based on rater's gender, student grade, and student ethnicity; there were significant differences based on rating setting, student gender, academic performance, and disability status.

The remainder of this chapter includes a detailed analysis of the results question by question, followed by a discussion of the limitations of the study. Then, the results of the study are discussed in light of new directions for future research and implications for application of these results regarding strength-based measurement and the SEARS assessment system in particular.

Research Questions

Research question 1. Results from this analysis show that variables on the SEARS-T cluster into four significant factors, comprising a total of 39 items. Looking at the items that correlated with the individual factors, suggestions as to what the factors measure are given. Factor One appears to be measuring traits and behaviors related to responsibility; Factor Two appears to be measuring traits and behaviors related to social competence; Factor Three appears to be measuring traits and behaviors related to self regulation; Factor Four appears to measuring traits and behaviors relating to empathy. Based on the variable clusters of these factors, the SEARS-T appears to measure behaviors and traits it purports to measure, those related to social and emotional strengths and resilience.

Results of the SEARS-T factor analyses are similar to and add to the research base on the Adolescent and Parent versions of the SEARS. Results of factor analyses for the SEARS-A have shown four-factors with items clustering in a similar manner as in

this study, though the factors explained differing amounts of variance than those on the SEARS-T. Results of the factor analysis of the SEARS-P showed a three-factor structure consisting of self-regulation/responsibility, social competence, and empathy (Felver-Gant & Merrell, 2010). Though the child version of the SEARS was analyzed in a similar fashion, a strong factor structure could not be determined.

Although four "clean" factors were identified through these analyses, it is important to note that the first factor explains the vast majority of the latent construct—Social and emotional strengths and assets. This large factor—responsibility (consisting of 10 items)—is composed of items assessing multiple skill sets including listening, impulse regulation, decision-making, and independent work skills. It was determined that these skills encompass traits and abilities relating to responsibility. These skills could also be interpreted as separate skill sets describing responsibility as another latent construct. This explanation is both reasonable and explorable. Exploring such an explanation would require other statistical analyses beyond the scope of this study—Hierarchical Linear Modeling. Given the culture of schools in the United States and the emphasis placed on acting responsibly, that responsibility explains the largest portion of variance is not surprising and appears consistent with the school culture. If this measure were to be used in a culture that emphasized other behaviors (e.g., empathy, self regulation) the amount of variance explained by each factor might vary to reflect the values of that culture. Each of the remaining three factors explain relatively little of the latent construct; Factors Three and Four explain less than five percent of what the

SEARS-T measures. When interpreting scores it will then be important to not overly emphasize scores on the factors and use them more as guidelines.

Given that the items on the measure were written for the purpose of assessing strengths related to the domains of social and emotional functioning, these results were consistent with theory. Established theory and research on other strength-based measures (see Epstein, 1999) suggested that the SEARS would measure areas of social competence, emotion regulation and empathy. However, with such little research on strength-based measures in general, it was difficult to anticipate how items would cluster together in terms of how much of the construct the factors would measure. It is possible that the responsibility factor accounted for most of what the SEARS-T measures because resilience is still a fairly nebulous construct—hard to define and measure. Literature on resilience also suggests that it is a complex construct encompassing multiple aspects of functioning (Schroeder & Gordon, 2002). Many of these aspects relate to one another enough to be considered a unified factor, though they may be separate skills. One factor can be composed of skills representing multiple domains of functioning.

Some of the constructs considered to be related to resilience can share similar qualities, creating an overlap of some items on more than one construct. For instance, some items loading highly on the Self-Regulation Factor could arguably be related to responsibility. From this study, several items on the Self-Regulation factor appear to be theoretically related to responsibility (e.g., "Knows how to set goals," "Disagrees without arguing," "Is good at solving problems"). Future research employing Hierarchical Linear Modeling could explain the relation of such items and factors in more detail. Such an

approach could also represent the construct of social emotional assets and resilience as composed of multiple latent constructs with different factors related to each latent construct. This could explain different skills sets related to different constructs all relating to what the SEARS measures, as discussed previously regarding the content of the Responsibility Factor. Other more clear and discrete domains (e.g., social competence and empathy) also accounted for areas of the measure, but were smaller subsets. Such domains, though related to resilience, are more concrete and thus readily identifiable and measurable.

Research questions 2 and 3. The results clearly show the SEARS-T is stable across items. Strong internal-consistency coefficients show that items on the SEARS-T are rated in ways that make sense. That is, teachers are rating items intended to measure similar constructs in similar ways while rating items measuring different constructs differently. Proving that the SEARS-T is internally consistent is important because it provides evidence that scores obtained have meaning beyond the quantity of the score. Providing support for internal consistency is also the foundation of further psychometric analysis and utility investigation. With strong internal consistency, the other results presented in this section have more strength because the basic structure of the measure—the items—is reliable at a very fundamental level.

Short-term temporal stability of the factor and total scores is robust. All correlation coefficients were well above the level considered to be strong. The pattern of results is also consistent with statistical theory and principles. That is, the factor with the least amount of items—Empathy—was less stable, although still strong, and the score

with the highest number of items—Total—was more stable than the other scores. Fewer items on a scale allow for more variability among scores.

