Social-Emotional and Behavioral Assessment in Culturally and Linguistically Diverse Learner

Populations: An Examination of the Validity and Reliability of the Social Skills Improvement

System in School-Aged Spanish-Speaking English Language Learners

by

Paige Lauren Mission

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Educational Psychology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2016

Date of final oral examination: 01/27/16

The dissertation is approved by the following members of the Final Oral Committee:
    Craig A. Albers (Chair), Associate Professor, Department of Educational Psychology
    Maribeth Gettinger, Professor, Department of Educational Psychology
    Thomas R. Kratochwill, Professor, Department of Educational Psychology
    Dan Bolt, Professor, Department of Educational Psychology
    Stephen Quintana, Professor, Educational Psychology, Counseling Psychology

Table of Contents

List of Tables

List of Figures

Abstract

Social-emotional development, psychosocial functioning, and relational experiences have been shown to impact academic achievement and psychological well-being. Social skills have been identified as being particularly critical in promoting mental health and life-long success. Children with well-developed social skills (e.g., sharing, being cooperative, taking responsibility, trying to understand how others are feeling) typically experience positive peer relationships, favorable psychological experiences, and higher levels of academic success than children without these skills. Due to the persistence of social skills deficits as caused by a lack of skill or ability to perform in a classroom setting, children who struggle to navigate social encounters frequently experience problems in school and social maladjustment. Children from culturally and linguistically diverse backgrounds have been identified as potentially experiencing difficulties in obtaining such skills, which may further contribute to poor academic and social outcomes. Early intervention is therefore necessary, but appropriate and valid assessment methods are first essential. Currently there is a lack of research on the screening and assessment of social-emotional and behavioral functioning for students from culturally and linguistically diverse backgrounds, especially for students identified as English language learners (ELLs). ELLs constitute the fastest growing population of culturally and linguistically diverse children and would likely benefit from more support. The *Social Skills Improvement System* (SSIS; Gresham & Elliott, 2008) is one assessment and intervention program that recently has been proposed as particularly promising for identifying social skills deficits in school-aged children. Social skills are measured across the areas of communication, cooperation, assertion, responsibility, empathy, engagement, and self-control. The SSIS also examines externalizing behaviors, bullying, hyperactivity/inattention, internalizing behaviors, autism spectrum behaviors, and levels of

academic competence. The validity and usability of the SSIS, however, have yet to be established in Spanish-speaking ELLs across the continuum of English language proficiency (ELP) levels. This study examined the SSIS's validity and reliability using item-response theory for the purpose of determining whether any items on the SSIS demonstrate differential item functioning in Spanish-speaking ELLs across the continuum of English language development.

CHAPTER I

Introduction and Literature Review

Prosocial behaviors have been widely shown to impact engagement, academic achievement, and mental health outcomes (e.g., Algozzine, Wang, & Violette, 2011; Gresham & Elliott, 2008; Juvonen, Espinoza, & Knifsend, 2012; Reinhard, & Linnenbrink-Garcia, 2012)[1]. According to a U.S. Department of Health and Human Services (1999) review, social-emotional development widely impacts relational experiences, communication skills, school readiness, self-esteem, academic achievement, and psychological functioning. Social competence and interpersonal functioning also promote adaptive functioning and serve as protective factors for psychopathology or deviant behaviors (Sheridan, Knoche, Edwards, Bovaired, & Kupzyk, 2010; Wentzel & Watkins, 2002). Furthermore, the Diagnostic and Statistical Manual (DSM-V-R; American Psychiatric Association, 2013) defines many psychiatric disorders in youth in terms of relational experiences within the context of dysfunctional environmental systems and social-emotional challenges (Evangelista & McLellan, 2004). With 20% or more of all children struggling with a clinically significant mental illness, valid and useful assessments that are linked with evidence-based practices are widely needed to support prevention and intervention efforts that promote psychosocial functioning (Adelman & Taylor, 2010; Burns et al., 1995).

Recently, attempts to promote social skills, social-emotional development, and behavioral health have been based in evidence-based and ecologically-focused programs that maintain integrity and accountability through the use of valid screening tools and monitoring systems

---

[1]   Algozzine et al. (2011) have warned practitioners and researchers that the relationship between social skills and academic success is unclear due to the fact that these findings have been based on correlational research. However, they acknowledge the importance of recognizing that social skills interventions have been associated with improvements in social-emotional functioning and academic achievement.

(Adelman & Taylor, 2010; Rhodes, Ochoa, & Ortiz, 2005). Universal screening measures and targeted assessments that evaluate social skills, emotional challenges, and behavioral problems at the elementary school level therefore are recognized as a best practice that facilitates the delivery of interventions, promotes achievement, and supports positive mental health outcomes (Elliott, Huai, & Roach, 2007).

Valid screening systems and assessment tools that identify social skills deficits and social-emotional functioning in culturally and linguistically diverse youth populations are essential in advancing comprehensive and effective prevention-based programs throughout educational and clinical systems (e.g., Bowman & Moore, 2005; Landau & Milich, 1990; Rhodes et al., 2005). As classrooms have become increasingly heterogeneous, public and private organizations have endorsed research that examines the development and validity of psychological assessments for students from culturally and linguistically diverse backgrounds (American Psychological Association Presidential Task Force on Evidence-Based Practice, 2006; U.S. Department of Health and Human Services, 1999).

English language learners (ELLs), who are students whose native language is something other than English and have limited academic knowledge in English, have been identified as the fastest growing and most heterogeneous group of school-aged children in the United States and are represented by hundreds of languages (Albers & Martinez, 2015; Artiles & Ortiz, 2002; Ortman & Shin, 2011). In the 21st century, schools are faced with new challenges in supporting both the academic and social-emotional development of students from culturally and linguistically diverse backgrounds. Fortunately, in an attempt to promote valid, reliable, and acceptable assessment practices, advancements have been made in the areas of test theory, psychometrics, measurement, and survey research (Brennan, 2006; Ortiz, 2002; Solano-Flores &

Nelson-Barber, 2001). However, a lack of scientifically based knowledge exists around the development and acceptability of measures that accurately identify social-emotional functioning in students from linguistically diverse backgrounds.

Prevention Science in Context

   Prevention science in educational systems. As prevention and intervention efforts have gained prominence in the delivery of human services, the essential components of prevention frameworks have become widely recognized across educational settings. The Commission on Chronic Illness (1957) originally defined primary prevention as the interception of a disorder before its advancement, secondary prevention as the prevention of a worsening of a condition, and tertiary prevention as the alleviation of the impact of the manifestation of the disorder. This multi-tiered approach has gained popularity as a Response-to-Intervention (RtI) paradigm in support of the identification and treatment of both academic and mental health challenges in youth in schools.

   Over the past decade, educational systems have increasingly focused on promoting responsive systems that are effective and efficient, as well as advancing prevention-based programming (Doll & Yoon, 2010). Under the revised *Individuals with Disabilities Education Act of 2004* schools are encouraged to offer instructional support through a multi-tiered system of delivery (U.S. Department of Education, 2004). At each intervention tier students receive increasingly intensive supports that address deficit areas and minimize the need for a special education diagnosis and corresponding individualized education plan (Brown-Chidsey & Andren, 2013). Within Tier 1 (also referred to as the universal stage), every child receives access to instruction and services that are culturally appropriate and evidence-based. Progress is monitored on a consistent basis. If data suggest that an individual is performing outside of the typical range

of functioning and not appropriately responding to universally delivered supports, then additional support is provided at Tier 2 (e.g., selected) level. As more intensive services are provided progress is monitored and analyzed. Lastly, if insufficient progress is observed, Tier 3 services are provided and can be as intensive as one-on-one instruction (Brown-Chidsey & Andren, 2013). Based on the central tenets of prevention science, this model focuses on reducing the number of students who might otherwise be placed in special education or other school-based pullout programs. Originally developed within prevention science research, this model has been widely used for academic, social-emotional, and behavioral purposes (Albers & Martinez, 2015; Hosp & Madyun, 2007).

       Bias in prevention systems. In considering multi-tiered systems of support, researchers and practitioners have expressed concerns regarding systemic issues of bias and prejudice (Rhodes et al., 2005; Solano-Flores, 2011). Adopting a deficit-based framework, which serves as the foundation for a medical model of diagnostics and treatment, precludes the potential value and information that could be obtained through a dualistic approach that examines risk and protective factors (Whitecomb, 2012). Adopting a strengths-based approach, in addition to considering individuals' challenges, allows for a comprehensive assessment of functioning and the creation of efficient and effective treatment planning (Friedman, Leone, & Friedman, 1999; McCurdy, Coutts, Sheridan, & Campbell, 2013). Focusing solely on psychopathology and skill-deficits further limits the extent to which effective interventions can be identified because of issues related to screening measures' sensitivity, reliability, and unrecognized environmental variables (Kaplan, 2000; Solano-Flores, 2011). Personal bias, even within the context of standardized tests, can also negatively impact the identification and implementation of instructional planning for youth (Kaplan, 2000). However, such a model can also be difficult to

implement in schools because students must meet certain classroom expectations, and by understanding a student's problem areas support appropriate supports can be provided. Therefore, when identifying supports for diverse learners, multiple factors must be taken into account (Solano-Flores & Nelson-Barber, 2001). Identifying the change mechanisms that directly impact behaviors and psychological outcomes have been shown to be more effective than simply recognizing the presence or absence of symptomatology (U.S. Office of the Surgeon General, 2001). Utilizing screening systems that maintain high levels of treatment utility and are ecologically focussed (e.g., take into account family structure and socio-political factors), strength-based, and accurate are essential for supporting educational planning, instruction, and the development of effective interventions (Hosp & Madyun, 2007).

Environmental factors must be considered to effectively conceptualize risk and protective indicators (Adelman & Taylor, 2010). In primary educational settings, researchers have encouraged educational professionals to recognize social-emotional functioning and adaptive behaviors, instead of the absence of deficits (Gettinger, Ball, Mulford, & Hoffman, 2010). However, the development and use of effective, efficient, and affordable systems that offer such recognitions are impacted by a lack of familiarity with screening systems, a dearth of resources, divisions between social-emotional learning and academic standards, and a poor understanding of effective assessment methods for culturally and linguistically diverse youth (Doll & Yoon, 2010; Kataoka, Zhang, & Wells, 2002; Rhodes et al., 2005; Solano-Flores, 2011).

Recognition of the barriers that impact the application of prevention efforts and the implementation of multi-gated assessments can facilitate the development of effective delivery systems. Due to a limited amount of resources and the importance of early intervention, the efficiency, cost-effectiveness, validity, and reliability of an assessment system must be

considered. In the context of early identification, evidence-based programming is critical for promoting mental health outcomes and academic achievement (Asos, Lieb, Mayfield, Miller, & Penucci, 2004). Educational systems must have culturally appropriate and valid systems of measurement for the purpose of establishing an effective system of prevention, advancing educational policy, facilitating the delivery of evidence-based programming, and promoting both academic achievement and mental health (Darney, Reinke, Herman, Stormont, & Ialongo, 2012; Glover & Albers, 2007; Kettler, Glover, Albers, & Feeney-Kettler, 2014; McCurdy et al., 2013).