Consistency in ratings of strengths over time without the effect of an intervention is not surprising. Although behavioral research indicates that behaviors are subject to change over time, changes generally result from the influence of an external factor (e.g., an intervention). One would not expect children's strengths to change much, if at all, over the course of a two-week period without external influences. This finding also supports the theory that the constructs being measured are more stable than would be expected of specific behaviors. In other words, social and emotional strengths may not be transient states, but are more akin to traits possessed by an individual, which manifest themselves behaviorally. Children are not likely to be responsible one week and irresponsible the next. If this were the case, constructs measured by the SEARS-T would better be interpreted as isolated occurrences of behavior rather than patterns of behavior represented in an immeasurable construct. For now, though, it can be stated that the SEARS-T is reliable over short time periods.

Research question 4. By assessing the correlations between teacher and self-report forms, the researcher was able to evaluate the consistency of scores across raters. As described in the Chapters 1 and 2, research regarding cross-informant correlations using strength-based measures is very sparse. This study was one of if not the first to evaluate cross-informant correlations of strength-based measures using a self report. Research on problem-focused rating scales indicates weak correlations between raters, but correlations of different informants on strength-based measures may be higher

(Achenbach, McConaughy, & Howell, 1987; Beaver, 2008; Friedman, Leone, & Friedman, 1999). Thus, prior to the analysis, the researcher did not know exactly what to expect. Results from this study indicate relationships among ratings between teachers and children are relatively weak. The correlation was higher than those found in deficit-based measures and more similar to other studies using strength-based measures. Several factors may be contributing to these findings. Perceptions of behavior may vary by context. Students and teachers are rating themselves based on different, albeit similar contexts.

In general teachers rate students based on what they see at school, while students are rating themselves based on all contexts in which they exist (i.e., home, school, community, with friends, with adults, and by themselves). For that reason one would expect to find differences in ratings to one degree or another. How youth perceive and thus rate themselves may also change with time due to cognitive development (Ray, Shelton, Hollon, Michel, Frankel, Gross et al., 2009). That is, older youth are likely to perceive themselves in a different, more complex manner, than younger children. These perceptions can be described by theories of cognitive development (e.g., Piaget's four stage theory of cognitive development; Shaffer, 2004). Researchers memory tasks and brain imaging regarding self-versus other-related information indicate differences between older and younger youth in how they recall information and how their brain actually functions. As youth develop they focus more on the self. That is, over time, "the representation of self and of others becomes more complex" (Ray et al., p. 1240). Second, though related, teachers and students have different perspectives. A child may

interpret actions differently than a teacher, and thus rate an item differently. For example, a student may think that other kids like him or her while a teacher does not think very many kids like him or her. In this case, it is the perspective of the rater, in addition to the context, that contributes to differences in ratings. Another related explanation involves interpretation of the items. Raters may interpret or understand the items they read in different ways. Listening, problem solving, and not arguing, can mean different things to different people. These differences may be even more apparent between teachers and students.

It was beyond the scope of this study to assess the exact reasons for the differences between raters. Differences were expected and desired, as one of the purposes of using multiple informants is to obtain unique information (Merrell, 2008). Given the large differences in teacher and student ratings, neither score should be considered a comprehensive assessment of strengths. These forms are intended for use in concert and combined with other information to create a picture of a student's social and emotional strengths. Scores should be interpreted in terms of the context for which they apply. That is, scores are inseparable from the context for which the rater has information. For instance, a student may display certain strengths at home or on sports teams not displayed in school. These strengths would be captured in student or parent reports but not in teacher reports. Such a disparity could support an intervention to generalize skills from one context to another. Differences in scores between teacher and self reports also speaks to the importance of using as many raters as possible to develop a useful intervention for students based on an ecological assessment of functioning. In sum,

the weak-moderate correlation found between teachers and students indicates that teachers and students provide unique information regarding a student's social and emotional assets. This unique information provides support for the usefulness of both student and teacher ratings and also suggests that one rating does not provide a complete and accurate measure of strengths and resilience.

Research question 5. Several analyses of group differences were carried out to answer this question. In discussing these results, it is important to note that results from this study represent differences between groups of individuals and not individuals themselves. Differences within a group are generally greater than differences between groups. Within each of the groups discussed is a range of strengths and resilience. Thus, the reader should not interpret these results and this discussion in terms of making statements about how all individuals in a particular group perform; rather, the results are based on mathematical averages used to make general comparisons between groups. Due to the high number of separate analyses, discussion of the results will be presented by independent variable.

Grade. Results of this study clearly suggest that there are no differences in scores on the SEARS-T between primary and secondary aged students, on all but one factor score—Self regulation. These results could support the idea that strengths and assets are relatively stable traits existing throughout life rather than transient states. This finding does not suggest that strengths and resilience cannot change over time, but it does suggest that age is not a factor to consider when evaluating teacher ratings of resilience and social-emotional strengths. It also appears that teachers recognize and rate strengths

consistently for children and adolescents. Thus, the construct of social emotional strengths and resilience is measured for both children and adolescents by the SEARS-T.

Although there was a difference between scores of students in primary and secondary grades on the self-regulation factor, the effect size of that difference was not meaningful. Items in this factor relate to understanding emotions, demonstrating problem-solving skills, and using coping skills to manage negative emotions. These skills require higher-order cognitive abilities, which may be less developed in younger children than older youth. Labouvie-Vief, DeVoe, and Bulka (1989) suggested that with an increase in cognitive complexity comes an increase in ability to identify and regulate emotions. Other researchers also indicate the ability to regulate one's self changes over the course of development (Campos, Frankel & Camras, 2004). Differences in scores on self-regulation between primary and secondary grades is consistent with other research an theory regarding cognitive and affective development—students in later grades have developed an increased ability to regulate themselves. This difference could also be a product of differing manifestations of emotion regulation. Research on depression and anxiety in children is clear that younger children can manifest depression and anxiety differently than older children and adolescents (Hammen & Rudolph, 1997). Often depressive symptoms appear as overt behaviors such as irritability and acting out. Such overt behaviors would be more noticeable to teachers and their existence may lead them to rate younger children lower on items related to emotion regulation.

Rater gender. As expected, all factor and the total scores were unaffected by rater gender. It would not be expected for raters of different genders to rate students

differently on a measure of strengths. Because the SEARS-T purports to measure strengths and resilience as a construct manifest through behaviors, not teacher's perceptions, rater gender should not affect scores. These results clearly indicate that male and female teachers rate students' strengths and resilience similarly.

Student gender. Results indicate that teachers clearly rate females as having more strengths and resilience than males on all four factors and the total score. The effect sizes of the differences were small, but meaningful, which indicates the differences have importance or meaning. This result is both consistent and inconsistent with other theories and research findings. Research clearly suggests depression is much more prevalent among adolescent females than males (Hammen & Rudolph, 1997). Based on this research, one might think that females would be rated as having fewer social and emotional strengths. On the other hand, prevalence rates of overt delinquent and disruptive behavior is lower for girls, which may lead teachers to rate girls higher on a measure of social-emotional functioning than boys who are engaging in more overt delinquent and non-compliant behavior (Schroeder & Gordon, 2002). Given the covert nature of problem behavior demonstrated by girls, teachers may be less likely to recognize emerging problem behaviors (Dishion & Kavanagh, 2005). Rater bias is always to be considered when using rating scales and could be a factor in the difference in ratings of male and female students. To account for these differences, it will be important to either create separate norms for males and females or note in the published assessment manual these differences so they are considered when comparing males and females in future research and applied use.

Special education status. Results clearly show that teachers rate children and youth receiving special education significantly lower than students receiving general education. Students receiving special education had at least one of several disabilities including emotional disturbance, specific learning disorder, communication disorder, orthopedic impairment and others. It was expected that teachers would rate students with disabilities as possessing fewer strengths and resilient assets. Many of the students with disabilities are unable to perform some of the skills assessed in the measure based on the limitations of their disabilities. This result is consistent with some research citing a history of special education as a possible risk factor for future behavior problems (Thornberry, 2005), which may be partially explained by possessing fewer social-emotional strengths.

Academic performance. Among all of the results, the most robust are the differences between individuals with differing perceived levels of academic performance. On all factors and the total score, teachers rated individuals with higher perceived levels of academic performance as having more strengths than those with lower levels. What is most impressive about these results is the effect sizes of the differences in mean scores. All of the differences were at least medium in terms of size of the differences and many of them were very large. Differences in means between students rated with low and high levels of academic performance were very large. These results are consistent with research regarding risk and protective factors and problem behaviors. That is, school success and academic performance are predictors of later problem behaviors, including juvenile delinquency (Loeber & Farrington, 2000).

Given that academic success requires a fair amount of skills related to responsibility, social interactions, and emotion regulation, it is not surprising that teachers rate students as having more skills and behaviors related to resilience as also being more successful academically. For example, students who have a hard time thinking about their problems in ways that help, listening, and interacting with peers are going to have a harder time navigating a classroom environment, successfully learning new academic skills, and completing assignments, especially more difficult assignments. These results also provide a means of recognizing patterns in a student's rating. That is, teacher ratings of academic performance can be paired with the score on the SEARS-T to identify if the student is matching norms. Students with low academic performance and high scores on the SEARS-T may be manifesting more protective factors, which may help predict future outcomes. Similarly, students with high academic performance and low SEARS-T scores may be an indication the student is struggling in other areas. This indicator may not be detected solely based on their academic performance.

Setting of rater. Closely linked to the findings regarding the effect of special education status on SEARS-T scores was setting of rater. The main finding indicates that teachers in special education settings rate students significantly lower than teachers in all other settings. Given that there was a significant difference in mean scores of students with special education status and students without, it is not surprising that special education teachers rate students in their special education classrooms lower than teachers rating students in other settings. This finding may provides more support for the finding about differences in students with special education status and students without, because

it is another way of assessing the same question. Differences in student scores could also be a result of the training of raters. Although these data do not allow for an analysis of rater training, it is possible that raters in different settings have different training regarding how to evaluate or rate behaviors. It could be the case that teachers in special education settings have different training than teachers in general education settings. For example, some of the raters may have had explicit training in behavioral observation and assessment, which would give them a different lens to look through filling out a behavior rating scale. This training factor, though not able to be evaluated in a controlled manner in this study, may be a factor affecting differences between ratings of raters in different settings.

Though these findings may appear redundant, it provides important qualitative information regarding what information to pay attention to when interpreting scores on the SEARS-T. Whether or not different norms should be established for different settings was not the purpose of this study. However, scores should be interpreted with the setting of the rater of in mind, in order to understand how that student's score compares to other students in that same setting.