Early Identification of Social-Emotional Challenges in School-Aged Youth

Social-emotional functioning and academic development. The development of social skills and social-emotional functioning begins in early childhood and is affected by biological factors, environmental variables, and interpersonal experiences (Gettinger et al., 2010; Rhodes et al., 2005; Whitcomb & Merrell, 2011). Social skills are acquired across settings throughout childhood and adolescence and are critical in promoting overall development. The essential components of social skills consist of (a) communication (e.g., effective and positive exchange of either verbal or physical reciprocal interaction with another peer), (b) assertion (e.g., the ability to ask for help), (c) responsibility (e.g., respects the property of others), (d) empathy (e.g., concern for others, tries to understand how others are feeling), (e) engagement (e.g., extends invitations to others, makes friends easily), (f) self-control (e.g., makes compromises, stays calm when teased), and (g) cooperation (e.g., follows directions, takes turns; Caprara et al., 2000; Feil & Walker, 1995; Gresham & Elliott, 2008).

In addition to home-based settings, schools have been identified as an environment that can positively impact the development of critical social skills, student performance, and psychological development (McCurdy et al., 2013). Mental health, experiences associated with

high levels of acceptance and support, and social-emotional development have been suggested to moderate academic outcomes (Diperna, 2005). Allowing children to "pay better attention to speakers, work more cooperatively with others, ask for help when needed…behave more responsibly…(and) promote positive interactions while simultaneously discouraging negative interactions when applied to appropriate social situations" social skills influence psychological functioning both intra- and interpersonally (Gresham & Elliott, 2008, p. 1). Caprara, Barbaranelli, Pastorelli, Bandura, and Zimbardo (2000) additionally found that prosocial skills (e.g., cooperating and consoling) in third grade served as a better predictor of eighth grade academic achievement than academic achievement itself (early academic achievement and prosocial behaviors, r = .75). Conversely students who are difficult to manage in the classroom, avoid work, 'act out', are unable to self-regulate, and struggle socially with their peers are continually observed to develop psychological disorders, which has been associated with negative impacts on their achievement later on in life (Campbell, 2002; Malecki & Elliott, 2002).

The process of acquiring social skills across environments has additionally been suggested to moderate social-emotional development, behavioral functioning, and school achievement (Diperna, 2005). Social skills functioning impact a child's capacity to develop effective study skills, motivation to learn, and engagement in academic content (academic competence and social skills, r = .66; Ray & Elliott, 2006). According to Reynolds and Walberg (1991, 1992) academic success is impacted by multiple variables (see Figure 1). Moreover, prosocial behaviors have been suggested to augment academic achievement (DiPerna & Elliott, 2000; DiPerna, Volpe, & Elliott 2005). Conversely, students that lack social skills and motivation experience peer rejection, fewer learning opportunities across various situations (e.g., loss of academic time), and miss out on the occasion to receive encouraging and supportive

feedback from teachers, family members, and peers (Noam & Herman, 2002). DiPerna, Volpe, and Elliott's (2002) model captures these dynamic relationships observed across these areas, which have been identified as academic enablers (e.g., prior achievement, interpersonal skills, prior achievement, motivation, engagement, study skills) and academic outcomes (see Figure 1).

Figure 1. Conceptual model of relationships among academic enablers and academic achievement
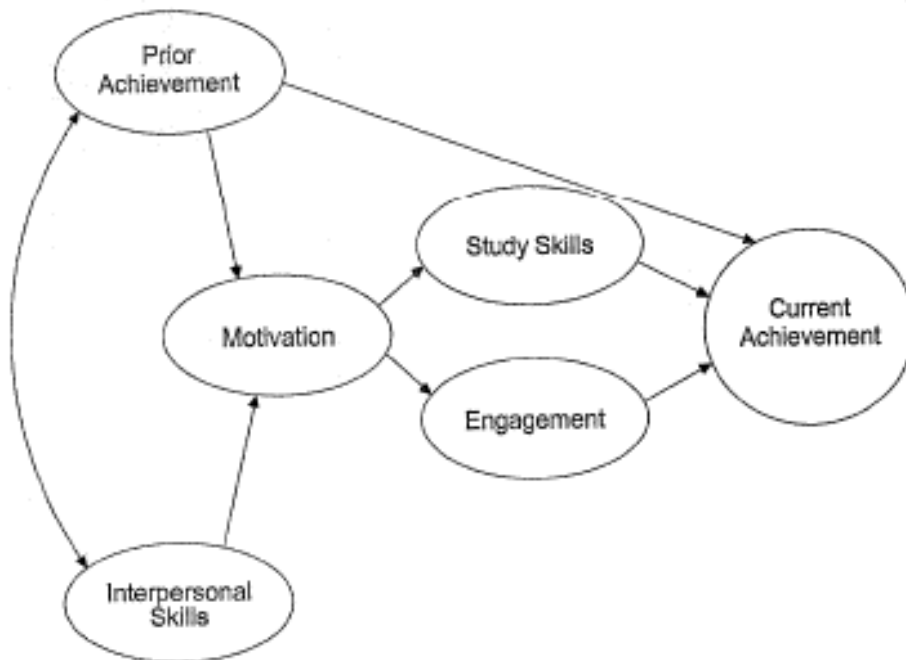
Note. From DiPerna, J. C., Volpe, R. J., & Elliott, S. N. (2002).

Elementary school students who do not develop critical social skills and adaptive behaviors are at a particularly high risk for academic failure and poor mental health outcomes (Zins & Elias, 2006). This is acutely true for children who have complex psychosocial problems, demonstrate maladaptive behaviors, and experience environmental challenges (e.g., poverty; Manz, Fantuzzo & McDermott, 1999). Longitudinal research suggests that young school-aged children with co-occurring academic and behavior problems are particularly more likely to be

placed in special education, require mental health services, demonstrate poor academic outcomes in secondary school, and drop out of school (Darney et al., 2012). Language development and early literacy skills have also been identified as influencing social-emotional development, behavioral functioning, and academic performance across student populations (Gettinger et al., 2010; Shonkoff & Phillips, 2000). The relationship between language development, which is a mechanism that operates on the development of social-emotional behaviors, and a child's understanding and display of the various elements that comprise social skills functioning is not well understood (Reynolds & Lowe, 2009). Consequently, the identification and measurement of social-emotional development in younger students (e.g., challenges, moderators, mediators) is more difficult due to the lack of understanding of critical elements associated with the development of prosocial functioning and adaptive skills in students from diverse backgrounds and the fact that universal screening systems and teacher referrals can be biased and insensitive to contextual variables such as language and cultural experiences (Abidin & Robinson, 2002; Rhodes et al., 2005). Albeit insufficient for a comprehensive assessment due to the amount of time teachers spend with students, teacher judgment has continued to be highly valued and utilized in the identification of students with academic, social-emotional, and behavioral concerns (Ollendick, Oswald, & Francis, 1989).

Defining features associated with the acquisition of social skills, and the relationship between social-emotional development and academic achievement, remain largely contextual and culturally-bound (Solano-Flores, 2011). Socio-linguistic factors can therefore distort responsive universal screening practices and evidence-based interventions. Gettinger et al. (2010) emphasized that individual characteristics, family history, and social-environmental features interact with one another to predict future academic and pro-social behaviors. For

example, within the context of a child's experiences and culture, self-regulatory capabilities have been associated with social-emotional growth, cognitive control, and academic proficiency (Pianta, Laparo, Payne, Cox, & Bradley, 2002). Moreover, since adaptive functioning and social skills have been shown to be highly contextualized and culturally bound, researchers have repeatedly called for a further investigation of the extent to which culture and language moderate student outcomes, influence social-emotional development processes, and impact the acquisition and display of social skills functioning (Baron, & Kenny, 1986; Bronson, 2000; Dever & Kamphaus, 2013; Diperna, 2005; Oakland, Wechsler, & Maree, 2013).

Social-emotional and behavioral measurement in schools. Encouraged by the U.S. Department of Health and Human Services (1999), educational programs are required to promote positive academic and mental health outcomes for all children. The No Child Left Behind Act (2001; PL 107-110) emphasizes the importance of early identification for students who do not display appropriate levels of social skills and academic readiness (e.g., social-emotional skills), with a special focus on supporting students identified as ELLs. Public policy has therefore influenced the extent to which school-aged children access and receive services (U.S. Department of Health and Human Services, 1999; Winter & Kelley, 2008).

Over the past decade, calls have been made for the advancement of early identification and warning systems for children that are at risk for academic failure and developing mental health disorders (Dever & Kamphaus, 2013; Reynolds & Lowe, 2009; U.S. Department of Health and Human Services, 1999). In response, researchers have attempted to validate both screening and assessment tools by examining measurement systems across racially diverse groups, promoting measures that demonstrate acceptable psychometric properties, and ensuring that teachers and parents have opportunities to offer their expertise (Whitcomb & Merrell, 2012).

However, there is a paucity of research on the assessment of social-emotional development and social skills functioning across linguistically diverse populations.

Downward extensions of measures that were meant for older students at the secondary level have contributed to multi-gated assessment systems that promote questionable identification practices at the elementary school level (Noam & Hermann, 2002; Whitcomb & Merrell, 2012). Due to the fact that students come to school with a diverse range of abilities (e.g., bilingualism, reading, information processing skills, capacity to moderate behavior), practitioners continually struggle to consistently and efficiently identify social-emotional functioning in young students (Rhodes et al., 2005). Screening processes also require an investment of time and resources before significant problems have developed or been identified; therefore, most assessment and intervention practices have remained largely reactive (Whitcomb & Merrell, 2012). However, reactive service delivery practices provide only the appearance of offering immediate benefits, because such processes are ultimately more resource intensive and taxing on both schools and families (Glover & Albers, 2007). Guided by research in the area of prevention science, there now exists a greater awareness of the inefficiency and costliness of reactive service delivery models (e.g., Noam & Hermann, 2002; White, Albers, DiPerna, Elliott, Kratochwill, & Roach, 2004).

Due to a lack of universal screening systems that have been validated for students from linguistically diverse backgrounds and belief systems that ELLs lack the necessary skills to access academic material and appropriately engage with peers in school environments, educational systems have struggled to identify and effectively provide services to diverse student populations that are in need of extra social-emotional and academic support (Albers & Martinez, 2015; Albers, Mission, & Bice-Urbach, 2013; Campbell, 2002; Rhodes et al., 2005). Many

intervention programs are also typically offered at higher grade levels even though parents often describe their concerns as having begun as early as preschool (Lopez et al., 2000). As a result, there is a need for systems that correctly identify social-emotional challenges in every young student.

Measurement practices for diverse learners. Student referrals have been widely based on teacher observations and reports (Whitcomb & Merrell, 2012). However, this process typically lacks consistency and continuity across student populations (Severson, Walker, Hope-Doolittle, Kratochwill, & Gresham, 2007). Faced with the task of managing, teaching, and supporting students from a wide variety of backgrounds, teachers struggle to effectively and efficiently identify youth who have academic and social-emotional challenges early on in their development (Elliott, Barnard, & Gresham, 1989; Elliott et al., 2007). Teacher training programs rarely offer guidance on working with culturally and linguistically diverse students, including training regarding effective identification practices for determining risk and protective factors that are related to poor mental health outcomes in elementary school-aged students (Dever & Kamphaus, 2013; Lee, 1998; Thorp, 1997). Therefore, even though teachers offer necessary information in supporting the identification of social-emotional and learning challenges, their reports are not sufficient (Whitcomb & Merrell, 2012). A lack of empirically supported evidence regarding the validity and effectiveness of universal screening systems for elementary school students with linguistically diverse characteristics has further put into question identification procedures (Abedi, 2013).