Ethnicity. Upon initial analysis of differences between ethnic groups on total scores, it appeared that that there were differences on total scores between those identifying as Asian/Pacific Islander and Latino, and Caucasian and Latino. Mean total scores for individuals identifying as Latino were lower than mean total scores for individuals identifying as either Asian/Pacific Islander or Caucasian Differences between other groups were found on different factor scale scores as well. These

differences were concerning because a measure of social-emotional strengths should not necessarily be expected to differ based on ethnicity alone. If these differences were true and not the artifact of extraneous variables, they could be the artifact of how social-emotional strengths manifest in different cultures. Benson et al. (2006) proposed the idea that assets are universal across cultures but differ in manifestation. It is thus hard to say whether or not these differences are true differences because they may be artifacts of extraneous variables (e.g., academic performance, gender, or special education status).

Knowing that social-emotional competencies may be influenced by academic performance, gender, and special education leads one to question the reality of differences based on ethnicity. When these three variables were controlled, the results indicated that initial differences found between ethnic groups were an artifact of confounding variables. This finding indicates that ethnicity does not affect scores on the SEARS-T, but that the distribution of the other independent variables, which do affect scores, is not proportional across ethnic groups, and may covary with ethnicity. Thus, the unequal distribution of confounding variables across ethnic groups creates an artificial difference between ethnic groups that appears with some statistical tests. Different norms do not appear to be necessary or desirable for different ethnic groups. There is no evidence to indicate that the SEARS-T would not be a useful and valid measure for individuals of all ethnic groups.

When discussing test bias, it is also important to note that assessment tools or procedures are considered bias if "their use resulted in systematic and improper diagnosis, classification, or service provision for a specific group of children or

adolescents (i.e., based on race/ethnicity)" (Merrell, 2008, p. 441). Data from the analyses looking at group differences by ethnicity reflect the assumption that ethnicity does not affect teacher ratings of social emotional strengths. Therefore, though not all ethnic groups were represented in the norming sample according to national census data, the measure is likely valid across ethnic groups. Even if the all ethnic groups were represented in the sample as they are according to national census data, some of the groups (e.g., Native American) would still only include a very small percentage of the sample and thus ratings on that very small sample would not generalize to all members of that ethnic group simply based on their representation in the norming sample (see Merrell, 2008). What is important in behavior rating scale development is establishing evidence that effort to include ethnic groups is made and systematic bias does not occur based on ethnicity. In fact, some evidence indicates that proportion of group representation in a sample has less effect on group differences than is often thought (Merrell).

Limitations

When evaluating the findings of these studies it is important to consider possible confounding influences on the results. Limitations to be considered include rater bias, under representation of some ethnic groups, limited correlational sample size, limited scope of correlational data, and method of data imputation for missing values. First, using a rating scale always introduces rater bias into the results. Rating scales can be reliable and valid, but because they involve human ratings, there is always the possibility of bias. In this case, the bias comes from teachers and also includes students in the cross-

informant correlation analysis. Reducing the impact of teacher bias was hoped to be accomplished by obtaining a large sample size. This was not the case for the correlational study, thus those results may be more subject to rater bias than the analyses including all cases.

Second, though considerable effort was given to obtaining a representative national sample, not all ethnic groups were represented in the sample according to the national census data. Specifically, the number of Latino, Native American and African American children and adolescents in the sample were not representative of these ethnic groups in the nation. However, as discussed previously, data suggest that the scores on the SEARS-T do not systematically differ based on ethnic group, so the measure appears to be valid for all ethnic groups.

Regarding the results of the cross-correlational analysis, the study is limited in both size and scope. A sample size of 31 is adequate but still a small sample. The concern with nesting is that the results may be confounded by uncontrolled variables in the classroom or school. Possible effects of nesting further decrease the generalizability of the results. Also, the sample was relatively homogenous and nested within one school and four classrooms. All of the participants were from one school, between grades three and five, and the majority were Caucasian. It was not the intent of this study to obtain results that would be generalizable to all school settings and populations. Another limitation of the correlational results regards time-limited sampling. Correlations are only snapshots in time. Correlations can, and in some cases do, change over time. Results from this study provide an idea of what the correlation of perceived strength and

resilience between students and teachers are at one time, but do not imply that all teacher-student correlations of strengths are always what these results indicate.

Future Research and Implications for Practice

Future research to continue the process of validating the SEARS assessment system should include longitudinal test-retest reliability analysis, cross-informant analysis with parents, teachers, and students, convergent validity with other strength-based measures, and use as an intervention outcome measure. To better understand the consistency of scores across times, it is necessary that the SEARS-T be used in a study spanning several months using the same individuals to assess the pattern of scores over time. These results would provide critical information regarding the use of the SEARS in intervention outcome research by having a norm for expected scores across time without interventions. To expand the results of the cross-informant correlations in this study, research obtaining parent, teacher, and student ratings for a single student will be instructive. Further research using a longitudinal design to measure correlations at several points in time would provide stronger evidence of the pattern of correlations over time. Such research would not only be entirely unique in the strength-based literature, but also informative regarding the degree of unique information provided across raters and time.

To provide evidence of the content validity of the SEARS-T—an area not addressed in this study--research should include a convergent correlation with another reliable and valid strength-based measure. Given the shortage of psychometrically sound strength-based measures, this research is essential not only for the SEARS but for strength-based assessment in general. After such fundamental reliability and validity

research has been completed, the next avenue of research to pursue should include use as an outcome measure for social-emotional interventions. Though the idea of using strength-based measures as an outcome measure has theoretical validity, the actual usefulness is yet to be determined. Such research will help to answer questions regarding the usefulness of this and other strength-based measures (see Messick, 1988). Such additional research studies are currently underway by members of the Oregon Resiliency Project team.