The Individuals with Disabilities Education Act (IDEA) promotes the identification and evaluation of students with developmental concerns, social-emotional problems, and behavioral challenges (U.S. Department of Education, 1999, 2004). According to the *child find* requirement

within IDEA, comprehensive and continuous efforts must be made to identify students who would benefit from special education and early intervention services (Hebbeler, Spiker, & Kahn, 2012). Encouraging the use of standardized, validated, and culturally appropriate screening systems, the *child find* requirement emphasizes the use of empirically supported identification procedures that consider teacher and parent input (Feil, Severson, & Walker, 1998). Researchers have expressed concerns regarding biased referral processes, disproportional rates of service delivery, and inequitable identification procedures for students with social-emotional and behavioral problems (Abedi, 2013; Gottleib, Gottlieb, & Trongue, 1991; Heller, Holtzman & Messick, 1982). Systems-level issues (e.g., limited funding, a focus on developmental processes that are based on the majority culture in training programs) have also hindered the development of programming for children from culturally and linguistically diverse groups who present with social skills deficits and social-emotional challenges.

Referral processes are impacted by several factors, including (a) teacher bias (among which includes differing perceptions of appropriate behaviors and levels of tolerance for behaviors, as well as preference toward certain peers), (b) changes in standards over time, and (c) expectations based on gender, age, language proficiency, culture, or ethnicity. These factors, in part or in aggregate, may further account for a lack of accuracy in the identification of linguistically diverse students (Goodman & Webb, 2006; Hosterman, Dupaul, & Jitendra, 2008; Sideridis, Antoniou, & Padeliadu, 2008).

Although Karabenick and Clemens-Noda (2004) reported that teachers generally maintained positive attitudes toward ELLs, teachers have also been identified as struggling with negative belief patterns regarding bilingualism. For example, teachers have occasionally expressed concerns that when a child is learning a second language their ability to learn in

academic settings is hindered (Barrera, Corso, & Macpherson, 2003; Walker, Shafer, Liams, 2004). Due to the lack of recognition of the impact of socio-linguistic factors and general cultural differences on student development and challenges related to the creation and implementation of effective universal screening systems for culturally and linguistically diverse children, research on the technical features of psychological testing, test development processes, and the general framework from which tests are based would assist in establishing a foundation for systems that support populations with diverse learners.

English language learners. Language development differs across academic and social contexts, as well as across student populations. Research suggests that social, instructional, and academic language development across ELL and native English-speaking populations is complex (Albers & Martinez, 2015; Anstrom, DiCerbo, Butler, Katz, Millet, & Rivera, 2010; Bailey & Butler, 2002; Francis, Lesaux, Kieffer, & Rivera, 2006). A variety of similarities exist between the two groups. For example, Takanishi (2004) demonstrated that social skills at the third grade level similarly predicted achievement and mental health functioning across both the general population and Spanish-speaking ELLs. Suggesting that early assessment of social-emotional functioning, and the consideration of a student's English language proficiency (ELP) level as occurring along a continuum may be critical for both developing and implementing assessment and intervention practices that are sensitive to the needs of every learner (Barrera et al., 2003; Gottlieb & Hamayan, 2007; Ortiz, 2002).

Research has identified various factors that place Spanish-speaking ELLs at risk for experiencing psychosocial and academic challenges. Chang et al. (2007) suggested that Spanish-speaking ELLs are more likely to experience increased victimization in kindergarten, struggle with being assertive, and withdraw from peer interactions. Conchas (2001) also reported that

Spanish-speaking ELLs experience heightened levels of social isolation due to perceptions of limited opportunities to interact with peers. Toppelberg, Medrano, Morgens, and Nieto-Castanon's (2002) examination of bilingual language skills and psychopathology suggested that children with clinically significant emotional/behavioral problems, demonstrated bilingual language skills that were strongly and inversely correlated with problem scores, particularly as related to global problems ($r = -0.67$, $p < .001$). In considering older students, Galindo and Fuller (2010) recently reported that by fourth grade, ELLs reported lower levels of school belonging as compared to their same aged English proficient peers. In an attempt to understand such challenges faced by Spanish-speaking ELLs, Barrera et al. (2002) and Spomer and Cowan (2001) have suggested that these students in particular struggle in social situations because of their lack of assertiveness. Subsequently assessment systems, which are constructed to serve students with a specific type of profile, have been documented as potentially over-identifying social skills deficits and social-emotional challenges, and under-identifying adaptive skills in children from culturally and linguistically diverse backgrounds (Dowdy, Dever, DiStefano, & Chin, 2011; Losen & Orfield, 2002).

Researchers therefore believe that social-emotional and behavioral functioning are considered to be rooted in cultural experiences (Greenfield, 1997a; Rhodes et al., 2005; Solano-Flores, 2011). It then follows that Spanish-speaking ELLs enter school with many adaptive behaviors and social skills. According to Crosnoe (2007), children with limited ELP enter kindergarten with strong social skills (e.g., cooperation) that facilitate their learning and academic achievement in school environments. Carlson and Meltzoff (2008) further found that Spanish-speaking ELLs with a limited understanding of the English language were able to manage conflicting demands by peers and adults, and sustain attention in social situations.

Importantly, this ability to maintain inhibitory control has strong implications for prosocial functioning and learning (Bialystok, 2001). Stipek and Byler (2001) also found that Latino students were less likely to be held back in kindergarten. Hair, Halle, Terry-Humen, Lavelle, and Calkins (2006) suggested that ELLs often display social skills (e.g., *respeto* – respect; *bien educado* – having good upbringing and polite behavior) that are not always valued by traditions observed in the United States, which include value systems that are rooted in competition, independence, and Socratic questioning.

This range of findings could demonstrate the complexity inherent in acculturative processes and the heterogeneity across Spanish-speaking ELL populations. For example, García Coll, and Marks (2009) suggested that exposure to "White" culture leads to challenging periods of transition that can foster a negative peer environment. Previously considered tangential, the lack of consideration regarding developmental pathways related to dual language learning, environmental processes, and the recognition of cultural differences in the construction of assessment systems has hindered the advancement of valid and effective screening programs for ELL populations (Dever & Kamphaus, 2013; Gottlieb & Hamayan, 2007; Rhodes et al., 2005). Furthermore, Dowdy et al. (2011) specifically found that a frequently used broad-band social-emotional assessment system might over-identify problem behaviors and under-identify adaptive functioning in Spanish-speaking ELLs. More specifically Dowdy et al. (2011) found that teachers more frequently endorsed better organizational skills for English-proficient students than Spanish-speaking ELLs, and indicated that Spanish-speaking ELLs scored significantly higher than non-ELLs as having social-emotional and behavioral problems on a universal screening measure. Dowdy et al. (2011) additionally found that items likely functioned differently for Spanish-speaking ELLs then for their English proficient peers on a broadband

assessment system. More specifically, teachers classified Spanish-speaking ELLs as having more school problems (i.e., attention problems and learning problems) and showing fewer adaptive skills (i.e., adaptability, social skills, leadership, study skills, and functional communication) at a rate that was statistically significant. Due to these discrepancies the authors concluded that it was unclear whether or not the measures were correctly identifying problem behaviors in students from linguistically diverse backgrounds, and that further research was widely needed to examine potential measurement bias in social-emotional and behavioral screening systems. Therefore, screening measures that are culturally sensitive, valid, useable, and able to effectively identify ELLs in need of early intervention are needed more than ever to advance culturally competent practices within a multi-tiered system, and across programs with limited resources.

Technical Considerations for Screening and Assessment Practices

   Measurement in a multi-tiered framework. Universal screening and assessment practices are significantly influenced by standards established by the American Educational Research Association (AERA), American Psychological Association (APA) and the National Council on Measurement in Education (NCME; AERA, APA, & NCME, 2014) for educational and psychological testing. Further encouraged by IDEA, as well as APA (2002) and the National Association of School Psychologists (NASP, 2000), assessments must be valid for the purposes for which they are being used and maintain technically adequate properties. For example, measures must demonstrate adequate levels of sensitivity and specificity, which means that the test must correctly identify an appropriate number of individuals that do and do not possess a given problem (American Educational Research Association et al., 2014; Lochman, 1995).

   Within a multi-tiered framework, universal screening procedures guide initial data-based decision-making (Kettler et al., 2014; Jimerson, Burns, & VanDerHeyden, 2007). Beginning at

the universal level, data are collected through screening systems and informal observations. Throughout the year, progress is continually monitored. Individuals that are identified as being at risk and are unresponsive to evidence based supports delivered at the universal level are further assessed and observed. At the second tier, additional resources are required and data are collected from formal rating scales and interviews (Severson et al., 2007). Fewer students are assessed at this second tier (e.g., approximately 15 – 20% of the population; Severson et al., 2007). In the third or final tier, the interventions that are provided and assessments (e.g., functional behavior assessments, in-depth interviews, and detailed observations) that are used require an extensive amount of time and resources. Therefore, even fewer individuals are assessed and given access to services at this level (e.g., approximately 5% of the population). At this third tier, if a student continues to demonstrate inadequate progress he or she could potentially receive a formal diagnosis or special education label (National Association of State Directors of Special Education [NASDSE], 2005). Adhering to such a framework can ultimately enhance service delivery for all children (NASDSE, 2005). Moreover, since research has demonstrated that individuals from culturally and linguistically diverse backgrounds frequently neither access nor receive mental health services, promoting equitable service delivery throughout this model at the elementary school level can potentially solve challenging problems (Huey & Polo, 2008).

Psychometric properties. Psychometrics traditionally function as the foundation for guiding the development of psychological testing. At the most fundamental level, test developers work to ensure that measurement systems maintain sufficient levels of reliability and validity (see Table 1; AERA et al., 2014).

Reliability. Proposed as a necessary, but insufficient, condition for validity, reliability is the consistency of measurements when the testing procedure is repeated in a population of individuals or groups (AERA et al., 2014). For the purpose of minimizing measurement error, precision and consistency are considered to be the foundational elements associated with reliability. Critical components of precision and consistency are internal consistency, part-whole reliability, and test-retest reliability.

Validity. As the most fundamental element, validity establishes the extent to which both evidence and theory can be considered to support the way tests are used and scores are interpreted (Sattler, 2008). The AERA et al. (2014) provide a comprehensive review of the validity-based evidence standards that are necessary for both developing and choosing an appropriate test. Recommending that researchers and practitioners consider the appropriateness of local norms, as well as unintended consequences from the test's sensitivity or failure to represent a particular construct, the AERA et al. (2014) highlighted the importance of considering rival hypotheses and culturally relevant practices in developing and administering assessment measures and screening systems. Although considered a unitary concept, various types of validity exist; however, the relationship between the construct and content of a test is regarded as a critical indicator in establishing the validity of a measure (AERA et al., 2014).