To more fully address the topic of differing scores related to ethnicity, research could be conducted using Differential Item Functioning and Item Response Theory procedures. Using these procedures it could be determined if members of different ethnic groups systematically differ in the way they endorse certain items. Given that differences between cultures exist in interpreting and perceiving certain phenomena, it would be insightful to know if and which items do members of different ethnic groups rate differently. Such research may also provide information about differing perceptions of strengths and resilience in different ethnic cultures by exploring how certain items function based on the ethnicity of given students.

In chapter 2 the researcher presented the case for using strength-based measures in intervention planning and outcome research. Based on the results of this study that case is bolstered to some degree. With a reliable and valid measure of children's and youth's social emotional strengths and resilience, outcomes of interventions designed to increase social skills, emotion identification and regulation, problem solving, impulse control, and other executive functions, could be assessed using a measure of strengths

and resilience. Given that the SEARS-T shows stability over a two-week period without an intervention, if changes were to occur during a two-week period, the conclusion that those changes are the result of an external factor has more strength. By purposefully changing a child's environment (i.e., setting up an intervention) we would expect changes in SEARS-T scores, which could be used as an indication of the effect of those changes. Though more research concerning the temporal stability and sensitivity to intervention effects is needed, the results on the temporal stability of the SEARS-T from this study provide a strong and positive first step in the direction for use as an outcome measure.

Measuring outcomes with strength-based instruments may provide insight regarding the process and mechanism driving intervention outcomes or resilience, a focus of research in developmental psychopathology (Cicchetti & Cohen, 1995). One of the strengths this study of the SEARS-T demonstrates is that the results are consistent over a short time and four solid factors exist. Factor scores allow for a more specific analysis of results rather than a general statement of strengths, which can also inform outcomes of an intervention. In essence, a strength-based outcome measure may shed light on what skills or attributes youth are acquiring and using rather than focusing on what symptoms or problems have abated. In addition, results from this study suggest that strength-based measures can be used to inform intervention development.

To add to the practice of identifying deficits as a means of building interventions, strength-based measures may help inform what interventions will be beneficial and useful. Though it is true that using a strength-based measure of resilience and social emotional assets can be seen as another way to determine deficient areas of child and

youth functioning, they may also provide information regarding those aspects to be included in an intervention. For example, knowing that a student has a high social competence score, a social interaction component of an intervention could be emphasized when trying to increase other skills. In this way interventions are developed not simply to support growth in deficient areas but to build upon and use strengths a student already demonstrates in order to support growth.

Along the same lines of supporting intervention development, results from this study indicate that the SEARS-T could be used as a strength-based measure in a battery of measures used to screen students who are at risk for future problem. Deficit-based or problem-focused measures often result in categorization into a type of problem without regard to specific skills a youth may be lacking. In these situations the SEARS-T could provide more reason for why a student demonstrates certain types of behavior. More importantly, the SEARS and strength-based measures could be used as prevention tools to identify students at risk of demonstrating difficult or problematic behavior. In all of these applications of strength-based measures, it is not being suggested that strength-based measures completely replace problem-focused measures. A realistic outcome of this and future research would be that assessments include both strength-based and problem-focused measures to provide a more ecological representation of students social, emotional, and behavioral functioning.

APPENDIX

SEARS-T QUESTIONNAIRE

SEARS-T

SOCIAL-EMOTIONAL ASSETS AND RESILIENCE SCALE
Teacher Rating Form, Item Tryout Version 1.0

PART 1: STUDENT AND RATER INFORMATION

Student's Grade Level _____ Student's Sex: Female Male

Student's Ethnic Group(s): White African American Hispanic or Latino Asian
Native American Other _____

Does this student receive special education services? Yes No If yes, please indicate the type of disability (example: leaming disability, autism, etc.)

_____

Class or setting in which you work with this student

_____

Your estimate of how well this student is doing in school: above average average below average

Your years of experience in education _____ Your sex: Female Male

PART 2: DIRECTIONS

Here are 54 items that describe some positive social and emotional characteristics of students. Please rate *how true* you think these items have been for this student *during the past 3 to 6 months.* Circle **N** for NEVER true, or if you have not observed that characteristic. Circle **S** for SOMETIMES true. Circle **O** for OFTEN true, and circle **A** if you think the item has been ALWAYS or ALMOST ALWAYS true for this student during the past few months. Please complete all items.

Remember: **N**EVER **S**OMETIMES **O**FTEN **A**LWAYS

1. Likes to do his/her best in school	N S O A
2. Feels sorry for others when bad things happen to them	N S O A
3. Knows when other students are upset, even when they say nothing	N S O A
4. Stays calm when there is a problem or argument	N S O A
5. Is good at understanding the point of view of other people	N S O A
6. Works independently on assignments, without help	N S O A
7. Tries to help other students when they need help	N S O A
8. Other people like to be with her/him	N S O A
9. Is comfortable talking to many different people	N S O A
10. Makes friends easily	N S O A
11. Expresses disagreement with other people without fighting or arguing	N S O A
12. Tries to understand how other students feel when they are not doing well	N S O A
13. Is a good listener	N S O A
14. Other students ask him/her to hang out with them	N S O A
15. People think she/he is fun to be with	N S O A
16. Is well-liked by teachers and other students	N S O A
17. Other students come to her/him for help	N S O A
18. Likes doing things for other people	N S O A
19. Is good at solving problems	N S O A
20. Is good at starting conversations	N S O A
21. Understands how other people feel	N S O A
22. Makes good decisions	N S O A
23. Feels okay with the way she/he is	N S O A
24. Is good at settling disagreements of other students	N S O A
25. Is comfortable telling other people how he/she feels	N S O A
26. Asks others for help when she/he needs it	N S O A
27. Gives compliments to others	N S O A
28. Understands how people could feel different about the same thing	N S O A
29. Stays in control when he/she gets angry	N S O A
30. Cares what happens to other people	N S O A
31. Thinks before she/he acts	N S O A
32. Is comfortable being in large groups	N S O A
33. Other people see him/her as a leader	N S O A
34. Likes who he/she is	N S O A
35. Is respected or "looked up to" by other students	N S O A
36. Is dependable, someone you can rely on	N S O A
37. Thinks of her/his problems in ways that help	N S O A
38. Accepts responsibility when she/he needs to	N S O A
39. Seems to like being at school	N S O A
40. Is good at telling stories and jokes	N S O A
41. Feels good about himself/herself	N S O A

42. Is able to handle problems on her/his own..... N S O A
43. Stands up for herself/himself..... N S O A
44. Knows how to calm down when stressed or upset..... N S O A
45. Knows how to identify and change negative thoughts ..... N S O A
46. Asks teacher for help when he/she does not understand an assignment..... N S O A
47. I trust her/him..... N S O A
48. Works well with other students on group projects..... N S O A
49. Can figure out whether or not negative thoughts are realistic..... N S O A
50. Appears to feel accepted and comfortable at school..... N S O A
51. Can identify errors in the way he/she thinks about things..... N S O A
52. Knows how to set goals for what she/he wants in life..... N S O A
53. Is able to handle problems that really bother other students..... N S O A
54. When life is hard, doesn't let things get to him/her..... N S O A

Remember: NEVER **S**OMETIMES **O**FTEN **A**LWAYS`

BIBLIOGRAPHY

Achenbach T. (1991a.). *Integrative Guide for the 1991 CBCL/4-18, YSR, & TRF Profiles.* Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach T. (1991d.). *Manual for the Youth Self-Report & Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T.M., McConaughy, S.H., & Howell, C.T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101(2)*, 213-232.

American Psychological Association (2002). American Psychological Association, Ethical principles of psychologists and code of conduct, American Psychologist 57 (2002), pp. 1060–1073.

American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1999). *Standards for educational and psychological tests.* Washington, DC: American Educational Research Association.

Anastasi, A. (1988). *Psychological testing (6th ed.).* New York, NY, England: Macmillan Publishing Co, Inc.

Arbuckle, J. L. (2005). *Amos 6.0 User's Guide.* Spring House, PA: Amos Development Corporation.

Beaver, B.R. (2008). A positive approach to children's internalizing problems. *Professional Psychology: Research and Practice, 39(2)*, 129-136.

Benson, P.L., Scales, P.C., Hamilton, S.F. & Sesma, A. (2006). Positive youth development: Theory, research, and applications. In R.M. Learner, W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development sixth edition* (6th ed., pp. 895-933). Hoboken, NJ: John Wiley & Sons Inc.

Bernat, D.H. & Resnick, M.D. (2006). Healthy youth development: Science and strategies. *Journal of Public Health Management Practice.* P 10-16.

Bonett, D.G. & Wright, T.A. (2000). Sample size requirements for estimating person, kendal and spearman correlations. *Psychometrika, 65(1),* 23-28.

Campos, J.J., Frankel, C.B. & Camras, L. (2004). On the nature of emotion regulation. *Child Development, 75(2),* 377-394.

Caplan, G. & Grunebaum, H. (1967). Perspectives on primary prevention. A review. *Archives of General Psychiatry, 17(3),* 331-346.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1,* 245-276.

Chen, G. & Astebro, T. (2003). How to deal with missing categorical data: Test of a simple Bayesian method. *Oranizational Research Methods, 6(3),* 309-327.

Cicchetti, D. (2006). Development and Psychopathology (Chapter 1). In D. Cicchetti & D. J. Cohen (Eds). *Developmental psychopathology, 3,* Risk, disorder, and adaptation (pp1-23). New York: Wiley.

Cicchetti, D., & Cohen, D.J., (1995). Perspectives on Developmental Psychopathology. In D. Cicchetti, & D.J. Cohen (Eds.), *Developmental Psychopathology Vol 1: Theory and Method.* Oxford, England:Wiley Press Inc.

Cicchetti, D., & Rogosch, F.A. (2002). A developmental psychopathology perspective on adolescence. *Journal of Consulting and Clinical Psychology, 70(1),* 6-20.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112(1),* 155-159.

Cohn, B., Merrell, K. W., Felver-Grant, J., Tom, K., & Endrulat, N. (2009, February). Strength-based assessment of social and emotional functioning: SEARS-C and SEARS-A. Presented at the Meeting of the National Association of School Psychologists, Boston.

Cox, K.F. (2006). Investigating the impact of strength-based assessment on youth with emotional or behavioral disorders. *Journal of Child and Family Studies, 15(3),* 287-301.

Deno, S.L., (2002). "Problem Solving as 'Best Practices.'" In A. Thomas & J. Grimes, (Eds.), *Best practices in school psychology IV,* (pp. 37-56). Bethesda, MD: The National Association of School Psychologists.