The predictive and incremental validity of assessment systems are particularly important to consider when screening at the universal level (Feil & Walker, 1995; Glover & Albers, 2007). According to Levitt, Saka, Romanelli, and Hoagwood (2007), information regarding the predictive validity of most psychological tests tends to be limited and highly variable. In accordance with a lack of information regarding the prevention of poor academic and mental health outcomes in children from linguistically diverse backgrounds, the differential validity and

Table 1

An Overview of Psychometric Properties

	Definition	Suggested Coefficient Values
Reliability		
Internal Consistency	The extent to which individual items measure the same construct, and are consistent within a scale. Internal consistency coefficients, or relationships across scores within a single measure. Internal consistency can be further established through an examination of item total reliability, which can be calculated by correlating a score on a single item and the total score. Cronbach's alpha is one type of coefficient that measures the homogeneity of test items by measuring the extent to which items relate within a measure.	At or above 0.80 (Sattler, 2008)
Part-Whole Reliability	The degree to which outcomes on an entire measure are consistent across individual items.	At or above 0.80 (Sattler, 2008)
Test-Retest Reliability	Also referred to as inter-rater reliability, this property measures the extent to which a measure is consistent over repeated administrations. The consistency of scoring across raters is also considered. Test-retest coefficients are determined after multiple administrations of the same test.	At or above 0.80 (Sattler, 2008)
Alternate-Form Coefficients	Values derived from administrations of similar tests at different time points.	At or above 0.80 (Sattler, 2008)
Validity		
Construct Validity	Relates to reliability, and is the first type of validity that is examined; internal structures and external relationships across variables must be considered. Referring to the extent to which a scale or test measures the theoretical construct that it claims to measure, this property considers a measure's internal structure and a measure's validity based on the extent to which the measure relates to other variables (American Educational Research Association et al., 2014). Includes content, concurrent and predictive validity.	At or above 0.80 (Sattler, 2008)
Content Validity	Measures specific features in a test and how well such properties measure the intended construct.	At or above 0.80 (Sattler, 2008)
Concurrent Validity	The degree to which a test correlates with measures that were previously validated (Glover & Albers, 2007).	At or above 0.80 (Sattler, 2008)
Predictive Validity	Measures the accuracy of a test and indicates whether or not a test correctly identifies individuals (Glover & Albers, 2007). Also referred to as incremental validity, this property depends on the context and availability of measures, and takes into account practical applications (Hunsley & Meyer, 2003). Based on whether an entire instrument or its components add construct validity to measures that are currently available, this property is assessed using a series of regression analyses that can determine whether a significant increase in the variance of an outcome variable could be accounted for by the new instrument.	At or above 0.75 (Glover & Albers, 2007)

usability of measures is critical in considering how systems identify students across diverse populations. Smith, Fischer, and Fister (2003) therefore emphasized that measures should only be developed if they possess superior psychometric properties, and secure additional necessary information efficiently and effectively.

    Sensitivity and specificity. Levitt et al. (2007) expressed concerns that broad- and narrow-band assessment instruments do not consistently possess high levels of validity across predictive validity indices, including those of sensitivity, specificity, negative predictive power (NPP), and positive predictive power (PPP). Levitt et al. (2007) therefore suggest that an in-depth examination of screening systems is necessary to decrease false negatives and positives. Maintaining appropriate levels of specificity is critical for preventing the misidentification of students and promoting effective resource allocation (Glover and Albers, 2007). For example, certifying that screening instruments have high NPP and sensitivity would ensure that students that are in need of additional support are able to access services (see Table 2).

    A review of *validity generalization* by AERA et al. (1999) suggests that the individual must also be considered, because individual and group differences can impact whether

Table 2
Possible Outcome Frequencies of Students Screened

		Eventual Outcome	
		Social Skills Deficits	No Social Skills Deficits
Screening Indicator	At-Risk	A	b
	Not At-Risk	C	d

Note. Sensitivity = a/(a+c), specificity = d/(b+d), PPP = a/(a+b), and NPP = d/(c+d).

assessment tools can be generalized across populations. In selecting a screening system, Glover and Albers (2007) recommended considering a series of key elements, including (a) cultural appropriateness or the 'goodness of fit' of the instruments in the desired population, (b) theoretical relevance and evidence base, (c) technical adequacy, and (d) usability. As a pre-requisite to addressing concerns regarding the validity and usability of screening systems, Glover and Albers (2007) emphasized the importance of examining the contextual and developmental fit of each measure within the intended population. In the case of elementary school populations, the goals of psychological screening systems are twofold: (a) to appropriately identify children who are at-risk for developing serious social-emotional and behavioral disturbances, and (b) to appropriately connecting students with evidence-based interventions. In addition to considering the standardization and relative norms of screening systems and assessment measures, the usability of the tool of interest must also be considered. Usability relates to the extent to which a system is accepted by stakeholders, cost-effective, able to be implemented (e.g., availability of resources for ELLs), and demonstrates high levels of treatment utility (Glover & Albers, 2007). Issues related to the fit between the measure and environment in which it will be used must therefore be fully considered.

Social-Emotional and Behavioral Measurement Systems

With mental health screening emphasizing the importance of classification (e.g., categorical or continuous), researchers have expressed concerns regarding the current status of such screening systems at the primary school level as well as for youth from culturally and linguistically diverse backgrounds (Achenbach & McConaughy, 1997; Rhodes et al., 2005). Historically, psychological tests have rarely taken into account cultural, linguistic, and developmental diversity across student populations. Therefore, the internal and external validity

of measurement systems continues to remain questionable (Rhodes et al., 2005; Ruffalo & Elliott, 1997).

Broadband measurement systems for elementary school populations. A wide range of multi-gated assessment systems currently exists for elementary school-age populations (Levitt et al., 2007; Severson et al., 2007; Whitcomb & Merrell, 2012). However, research suggests that the usability, validity, and reliability of measures for Spanish-speaking ELLs are largely unknown (Reynolds & Lowe, 2009; Severson et al., 2007). A review of current social-emotional and behavioral assessment systems revealed three particularly promising measures that can be used at the elementary school level, including the the Achenbach System of Empirically Based Assessment (ASEBA; Achenbach & Rescorla, 2001), the Systematic Screening for Behavior Disorders (SSBD; Walker & Severson, 1999), and the Social Skills Rating System (SSRS; Gresham & Elliott, 1990) and the Social Skills Improvement System (SSIS; Gresham & Elliott, 2008). Each of these measures has demonstrated a high incidence of use and is (a) a broadband assessment tool that screens for social-emotional functioning, (b) still in publication, and (c) used at the elementary school level.

Achenbach System for Empirically Based Assessment (ASEBA; Achenbach & Rescorla, 2001) and the Child Behavior Checklist (CBCL 6-18; Achenbach, 1991). The ASEBA for school-aged children includes the Teacher Rating Form (TRF), Youth Self-Report (YSR, 11-18), and Child Behavior Check List (CBCL/1½-5/6-18; Achenbach, 1991). Teachers complete the TRF, parents complete the CBCL, and students aged 11-18 complete the self-report. The measures reflect patterns of behavior that are symptomatic of empirically based syndromes. Each measure takes approximately 15 minutes to complete. The CBCL/6-18 contains 118 items rated on a three point Likert-type scale and a limited number of open-ended questions. The

measure is based on DSM-oriented scales, with subscales including Aggressive Behavior, Anxious/Depressed, Attention Problems, Rule-Breaking Behavior, Social Problems, Somatic Complaints, Thought Problems, Withdrawn/Depressed, Affective Problems, Anxiety Problems, Somatic Problems, Attention Deficit/Hyperactivity Problems, Opposition Defiant Problems, Conduct Problems, Obsessive-Compulsive Problems, Post-Traumatic Stress Problems, Sluggish Cognitive Tempo, and Positive Qualities. Each subscale loads onto one of three broad scales (i.e., Internalizing, Externalizing, and Total Problems).

For the purpose of assessing the psychometric properties of the CBCL/6-18, two samples of school-aged children of various sample sizes of children from across various cultural and ethnic backgrounds were used (Achenbach & Rescorla, 2001). One sample consisted of typically developing children, while the other sample was comprised of students with behavioral concerns that were referred by teachers. Demographic data were not reported. An analysis of item consistency with the DSM was conducted and the content validity was developed through research and examination of item endorsement on previous versions of the measures (Achenbach & Rescorla, 2000). The internal validity was established based on factor analyses drawn from eight factors on the CBCL/6-18 and the TRF [Aggressive Behavior, Attention Problems, Delinquent Behaviors, Social Problems, Somatic Complaints, Thought Problems, Externalizing, and Internalizing]. Internal consistency ranged from .78 to .97, test-retest reliability ranged from .95 to 1.00, and inter-rater reliability ranged from .93 to .96 between parents and teachers. The measures have been reported as being more complex to interpret and administer, but multiple norms drawn from parents from culturally and linguistically diverse backgrounds were established to determine the degree of deviance displayed by the child within his or her cultural group. Additional research on the sensitivity and specificity of the CBCL has revealed that a very

small percentage of teachers and parents reported data that overlapped for children that were selected as cases (Garrison & Earls, 1985). Due to the fact that the authors report that a limited number of children from diverse backgrounds were included in the validation of the CBCL, researchers do not assume that these norms are acceptable for culturally and linguistically diverse individuals.

Although the Office of Special Education Programs (OSEP) and the Stanford Research Institute (SRI) panel have identified the CBCL as a selected measure for social-emotional screening (Severson et al., 2007), the underrepresentation of Spanish-speaking ELLs in the development of the CBCL/6-18 indicates that this measure may not be appropriate for use in diverse populations. Concerns regarding the consistency of items throughout the internalizing subscales, and outdated norms further put into question the appropriateness of this measure for use in specific populations. A review of research that examined the use of the CBCL in Spanish-speaking ELL populations yielded 11 studies, which elicited mixed results (see Table 3). Therefore, researchers have questioned the validity and reliability of the CBCL for Latino populations and elementary school-aged ELLs.

Systematic Screening for Behavior Disorders (SSBD; Walker & Severson, 1990, 1992, 2002). The SSBD (Walker & Severson, 1990, 1992, 2002) is one of the most prominent screening systems used in elementary schools across the nation. This multi-gated system assists in the identification of children in first through fifth grade who are at risk for developing social-emotional difficulties. The SSBD begins with (a) the nomination phase; (b) followed by a targeted rating scale phase, which is estimated to take one hour to complete; and (c) a standardized direct behavior observation stage that is completed by a trained observer in a classroom and playground setting. In the first phase teachers nominate 20 students (i.e., ten

students with externalizing behavior problems, and ten students with internalizing behavior problems), create two rank ordered lists based on the severity and frequency of the problems, and choose three students within each dimension. In the second stage teachers complete the critical behavioral disorders scale. The SSBD contains a teacher form that includes an Events Checklist, which contains a 33-item checklist regarding both internalizing and externalizing behaviors, and the Combined Frequency Index for Adaptive and Maladaptive Behavior. The Combined Frequency Index contains a 23-item list (i.e., 12 appropriate and 11 inappropriate behaviors) of 5-point Likert-type questions. A trained observer then observes the three students with the most critical scores.

The SSBD has been widely standardized and norm referenced across elementary school populations. High levels of reliability and validity have been reported. The internal consistency of the targeted rating scale, as based on Cronbach's alpha, has been estimated as being above .80 (Walker & Severson, 1992). Based on Spearman's ρ, the test-retest reliability within the teacher nomination phase has been reported as being .72 for internalizing behaviors and .79 for externalizing behaviors, and the inter-rater agreement on the internalizing and externalizing dimensions within the first stage ranged from .82 to .94 (Walker & Severson, 1992). (For more information regarding the psychometric properties of the SSBD see Caldarella, Young, Richardson, Young, & Young, 2008; Walker, Severson, Nicholson, & Kehle, 1994; Walker, Severson, Todis, & Block-Pedego, 1990.) The SSBD was also identified by the OSEP and SRI panel as one of the best available options for social-emotional screening (Severson et al., 2007). Although the authors did not report sensitivity and specificity levels, this screening system is considered to be cost-effective and connected with evidence-based practices. Recently

researchers and practitioners have reported concerns regarding the amount of time required to complete the SSBD and its heavy reliance on teacher ratings (Whitcomb & Merrell, 2012).