Dishion, T.J. & Kavanagh, K. (2005). *Intervening in Adolescent Problem Behavior: A Family-Centered Approach.* New York, NY: The Guilford Press.

Elliott, S.N. & Gresham, F.M. (2007). Social Skills Improvement System (SSIS). Minneapolis, MN: Pearson Assessments.

Endrulat, N. R., Tom, K., Ravitch, K., Wesley, K., & Merrell, K. W. (2010, March). Gender differences in positive social emotional functioning. Presented at the meeting of the National Association of School Psychologists, Chicago.

Epstein, M.H. (1999). The development and validation of a scale to assess the emotional and behavioral strengths of children and adolescents. *Remedial and Special Education, 20(5),* 258-263.

Epstein, M.H., Dakan, E., Oswald, D.P. & Yoe, J.T. (2001). Using strengths-based data to evaluate children's mental health programs. In M. Hernandez, S. Hodges, (Eds). Developing outcome strategies in children's mental health. Systems of care for children's mental health. (pp. 153-166). Baltimore, MD, US: Paul H Brookes Publishing.

Epstein, M.H., Harniss, M.K., Pearson, N. & Ryser, G. (1999). The behavioral and emotional rating scale: Test-retest and inter-rater reliability. *Journal of Child and Family Studies, 8(3)*, 319-327.

Epstein, M.H., Harniss, M.K., Robbins, V., Wheeler, L., Cyrulik, S., Kriz, M., Nelson, R.J. & Weist, M.D. (2003). Strength-based approaches to assessment in schools. In M.D. Weist, S.W. Evans, N.A. Lever, (Eds.). Handbook of school mental health: Advancing practice and research. Issues in clinical child psychology. (pp. 285-299). New York, NY, US: Kluwer Academic/Plenum Publishers.

Epstein, M.H. & Sharma, H.M. (1998). *Behavioral and Emotional Rating Scale: A Strength Based Approach to Assessment.* Austin, TX: PRO-ED.

Felver-Gant, J., & Merrell, K. W. (2009, February). Teacher's ratings of student's assets and resilience: Development of the SEARS-T. Presented at the Meeting of the National Association of School Psychologists, Boston.

Felver-Gant, J. & Merrell, K. (2010, March). Parent Ratings of Student's Assets and Resilience: Sears-P Factor Structure. Presented at the Meeting of the National Association of School Psychologists, Chicago.

Floyd, F.J. & Widaman, K.F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7(3),* 286-299.

Friedman, K.A., Leone, P.E., & Friedman, P. (1999). Strengths-based assessment of children with SED: Consistency of reporting by teachers and parents. *Journal of Child and Family Studies, 8(2)*, 169-180.

Garmezy, N. (1993). Children in poverty: Resilience despite risk. *Psychiatry, 56,* 127-136.

Good, R.H. & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 61-88). New York: Guilford Press.

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of American Academics Child Adolescent Psychiatry, 40(11),* 1337-1345.

Greenberg, M.T., Weissberg, R.P., O'Brien, M.U., Zins, J.E., Fredericks, L., Resnik, H., et al., (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist, 58(6/7),* 466-474.

Greenspoon, P. J., & Saklofske, D. H. (2001). Toward an integration of subjective well-being and psychopathology. *Social Indicators Research. 54*, 81-108.

Gregory, R.J. (2004). *Psychological testing: History, principles, and applications,* Fourth Ed. Boston, MA: Pearson Education Group, Inc.

Gresham, F.M., Watson, S., & Skinner, C. (2001). Functional behavioral assessment: principles, procedures, and future directions. *School Psychology Review, 30,* 156-173.

Hammen, C. & Rudolph, K.E. (1997), Childhood Depression. In E.J. Mash & R.A. Barkley (Eds). *Child Psychopathology, 1st Ed.* (pp 153-195).

Harlacher, J. E. (2008). Social and emotional learning as a universal level of support: Evaluating the follow-up effect of Strong Kids on social and emotional outcomes. Unpublished doctoral dissertation, University of Oregon, Eugene.

Hu, L., & Bentler, P. M. (1999). Cut off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-31.

Huebner, E. S. (1991). Initial development of the Students' Life Satisfaction Scale. *School Psychology International, 12*, 231-240.

Jimerson, S.R., Sharkey, J.D., Nyborg, V., & Furlong, M.J., (2004). Strength-based assessment and school psychology: A summery and synthesis. *The California School Psychologist, 9*, 9-19.

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

Kazdin, A.E., (1978). *Evaluation of behavior therapy: Issues, evidence, and research strategies.* Cambridge, MA: Ballinger Pub. Co.

Keppel,G.,&Zedeck,S.(1989). *Data analysis for research designs: Analysis of variance and multiple regression/correlation approaches.* New York: Freeman.

Kline, R. B. (2005). *Principles and practice of structural equation modeling, 2nd edition.* New York: Guilford Press.

Leffert, N., Benson, P.L., Scales, P.C., Sharma, A.R., Drake, D.R., & Blyth, D.A. (1998). Developmental assets: Measurement and prediction of risk behaviors among adolescents. *Applied Developmental Science, 2(4)*, 209-230.

Labouvie-Vief, G., DeVoe, M., & Bulka, D. (1989). Speaking about feelings: Conceptions of emotion across the life span. *Psychology and Aging, 4*, 425-437.