Further examination of the standardization process of the SSBD suggests that a nationally representative sample was not used in the development and validation of this multi-staged screener. Additionally, the manual does not report demographic information beyond identifying "white" and "non-white" participants. Therefore, the usability of this screening system in Latino or Spanish-speaking ELL populations remains unknown. A review of current research on the SSBD that has included Latino populations only includes one dissertation study (Jenkins, 1998). Additional studies that have used the SSBD, but did not directly investigate the validity and usability of the system across Spanish-speaking ELLs, reveal that only a negligible number of Latino students have been examined in general. Jenkins' (1998) dissertation study assessed the use of the SSBD in two culturally diverse schools in west Texas.

Jenkins (1998) determined that the mean scores of students from culturally and linguistically diverse backgrounds differed significantly from the normative data provided by Walker and Severson (1992) within the SSBD manual. Potentially due to the fact that the authors were not sure how to classify culturally and linguistically diverse students (i.e., ELLs), Jenkins (1998) concluded that the SSBD may not appropriately classify such students and may need to be re-normed for these students. In an attempt to address the limitations of the SSBD and create a more valid and reliable screening system, Gresham and Elliott (1990, 2008) developed a multi-tiered model of assessment and refined the link between evidence-based interventions with assessment results.

Social Skills Rating System (SSRS; Gresham & Elliott, 1990). Developed by Gresham and Elliott (1990), the SSRS served the purpose of identifying social-emotional and behavioral

challenges and strengths in children ages 3 to 18 years. Based on a series of norm-referenced

measures, the SSRS includes teacher (SSRS-T), parent (SSRS-P), and child (SSRS-C) behavior

rating scales. Each measure has been widely used by researchers and practitioners to assess pro-

social behaviors, social-emotional challenges, and treatment outcomes across students in

Table 3
Research on the CBCL/ASEBA in Latino Populations

Authors	Purpose	Sample	Limitations	Psychometrics and Findings
Achenbach et al. (1990)	• Examination of the CBCL/4-16, TRF/6-16, and YSR/12-16 in U.S. mainland and Puerto Rican populations	• U.S. mainland sample: 724 (CBCL), 567 (TRF), 274 (YSR) • Puerto Rico sample: 724 (CBCL), 567 (TRF), 274 (YSR)	• Samples obtained 3 years apart • Children in special education or remedial classrooms were excluded • English language learner status was not examined	• 4 items showed large cross-cultural effects (Var(x)>.13): dependency, sulking, concern with cleanliness and secretive for Puerto Rican children • High Total Problem Scores for Puerto Rican children (p = 0.02), but children rated themselves as having significantly fewer problems than U.S. mainland children • Parents rated Puerto Rican children as not being involved in organizations, jobs, and making friends, and scored lower on ratings of time and skill in these areas (Var(x) > .13) • Teachers rated U.S. mainland children higher on adaptive behaviors (i.e., working hard, learning, and happy) than Puerto Rican children (p < .01); overall activities, social and total competence scores were rated more highly for U.S. mainland children • Recommend higher cutoff points due to concerns of over identification of problem areas in Puerto Rican children
Crijnen, Achenbach, & Verhulst, (1997); Crijnen, Achenbach, & Verhulst (1999)	• Examined parent-reported problem behaviors in children across 12 cultures	• 13,697 children and adolescents aged 4 through 17 from a population in Australia	• U.S. mainland sample: 724 (CBCL), 567 (TRF), 274 (YSR) • Sample sizes across cultures differed	• Across age levels children from Puerto Rico (μ=12.8, 13.5), Greece (μ=11.2), and China (μ=4.4) scored significantly above the established mean (p < .01) on problem behaviors and internalizing problems
Lequerica, & Hermosa (1995)	• Reviewed CBCL and ASEBA data to inform epidemiological research on Hispanic children	• 52 Hispanic mothers or preschool children aged 2 through 5 (59.6% Spanish speaking)	• A limited number of items were used from the ASEBA and CBCL 6 • English language learner status was reported but not examined	• Ratings on 13 externalizing behaviors and 5 internalizing dimensions: > .6 (higher frequency ratings in Hispanic samples) • Ratings of developmentally appropriate behaviors: .06 – .7 (comparable to non-Hispanic samples) • Recommend higher cutoff points due to concerns in the over identification of problem behaviors in Hispanic children

Table 3

Research on the CBCL/ASEBA in Latino Populations (cont'd)

Authors	Purpose	Sample	Limitations	Psychometrics and Findings
Aisenberg (2001)	• Examined the effects of exposure to community violence in Latina mothers and their children	• 31 Latina children (58% female) aged 4 through 5	• Small sample size • English language learner status was not examined • Universal exposure to community violence	• Clinically significant differences were observed in Latino children on the CBCL for total problem behaviors (p = .01), externalizing problems (p < .01), and internalizing problems (p = .01)
Leadbeater & Bishop (1994)	• Reviewed CBCL/2-3 data to evaluate preschool children with African American and Puerto Rican adolescent mothers	• 83 children aged 1 to 3 years of age (56% African-American, 38% were Puerto Rican, and 4% other)	• Concerns with an over identification of behavior problems due to depressive symptoms and parenting styles • Administered the CBCL to a subset of the sample (n = 62)	• A total of 13% of children had clinically significant behavior problems, which falls in the high average range compared to estimates of serious behavioral problems in large-scale research studies • Inter-rater reliability: .93 • Total child difficult behaviors modestly correlated with total CBCL scores (r = .23) • Ratings for cooperative, compliant and passive behaviors didn't correlate with CBCL scores (r = -.03, .09, & -.14)
Bird et al. (1987)	• Reviewed CBCL data to inform epidemiological research on Puerto Rican children	• 191 children aged 4 to 16 • 62 children were referred participants and 105 participants met DSM-III criteria for a diagnosis	• English language learner status was not considered • Limited by sample size	• Test-retest reliability: .83 agreement, .54 agreement on positive scales, & .28 agreement on negative scales • Sensitivity based on presence or absence of a DSM-III diagnosis: .67 (parent), .45 (teacher) • Specificity based on presence or absence of a DSM-III diagnosis: .74 (parent), .91 (teacher) • Recommend higher cutoff points due to concerns of over identification of problem behaviors in Puerto Rican males

Table 3

Research on the CBCL/ASEBA in Latino Populations (cont'd)

Authors	Purpose	Sample	Limitations	Psychometrics and Findings
Lacalle, Ezpeleta, & Doménech (2012)	• Examined the validity of the CBCL and YSR for Spanish speaking children	• 420 Spanish-speaking students ages 8 to 17 • Clinically referred sample	• English language learner status was not examined • Compared to previous findings coefficients attained were lower, particularly in the areas of Behavior Problems and Affective Problems	• Internal consistency: .58 – .86 • Anxiety Problems Scale: < .6 • Behavior Problems Scale : > .8 (CBCL) • Affective Problems Scale: > .8 (YSR) Inter-rater reliability: .25 – .44, with raters disagreeing the most on the Behavior Problems Scale • Construct validity: .3 – .8 (the validity of the social phobia scale was the lowest for the CBCL; and specific phobia for the YSR)
LeBoeuf, Fantuzzo, & Lopez (2010)	• Examined the CBCL in diverse community-based sample of young, low-income children	• 2,757 children (28% Latino, 42% African American, 27% Caucasian)	• Community-based sample that was not nationally representative	• Internal consistency: .91 • The reliability and validity of the CBCL varied across groups (e.g., total score, externalizing problems and internalizing concerns); over-identified behavioral problems in Latino student populations • Recommend that the CBCL may be inappropriate for measuring behavioral problems in low-income preschoolers from culturally and linguistically diverse backgrounds
Gross et al. (2006)	• Examined the equivalence of CBCL/1½-5 of young children from diverse ethic, socio-economic, and linguistic backgrounds	• 682 African-American, White and Latino children aged 2 to 4	• English language proficiency level was not examined	• Internalizing subscale scores varied depending on the child's ethnicity • Latino children scored higher than non-Latino White and African-American children (p <.01) • DIF was detected in 1 item for the Latino middle upper income and non-Latino White low-income groups. • 7 items displayed DIF for the Latino low-income Spanish-speaking group
Kang (2005)	• Examined the equivalence of the CBCL in White, Black and Hispanic populations		• Small sample size	• The reliability of the CBCL was equivalent across ethnic groups • The validity varied among the groups, and significant differences existed in internalizing and externalizing problems for Hispanic children (p < .05)

preschool through 12th grade (Antshel & Remer, 2003; Conoley & Impara, 1995; Pfiffner & McBurnett, 1997). The SSRS was norm referenced across a predominately White sample that consisted of 4,170 children, 1,027 parents, and 259 teachers. Factor analyses revealed that the average internal consistency of the measures were .90 for the Social Skills scale, .84 for the Problem Behaviors scale, and .95 for the Academic Competence Scale. The SSRS self-rating form was also reported as demonstrating adequate internal consistency (Diperna & Volpe, 2005). An examination of the test-retest reliability of the SSRS revealed that the measure demonstrated correlations that ranged from .84 to .93 on the teacher scales and .65 to .80 on parent scales. Correlations that fell below .70 were found on student self-rating scales.

When the SSRS was compared to the Social Behavior Assessment (Stephens, 1992) and the CBCL-TRF, correlations fell between .50 and .60, which suggested moderate levels of construct validity. When additional scores were compared against parent and student measures, the measures were shown to maintain moderate to low levels of construct validity (Gresham & Elliott, 1990). Although recently revised and updated due to outdated norms and concerns regarding the usability and validity of the measure, the SSRS has widely advanced knowledge regarding the assessment of social skills.

Two studies have examined the use of the SSRS in culturally and linguistically populations. Manz et al. (1999) examined the psychometric properties of the preschool versions of the SSRS-P and SSRS-T across low-income African-American children. This study revealed several psychometric shortcomings of the SSRS for use with African-American children. First, a lack of construct validity was found to exist consistently across the measures. Second, the social skills scales and problem behavior scales did not offer distinct amounts of information, and both

self-control and interpersonal skills factors were found to be strongly and negatively correlated to problem behavior factors. Third, the SSRS-P and SSRS-T only had a mean correlation of .13, suggesting the presence of low inter-rater reliability.

Gonzalez (2000) used the SSRS-T to examine the relationship among classrooms settings, language, and social skills in preschool-aged ELLs. Through the monitoring of vocabulary skills, language development, and social skills using the SSRS, Gonzalez found greater levels of growth in the receptive and expressive language skills of Spanish-speaking preschoolers with language disorders in a segregated program. Although the psychometric properties of the SSRS-T were not extensively examined in this study, Spanish-speaking preschoolers enrolled in bilingual programs were identified as functioning at a higher level in terms of their social skills. Although the SSRS has been shown to maintain acceptable levels of validity when used in older Spanish-speaking ELL populations (Elliott, Barnard, & Gresham, 1989; Powless & Elliott, 1993; Van Horn, Atkins-Burnett, Karlin, Ramey, & Snyder, 2007) research has since emphasized the importance of validating social skills measures in culturally and linguistically diverse populations (Braccio, 2009).

Social Skills Improvement System (SSIS; Gresham & Elliott, 2008). Gresham and Elliott (2008) created the SSIS to address concerns regarding outdated norms and questionable psychometric properties. Both the Social Skills Intervention Guid (SSIG; Gresham & Elliott, 2008) and SSRS were updated and combined to create this new screening system (i.e., SSIS). Gresham and Elliott (2008) defined social skills as behaviors that are made up of seven constructs: communication, cooperation, assertion, responsibility, empathy, engagement, and self-control (see Table 4). Following a multi-tiered service of delivery and being used for universal screening and progress monitoring purposes, the SSIS retains the strengths of the SSRS

and includes additional subscales (Gresham & Elliott, 2008). The SSIS contains a Performance Screening Guide (PSG), a Class-wide Intervention Program (CIP), and a Selected/Targeted Assessment system, which includes teacher, parent, and student rating scales, as well as an intervention guide. The SSIS is considered to be one of the more promising screening and assessment tools for identifying social-emotional challenges in young children.

SSIS-Performance Screening Guide (PSG). Designed to function within a RtI system, the PSG offers a "time efficient, standardized, class-wide or school-wide screening of key social, motivational, and academic skills" (p.5, Gresham & Elliott, 2008). The PSG is a criterion-referenced measure that can be used to identify strengths and areas of concerns and also used as a progress-monitoring tool. For children ages 3 to 18, the measure is based on four skill areas, including Prosocial Behavior (i.e., effective communication, cooperative, exhibits self-control, and demonstrates empathy towards others), Motivation to Learn (i.e., engagement in instructional activities, ability to stay on task, effort displayed, and attentiveness), Reading skills, and Math skills (i.e., attending, participation, and general competence in the application of each skill area). The PSG ratings range from 1 (limited communication and/or cooperation skills, poor self-control, and difficulty in interacting with others) to 5 (excellent communication skills, self-control, and skills to interact with others). Teachers complete the ratings, which takes between 20-30 minutes to complte, by entering up to 25 student names, circling the matching level within each domain, and then transferring students with ratings of either a 1 or a 2 into two separate boxes. Besides demonstrating a close connection to previously developed multi-gated screening systems (e.g., SSBD and SSRS), the usability, reliability, and validity levels of the PSG have been established as being at least moderately adequate. Gresham and Elliott (2008) found that

98% of teachers agreed that the behaviors addressed in the PSG were valuable and that it was both useful and clearly written.

The authors demonstrated that the PSG had acceptable levels of test-retest reliability and inter-rater reliability. Using a sample of 543 students, over a period of 74 days, test-retest reliabilities were found to fall in a moderately acceptable range, .53 to .62, for preschoolers. Secondly, inter-rater reliabilities for a total of 434 students were approximately .73. Lastly, 85 students were evaluated using both the PSG and the full rating scales. A series of correlations were analyzed between the social skills rating scales and each subscale within the PSG. The rating scales correlated with the Pro-Social behavior scale at .63, Motivation to Learn at .67, early Reading skills at .61, and early Math skills at .81, suggesting moderate to substantial relationships at the preschool level. The authors did not report demographic data, and little is known regarding the development and psychometric properties of the PSG. In the validity study, 72% (i.e., 16 children) of the preschool-aged sample was white and only 13% (i.e., 3 children) was Hispanic. It is unclear whether or not these students were ELLs, and the appropriateness of the use of the PSG for students from culturally and linguistically diverse backgrounds is unknown. As previously observed across social-emotional and behavioral assessment tools, a lack of research currently exists regarding the use of the PSG in ELL populations, and particularly across ELP levels.

SSIS-Rating Scales. Three rating scales exist within the SSIS, including the (a) Teacher Rating Scale (SSIS-TRS), (b) Parent Rating Scale (SSIS-PRS), and (c) Student Rating Scale (SSIS-SRS). Two sets of teacher and parent rating scales were developed for preschool-aged students (i.e., three to five year olds) and kindergarten through 12th grade students. Student rating scales are available for 8-12 year olds and 13-18 year olds. Taking 15-20 minutes to complete,

the SSIS-TRS has 83 items, and the SSIS-PRS has 79 items. These new rating scales, which were adapted from the SSRS Rating Scales and Intervention Guide, offer improved norms for elementary school children, additional scales (i.e., Communication and Autism Spectrum), validity scales, and improved alignment across each form. (See Table 4 for an overview of the content of each subscale.) In total, the rating scales offer scores in four major domains: Social Skills, Problem Behaviors, Autism Spectrum, and Academic Competence.

Available in both Spanish and English, the SSIS rating scales were standardized on a nation-wide sample that consisted of 4,700 students ages 3 through 18 and was representative of race/ethnicity, socioeconomic status, and geographic region (Gresham & Elliott, 2008). Generally, the rating forms have been shown to have moderate to strong psychometric properties. The internal consistency reliabilities across the social skills and problem behaviors scales fell in the mid to upper .90s, which indicates that the scale scores are not influenced by sources of random error in the population that was sampled. Adjusted test-retest reliability coefficients for the SSIS-TRS across age levels fell in a moderately high range (i.e., .68 to .74), which indicates that raters' perceptions of social skills and problem behaviors are fairly stable. In an additional study completed by Gresham and Elliott (2008), inter-rater reliability ratings ranged from .37 for secondary students to .73 for preschool students, with most coefficients falling between .50 and .60.

The validity of the SSIS rating scales was measured using a sample of 85 students using both the PSG and the SSIS-TRS. The overall SSIS-TRS Social Skills subscale was correlated with the PSG Prosocial Behavior scale for preschoolers at .63, Motivation to Learn at .67, Reading skills at .61, and Math skills at .81, suggesting moderate to substantial relationships. At the elementary and secondary level correlations were .70, .58, .43, and .34 respectively. This

suggests that in elementary school students, the relationship between the PSG Prosocial Behavior scale and the SSIS-TRS is strong, and the relationships between the SSIS-TRS and Motivation to Learn, Reading skills, and Math skills scales are relatively weaker. Gresham and Elliott (2008) additionally found that the Assertion scale correlated the least across the SSIS-PRS and SSIS-TRS, and the Internalizing scale was only weakly and negatively correlated with the Motivation to Learn across the rating scales.

Although an in-depth analysis was conducted on the validity and reliability of the SSIS, a further examination of the rating scales is necessary to determine the appropriateness of using such a measure for ELLs in general. The convergent and discriminant validity of the SSIS-PRS and SSIS-TRS was found to be modest. Therefore, the authors emphasized the importance of considering context and rater effects. A differential item functioning (DIF) analysis found the items to be largely invariant across male and female, Caucasian and African American, and Caucasian and Hispanic children. However, ELL status was not taken into account, yet the measure was considered to have been validated on a nationally representative sample and is recommended for use across culturally and linguistically diverse students. In considering that Gresham and Elliott (2008) reported a wide range of coefficients that were based on data collected from a predominantly non-Hispanic, native English-speaking, Caucasian sample, differences might be even more exaggerated upon further examination of the measure in a predominantly ELL populations.

Research suggests that the SSIS is a promising measurement tool for elementary school-aged children. Therefore, this assessment system offers an ideal opportunity for researchers to examine its usability, reliability, and validity in ELL student populations. An unpublished thesis project (Felt, 2011) examined the validity and usability of the SSIS for Spanish-speaking ELLs;

although affected by a small sample size, Felt found reported data that suggested that the PSG maintains good part-whole reliability when used with Spanish-speaking ELLs and demonstrates higher levels of sensitivity than teacher referrals. The PSG was also found to have stronger NPP, but teacher referrals had better PPP. The Social Skills scale score within the SSIS-TRS and the teacher referrals only approached significance when correlated with the PSG classification. Significantly, negative correlations were found to exist between teacher referrals and the PSG Pro-social and Motivation scores, suggesting that Spanish-speaking ELLs who received low scores on the PSG were more likely to be referred. However, the PSG correlated highly with English proficiency level within the Pro-social and Motivation domains. Due either to an over-identification of students with lower English proficiency levels or to the fact that ELLs have been reported to have under-developed social skills, these results suggest that a significant difference exists in the way the social skills of Spanish-speaking ELLs are identified (Spomer & Cowan, 2001). Due to the small size and homogeneity across the sample, the extent to which the PSG influences teacher biases and the degree to which teacher referrals and the PSG account for a significant amount of variance across a student's social skills score on the SSIS-TRS is unclear. Felt (2011) ultimately reported that the PSG might inaccurately identify Spanish-speaking ELLs for social skills interventions. Thus, the usefulness of the PSG in coordination with the SSIS-TRS in linguistically diverse populations is unclear and requires further examination.

Table 4
SSIS Rating Scale Domain Areas and Subscales

Scale		General Description of Items Across Teacher and Parent Forms	Number of items – Teacher Form	Number of items – Parent Form
Social Skills			46	46
Communication		Responds well when others start a conversation; makes eye contact; uses appropriate tone of voice and body language; takes turns	7	7
Cooperation		Completes tasks without disturbing others; participates appropriately; pays attention to instructions; follows rules; helps others; ignores distractions in class	6	6
Assertion		Asks for help; expresses self when something is unfair; stands up for others; says nice things about self without bragging	7	7
Responsibility		Well-behaved when unsupervised; respects others' property; takes responsibility	6	6
Empathy		Tries to comfort or understand the point of view of others; feels bad when others are sad; shows concern for others	6	6
Engagement		Easily makes friends; invites others to join activities; starts conversations; introduces self to others	7	7
Self-control		Remains calm when teased or in a disagreement; appropriately accepts criticism; accepts comprise during a conflict; tolerates irritating peers	7	7
Problem Behaviors			30	33
Externalizing		Struggles to wait for turn; fidgets; bullies; talks back; fights with others; cheats/lies in games; disobeys rules; has temper tantrums	12	12
Bullying		Bullies others; scares others; forces others to act against their will; aggressive toward others	5	5
Hyperactivity/ Inattention		Acts without thinking; has difficulty waiting for turn; inattentive; breaks into or stops games; gets distracted; fidgets or moves around too much	7	7
Internalizing		Withdraws; acts anxious/lonely/sad; has low energy; says bad things about self	7	10
Autism Spectrum		Takes turns in conversations/activities; makes eye contact; uses appropriate body language; has nonfunctional routines; repeats things over and over; preoccupied with object parts; exhibits stereotyped motor behaviors	Part A: 8 Part B: 7	Part A: 8 Part B: 7
Academic Competence			N/A	7
Academic Competence		Overall academic performance; reading; mathematics; grade-level expectations; motivation to learn	N/A	7

Note. Adapted from Gresham, F. M., & Elliott, S. N. (2008). *Rating Scales Manual: SSIS – Social Skills Improvement System*. Minneapolis, MN: NCS Pearson, Inc.