Luthar, S.S. (1991). Vulnerability and Resilience: A study of high-risk adolescents. *Child Development*, 1991, 62, 600-616

Luthar, S.S. & Cicchetti, D., (2000). The construct of resilience: Implications for interventions and social policies. *Development and Psychopathology, 12,* 857-885.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111,* 490-504.

Merrell, K. W. (2008). *Behavioral social, and emotional assessment of children and adolescents* (3rd ed.). Mahwah, NJ: Erlbaum.

Merrell, K. W. (2007). *Social emotional assets and resiliency scales.* Retrieved February, 2009 from http://strongkids.uoregon.edu/SEARS.html.

Merrell, K. W. (2002). *School Social Behavior Scales, Second Edition.* Baltimore: Paul H. Brookes Publishing.

Merrell, K.W., Ervin, R.A., & Gimpel, G.A., (2006). School psychology for the 21st century: Foundations and practices. New York, NY: The Guilford Press.

Messick, S.J., (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer and H.I. Braun (Eds.), *Test validity,* (pp. 33-45). Hillside, NJ: Lawrence Erlbaum.

National Association of School Psychologists. (2000a). *Professional conduct manual:Principles for professional ethics and guidelines for the provision of school psychological services.* Bethesda, MD: Author.

Ogden, T. (2003). The validity of teacher ratings of adolescents' social skills. *Scandinavian Journal of Education Research, 47(1),* 63-76.

Preacher, K. J., & MacCallum, R. C. (2002). Exploratory factor analysis in behavior genetics research: Factor recovery with small sample sizes. *Behavior Genetics, 32(2),* 153-161.

Ray, R.D., Shelton, A.L., Hollon, N.G., Michel, B.D., Frankel, C.B., Gross, J.J. et al. (2009). Cognitive and neural development of individuated self-representation in children. *Child Development, 80(4),* 1232-1242.

Reynolds, C. R., & Kamphaus, R. W. (2004). Behavior Assessment System for Children-Second Edition manual. Circle Pines, MN: American Guidance Service Publishing.

Rhee, S., Furlong, M.J., Turner, J.A., & Harari, I. (2001). Integrating strength-based perspectives in psychoeducational evaluations. *The California School Psychologist, 6*, 5-17.

Salvia, J. & Ysseldyke, J., (2004). *Assessment: In special and inclusive education (9th ed.).* Boston, MA: Houghton Mifflin.

Schroeder, C.S. & Gordon, B.N. (2002). *Assessment & Treatment of Childhood Problems: A Clinicians Guide (2nd ed.).* New York, NY: The Guilford Press.

Sha, F. & Morgan, S.B. (1996). Teacher's ratings of social competence of children with high versus low levels of depressive symptoms. *Journal of School Psychology, 34(4),* 337-349.

Shaffer, D.R. (2004). *Social and Personality Development (5th ed.).* Belmont, CA: Wadsworth Publishing.

Snyder, C.r., Lopez, S.J., Edwards, L.M., Pedrotti, J.T., Prosser, E.C., Walton, S.L., et al. (2003). Measuring and labeling the positive and the negative. In S.J. Lopez & C.R. Snyder (Eds.). *Positive psychological assessment: A handbook of models and measures.* (pp. 21-39). Washington, DC: American Psychological Association.

SPSS Graduate Student Version for Mac, Rel. 16.0.1. 2007. Chicago: SPSS Inc.

Stevens, J. (2002). *Applied multivariate statistics for the social sciences (4th ed.).* Mahway, NJ: Lawrence Erlbaum Associates.

Stormshak, E.A. & Dishion, T.J. (2002). An ecological approach to child and family clinical and counseling psychology. *Clinical Child and Family Psychology Review, 5(3),* 197-215.

Suldo, S.M. & Shaffer, E.J. (2008). Looking beyond psychopathology: The dual-factor model of mental health in youth. *School Psychology Review, 37(1),* 52-68.

Synhorst, L.L., Buckley, J.A., Reid, R., Epstein, M.H. & Ryser, G. (2005). Cross informant agreement of the behavioral and emotional rating scale-2nd edition (BERS-2) parent and youth rating scales. *Child and Family Behavior Therapy, 27(3),* 1-11.

Tabachnick, B. G., and Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed.). Boston: Allyn and Bacon.

Tilly, D. V. III, (2002). Best practices in school psychology as a problem-solving enterprise. In A. Thomas & J. Grimes, (Eds.). *Best practices in school psychology IV,* (pp. 21-36). Bethesda, MD: The National Association of School Psychologists.

Tom, K., Merrell, K. W., Endrulat, N. R., Cohn, B., & Felver-Gant, J. C. (2009, February). Assessing positive youth characteristics: Development and structure of the SEARS-P. Presented at the Meeting of the National Association of School Psychologists, Boston.

United States Census Bureau (1996). Annual Estimates of the Population by Selected Age Groups and Sex for the United States: April 1, 2000 to July 1, 2006. Retrieved January 5, 2009, from http://www.census.gov/popest/national/asrh/NC-EST2006-sa.html.

United States Census Bureau (1995). Population distribution in 2005. Retrieved January 5, 2009, from www.census.gov/population/www/pop-profile/files/dynamic/PopDistribution.pdf

Walker, H.M. & Stieber, S. (1991). Teacher ratings of adolescent social skills: Psychometric characteristics and factorial replicability across age-grade ranges. *School Psychology Review, 20(2),* 301-315.

Widaman, K.F. (2006). Missing data: What to do with or without them. *Monographs of the Society for Research in Child Development, 71(3),* p. 42-64.