Summary and Implications

Overall, research has demonstrated that social skills are clearly associated with academics, social-emotional functioning, and mental health. Early social skills programs have been linked with improved outcomes across pre-school and elementary school populations, but students must first be accurately identified so that they can receive the necessary interventions. Presently, the assessment of social skills development within a prevention-based framework (e.g., RtI model) appears promising. However, inaccurate identification rates and possible bias in measurement systems for ELLs have led to the recognition of the need for research on assessment protocols that measure the development of social skills. Spanish-speaking students are the fastest growing demographic across ELL populations, and exhibit distinct and dynamic social and behavioral characteristics that are different from their native English-speaking majority peers. Spanish-speaking ELLs may therefore be at risk for neither being screened effectively nor receiving services appropriately. A further examination of social skills assessment measures is particularly necessary because such assessment systems have neither been validated nor demonstrated as being useful for identifying strengths and areas of concern in Spanish-speaking ELLs. The SSIS, which is a promising new screening system, is one measurement tool that would benefit from further examination of its psychometric properties, validity, and usability for young Spanish-speaking ELLs.

CHAPTER II

Statement of The Problem and Research Questions

Statement of The Problem

As the United States struggles to improve educational outcomes and deliver mental health services to a diverse citizenry, stakeholders have widely promoted prevention efforts (Chapman, Laird, & KewalRamani, 2010; Huey, & Polo, 2008). Early identification and programming that address problem behaviors and academic skill deficits are essential for solving costly problems, decreasing school drop out rates, and preventing the advancement of psychiatric disorders (Adelman & Taylor, 2010). Children with poor social skills and psychosocial functioning tend to display low academic achievement and social maladjustment (Darney et al. 2012; Malecki & Elliot, 2002). Well-developed social skills are linked with positive relationships, appropriate mental health functioning, and academic success (Algozzine et al., 2011). Academic skill development is critical for physical, emotional, and vocational wellness (Brown-Chidsey, 2005). Academic skill deficits are associated with youth pregnancy, incarceration, mental illness, and poor health (CDC, 2005; Matson & Haglund, 2000; Strom, 2000). Ethnically diverse students experience difficulties in obtaining such skills, which has lead to negative outcomes that have persisted over the past decade. Recognizing that social and relational experiences impact school performance and mental health, researchers have begun examining tools that effectively identify and facilitate service delivery for ethnically diverse students (Diperna, 2005; Landy, 2009). Fortunately, social skills can be addressed through early intervention, but appropriate and valid assessment systems are first needed (Dever & Kamphaus, 2013; Severson et al., 2007).

There currently is a lack of empirical evidence regarding the validity and reliability of social skills measurement systems for students from linguistically diverse backgrounds,

particularly when considering the measurement of social skills across the wide continuum of ELP levels (Reynolds & Lowe, 2009; Solano-Flores, 2011). The current study examined the extent to which the SSIS' PSG and SSIS-TRS identified strengths and areas of concern in Spanish-speaking ELLs across varying ELP levels, as compared to native English-speaking students, for the purpose of determining which students are in need of further social-emotional or behavioral support. In considering systems that examine social, emotional, and behavioral functioning, researchers have generally failed to account for language proficiency levels in the construction of scale norms and the analysis of the psychometric properties of assessment systems (Dever & Kamphaus, 2013; Reynolds & Lowe, 2009). Specifically, there is a paucity of data regarding the reliability and validity of the SSIS-TRS and PSG in Spanish-speaking ELL populations (Gresham & Elliott, 2008). Although the normative sample that was used to validate the SSIS included a nationally representative sample of *Hispanic* students according to the 2006 U.S. Census Bureau, ELL status was neither reported nor apparently considered. Due to the high proportion of Spanish-speaking ELLs compared to other ELLs enrolled in U.S. public school systems, this study focused on this particular subpopulation (Rhodes et al., 2005). To determine the usefulness of the PSG in identifying ELLs in need of further assessment, and the validity of the SSIS-TRS across native Spanish-speaking ELLs, this research also examined outcome and item scores across this particular linguistically diverse group. Ultimately, this study attempted to further establish a research base for the use of the PSG and SSIS-TRS in Spanish-speaking ELL populations, and advance our current understanding of the way social-emotional assessment systems measure functioning in culturally and linguistically diverse learners.

Research Questions

The goal of the present study was to evaluate the validity of the PSG and SSIS-TRS

across Spanish-speaking ELLs and native English-speaking students. First, the reliability, construct validity, predictive validity, concurrent validity, specificity, and sensitivity of the PSG and SSIS-TRS were examined through a review of correlation-based analyses, chi-square tests, and regression analyses. Second, the researcher examined the extent to which the PSG and the SSIS-TRS were invariant across groups that vary by ELP level through the application of logistic regression as rooted in DIF. An exploratory analysis using mean level comparisons was used to examine differences across ELP levels and grade levels. This study sought to inform the psychometric properties of the PSG and SSIS-TRS in classifying social skills functioning, academic achievement, and social-emotional and behavioral performance in Spanish-speaking ELLs. Based on this research design the following research questions were investigated:

1) Do the PSG and SSIS-TRS demonstrate appropriate levels of reliability, in Spanish-speaking ELL populations as evidenced through:

    a. Internal consistency coefficients; and

    b. part-whole reliability coefficients?

2) Do the PSG and SSIS-TRS demonstrate appropriate levels of validity in Spanish-speaking ELL populations as evidenced through:

    a. Statistically significant correlations across the measures (i.e., PSG, teacher referral form and SSIS-TRS);

    b. moderate to high levels of construct validity as demonstrated by no statistically significant correlations with grade, gender, percentage of ELLs in the classroom, ELL status, and ELP level;

    c. adequate levels of predictive validity when comparing the PSG to outcomes on the SSIS-TRS, above what is observed with teacher referral;

      d. appropriate levels of sensitivity and specificity?

3) Does the PSG function similarly across native English-speaking and Spanish-speaking ELL populations, and demonstrate adequate levels of construct validity, as evidenced by scores that are invariant across across scales (i.e., Math skills, Reading skills, Motivation to Learn, and Prosocial Behavior) on the PSG?

4) Does the SSIS-TRS function similarly across native English-speaking and Spanish-speaking ELL populations and demonstrate adequate levels of construct validity as evidenced by scores that are invariant across items, 11 subscales (i.e., Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, Self-control, Externalizing, Bullying, Hyperactivity/Inattention, and Internalizing), and four major scales (i.e., Social Skills, Problem Behaviors, Autism Spectrum, and Academic Competence)?

5) Does the SSIS function similarly across students of varying ELP levels as demonstrated by:

      a. Invariance in scores across domain areas (i.e., Math skills, Reading skills, Motivation to Learn, and Prosocial Behavior) on the PSG; and

      b. invariance in scores across items, 11 subscales (i.e., Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, Self-control, Externalizing, Bullying, Hyperactivity/Inattention, and Internalizing), and four major scales (i.e., Social Skills, Problem Behaviors, Autism Spectrum, and Academic Competence)?

CHAPTER III

Method

Sample and Setting

The purpose of this research was to determine the reliability and validity of the PSG and SSIS-TRS across elementary school Spanish-speaking ELLs. Elementary school students were recruited and teachers were asked to rate students within their classrooms using the PSG and SSIS-TRS. Teachers from kindergarten through fifth grade classrooms were selected. These grades were chosen because the teacher rating scale for elementary students begins in kindergarten, and developmental differences are noted both at the student level and across educational systems. Gresham and Elliott (2008) further suggest that the SSIS functions differently across elementary and secondary levels. To reduce bias, teachers were neither recruited nor screened based on any particular characteristic (e.g., classroom structure, training history, teaching style, school type).

For this study, a number of considerations were taken into account to determine an appropriate sample size. First, preliminary research suggests that social emotional rating scales may function differently across ELL and native English-speaking students (e.g., Dowdy et al., 2011; Felt, 2011). Second, given the nature of this study, resources and time were limited. Third, the feasibility of detecting DIF in small sample sizes is not definitively known. Therefore, given the number of students recruited in other small scale studies this research study ultimately aimed to collect a sample of at least 200 students in total (i.e., 100 ELLs, and 100 native English-speaking students). To maintain an equal number of ELLs and native English speakers, teachers were expected to complete the PSG and SSIS-TRS for 6 students, and at least 34 teachers were targeted. The order in which teachers completed the measures was counterbalanced (e.g.,

Teacher Referral form and PSG, see below for further description of measures) to control the effects that order of completion may have had on the results.

Demographics, including gender, ethnicity, bilingual status, and total years of teaching experience were collected for each participating teacher. Teachers also reported the total number of students and number of ELLs in their classroom. Demographic data about students who were rated by their teachers using the SSIS-TRS were collected. Student information that was collected included gender, ELP level, history of grade retention, and current grade level.

Recruitment

To access student ELP levels in a consistent and uniform manner, teachers were recruited from a Midwestern state that is a member of the World-Class Instructional Design and Assessment (WIDA) Consortium. This particular state within the WIDA Consortium utilizes the Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs®; WIDA Consortium, 2010) that measures ELP levels on a 6 level proficiency continuum, with students typically exiting ELL programming when they achieve an ELP level of 6 (see Table 5).

Due to a limited amount of resources, a convenience sample was used to recruit participants. The researcher targeted school districts throughout a Midwestern state. Additional districts with previous connections to research projects conducted through the University of Wisconsin-Madison were also contacted. District level superintendents were first contacted through phone and email. Once consent was given the building-level principals were contacted. Participating teachers received an honorarium in the form of $50 dollars.

Informed Consent

Approval for this study was obtained from the University of Wisconsin-Madison

Education Research Institutional Review Board. District level superintendents, building

principals, teachers, and parents were asked to complete consent forms.

Sample and Descriptive Analyses

Descriptive statistics for student demographics are presented in Table 6 with a total

sample size of 238 students and 33 teachers. A total of 17 teachers completed the Teacher

Referral form following the PSG and 16 completed the PSG followed by the Teacher Referral

form (see descriptions of the measure below)**.** Social skills ratings were ultimately provided for a

total of 238 students, including 120 native English-speaking students with a total of 62 girls

(51.6%) and 58 boys (48.4%), and 118 ELLs with a total of 55 girls (46.6%) and 63 boys

(53.3%). ELP levels for ELL students were measured using the ACCESS for ELLs® (WIDA

Consortium, 2010). Results show that 11.9**%** of ELL participants were classified at an Entering,

or Level 1 proficiency of English; 22.9**%** at Level 2, Beginning; 39**%** at Level 3, Developing;

12.7**%** at Level 4, Expanding; 11.9% at Level 5, Bridging; and 1.7% at Level 6, Reaching. Most

of the ELLs (n=77; 65.2%) were in classrooms with 26-75% of the students being ELLs, but the

native English-speaking students (n=75; 62.5%) were predominately in classrooms with less than

25% of the students being ELLs.

Teacher demographics are presented in Table 7. Thirty three teachers participated in this

study. Of these, the majority (81%) had at least 5 years of teaching experience, with

approximately 45% of the total sample having 5-10 years, two teachers having 3-5 years of

teaching experience, and four teachers having less than three years of experience. Although male

teachers were recruited none were retained in this sample. Three teachers self-identified as

Latino, and one teacher reported being American Indian. While 12 teachers (36%) reported being

monolingual, 21(63%) teachers reported being at least bilingual, with 14 of these teachers

reporting fluency in both English and Spanish. Teachers reported a varying range in terms of the

number of Spanish-speaking ELLs in their classroom, with 45% of the teachers (n=15) having a

classroom that consisted of no more than 25% Spanish-speaking ELLs, while 3 teachers reported

having at least 76% of their class made up of Spanish-speaking ELLs. Of the 33 classrooms 21

(63%) were dual language immersion classrooms. The majority of teachers (78%) in the sample

had also obtained a Masters degree.

Table 5
Student Sample Demographics (n=238 total; Spanish-speaking English Language Learners
n=118, Native English Speakers n =120; Demographics for Spanish speaking English language
learners are indicated in parentheses)

	N	%
Gender		
Male	57(63)	47.5(53.4)
Female	63(55)	52.5(46.6)
ELLs		
Level 1, Entering	(14)	(11.9)
Level 2, Beginning	(27)	(22.9)
Level 3, Developing	(46)	(39.0)
Level 4, Expanding	(15)	(12.7)
Level 5, Bridging	(14)	(11.9)
Level 6, Reaching	(2)	(1.7)
Grade		
Kindergarten	6(8)	5.0(6.8)
1	35(39)	29.2(33.1)
2	30(31)	25.0(26.3)
3	31(15)	25.8(12.7)
4	12(14)	10.0(11.9)
5	6(11)	5.0(9.3)
Percentage of ELLs in classroom		
<25%	75(14)	62.5(11.8)

26-50%	20(36)	16.7(30.5)
51-75%	17(41)	14.2(34.7)
>76%	8(27)	6.7(23.0)
Type of Classroom		
Dual-Language Program	27(60)	22.5(50.8)
Non-Dual-Language Program	93(58)	78.8(49.2)

Table 6
Teacher Sample Demographics (n=33)

	N	%
Years of Teaching Experience		
< 3	4	12.1
3-5	2	6.1
5-10	15	45.5
10-20	7	21.2
>20	5	15.1
Gender		
Male	0	0.0
Female	33	100.0
Grade		
Kindergarten	1	3.0
1st	10	30.3
2nd	13	39.4
3rd	4	12.1
4th	2	6.0
5th	3	9.1
Race		
American Indian	1	3.0
White	30	90.9
Other	2	6.12
Ethnicity		
Latino	3	9.0

Non-Latino	30	90.9
Type of Classroom		
Dual-Language Program	21	63.6
Non-Dual-Language Program	12	36.4
Bilingual Status		
Not bilingual	12	36.4
Bilingual in Spanish	14	42.4
Bilingual in other language	7	21.2
Percentage of ELLs in classroom		
<25%	15	45.5
26-50%	8	24.2
51-75%	7	21.2
>76%	3	9.1
Highest Degree Achieved		
BA/BS	6	18.2
AA	1	3.0
Masters	26	78.8

Table 8 provides classification rates for all students based on teacher referral and the SSIS-PSG. A total of 187 students (i.e., 78 ELLs and 109 native English-speaking students) were screened using the teacher referral form and PSG. The teacher referral form identified a total of 28 (23.3%) native English speaking and 23 (19.5%) Spanish-speaking ELLs as students in need of social skills intervention. The PSG identified a total of 79 (65.8%) native English-speaking students and 74 (62.7%) Spanish-speaking ELLs in need of social skills intervention, with the majority of each group consisting of students as students experiencing moderate difficulty. A total of 15 native English-speaking students and 12 Spanish-speaking ELLs were identified as experiencing significant difficulty. Finally, the SSIS Teacher Rating Scale (TRS) identified 15 native English-speaking students and 42 Spanish-speaking ELLs as students in need of social

skills intervention, according to the TRS social skills scores that were in the below average range of functioning.

Regarding missing data, of all of the students who were assessed, some did not go through the screening process (e.g., 12 native English-speaking students and 46 Spanish-speaking students were not screened using the teacher referral form and 11 native English-speaking students and 40 Spanish-speaking ELLs were not screened using the PSG). Subsequently, a total of 187 students were analyzed across the SSIS-TRS, PSG and Teacher Referral, with 78 of these students being Spanish-speaking ELLs.

Table 7
Demographics: Students Classified as in Need of Social Skills Intervention
(Deomographics for Spanish speaking ELLs are indicated in parentheses)

	N	%
Teacher referral	28 (23)	23.3 (19.5)
Missing	12 (46)	10.0 (39.0)
PSG		
PSG – Moderate Social Skills difficulty	64 (62)	53.3 (52.0)
PSG – Significant Social Skills difficulty	15 (12)	12.5 (10.2)
Missing	11 (40)	9.2 (33.9)
TRS	15 (42)	12.5 (35.6)

Measures

The measures that were used included (a) the ACCESS for ELLs®, (b) the SSIS Performance Screening Guide, (c) the SSIS Teacher Rating Scale (K-5th grade), (d) a Teacher Referral Form, and (e) a teacher survey. Each of these are described in more detail below.

ACCESS for ELLs®. ELP levels were identified for Spanish-speaking ELLs using the ACCESS for ELLs®. Although a wide variety of ELP measures currently exist, many were developed before the enactment of No Child Left Behind (NCLB; No Child Left Behind [NCLB],

2002). Whereas these pre-NCLB measures (e.g., Language Assessment Scales, Idea Proficiency Test, and Language Proficiency Test Series) tended to focus on social language proficiency, the passage of NCLB required that ELP measures examine academic language proficiency, resulting in a new generation of such measures, such as the ACCESS for ELLs[®] (Albers et al., 2009) and its new 2.0 version, which is computer-based ass of 2015. Each student's ELP level is determined by scores obtained annually on the ACCESS for ELLs[®], the language proficiency measure used with ELL students in the 35 WIDA Consortium states. Based on the ELP standards, the ACCESS for ELLs[®] performance indicators describe expectations for ELLs at multiple grade clusters (K, 1, 2–3, 4–5, 6–8, and 9–12) and across five content areas: social and instructional language, language arts, math, science, and social studies. Additionally, within each of these content areas, four language domains are measured: speaking, listening, reading, and writing. The test assesses progress within each of these domains using six proficiency levels: (a) entering, (b) beginning, (c) developing, (d) expanding, (e) bridging, and (f) attained. Designed to assess annual gains in English proficiency, ACCESS for ELLs[®] is used by schools for making ELL placement decisions and evaluating annual measurable achievement objectives (AMAOs). Schools are required to annually assess a student's ELP level if the parent reports that a language other than English is spoken at home. Therefore, the researcher collected the student's most recent available ACCESS for ELLs® score. The researcher contacted the school to obtain student ACCESS for ELLs® scores. Teachers were asked to provide the overall composite score from the ACCESS for ELLs®.

SSIS Performance Screening Guide. The SSIS Performance Screening Guide (PSG; Elliott & Gresham, 2008) social skills screener was of interest for this study. As previously discussed, the PSG is a standardized, criterion-referenced measure of socio-emotional behavior

that reflects age- and grade-level expectations. The PSG has four main scales: Prosocial Behavior, Motivation to Learn, Reading skills, and Math skills. Ratings are based on a 5-point grading scale with a score of 1 in any skill area indicating a high level of concern, need for further evaluation, and potential instructional evaluation. A score of 2 indicates a moderate level of concern, with a potential need for further evaluation and instructional support; a score of 3 or 4 or higher indicates the student is functioning at or above average and further evaluation is not necessary. Students therefore fall into three categories: experiencing a high level of difficulty, experiencing a moderate level of difficulty, and demonstrating average or above average levels of functioning as compared to their same age peers. The elementary/secondary level version of the PSG was used because this research focused on elementary school-aged students. As previously noted teachers can complete the survey within 20-30 minutes by entering up to 25 student names, circling the matching level within each domain, and then transferring students with ratings of either a one or a two into two separate boxes. Besides demonstrating a close connection to previously developed multi-gated screening systems (e.g., SSBD and SSRS), the usability, reliability, and validity levels of the PSG have been established as being at least moderately adequate, with appropriate test-retest reliability and inter-rater reliability. Gresham and Elliott (2008) found that 98% of teachers agreed that the behaviors addressed in the PSG were valuable and that it was both useful and clearly written. The authors examined the reliability and validity of the measure using a sample of 543 students. Inter-rater reliabilities for a total of 434 students were approximately .73. The authors did not report demographic data, and little is known regarding the development and psychometric properties of the PSG. It is unclear whether or not any students included in the sample were ELLs.

SSIS-Teacher Rating Scale. The SSIS-TRS form for students in grades kindergarten through 12th grade (TRS; Elliott & Gresham, 2008) was also examined in this study in order to inform the ways in which this measure functions in ELL student populations versus native English-speaking student populations. The use of this social skills measure is supported by a standardized sample and maintains adequate norms and psychometric properties. Developed based on a previous widely accepted assessment system (i.e., SSRS), the current status of this measure suggests it can be of use in ELL populations. However, as previously mentioned, a further examination of the technical adequacy of this measure in ELL populations is necessary. The SSIS-TRS includes major scales: Social skills, Problem Behaviors; and 12 subscale scores: Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, Self-control, Externalizing, Bullying, Hyperactivity/Inattention, Internalizing, and Autism. As previously mentioned the SSIS rating scales are available in both Spanish and English and the scales were standardized on a nation-wide sample that consisted of 4,700 students ages 3 through 18 and was representative of race/ethnicity, socioeconomic status, and geographic region (Gresham & Elliott, 2008). Generally, the rating forms have been shown to have moderate to strong psychometric properties. The internal consistency reliabilities across the social skills and problem behaviors scales fell in the mid to upper .90s, which indicates that the scale scores are not influenced by sources of random error in the population that was sampled. Adjusted test-retest reliability coefficients for the SSIS-TRS across age levels were reported, with values that fell in a moderately high range (i.e., .68 to .74), which indicates that raters' perceptions of social skills and problem behaviors are fairly stable. The validity of the SSIS rating scales was measured using a sample of 85 students using both the PSG and the SSIS-TRS. The overall SSIS-TRS Social Skills subscale was correlated with the PSG Prosocial Behavior scale at .70, Motivation to

Learn at .58, Reading skills at .43, and Math skills at .34, suggesting moderate to substantial

relationships. Gresham and Elliott (2008) additionally found that the Assertion scale correlated

the least across the SSIS-PRS and SSIS-TRS, and the Internalizing scale was only weakly and

negatively correlated with the Motivation to Learn across the rating scales. Although an in-depth

analysis was conducted on the reliability and validity of the SSIS, a further examination of the

rating scales is necessary to determine the appropriateness of using such a measure for ELLs in

general. The convergent and discriminant validity of the SSIS-PRS and SSIS-TRS was found to

be modest. Therefore, Gresham and Elliott (2008) emphasized the importance of considering

context and rater effects. A differential item functioning (DIF) analysis found the items to be

largely invariant across male and female, Caucasian and African American, and Caucasian and

Hispanic children. However, ELL status was not taken into account, yet the measure was

considered to have been validated on a nationally representative sample and is recommended for

use across culturally and linguistically diverse students. This measure therefore served as the

classification tool to which PSG screening outcomes were compared. This study also examined

the major scales, subscales, and items throughout the SSIS-TRS across both Spanish-speaking

ELLs and native English-speaking students.

      Teacher Referral Form. Teacher referral is a widely used practice for identifying

students with possible socio-emotional and academic difficulties (Rhodes et al., 2005; Whitcomb

& Merrell, 2012). Teacher judgment has been found to be useful in evaluating academic and

behavioral functioning. According to Elliott et al. (2007), teacher referrals have also been used in

the creation of many standardized tests, including both IQ measures and mental health

assessment systems. To compare students that would be identified using the PSG, teachers were

Table 8. ACCESS for ELLs® ELP Levels

	Performance Level Descriptions WIDA English Language Proficiency Level Standards
	At the given level of ELP, a student will process, understand, produce or use:
Level 6 Reaching	• Specialized or technical language reflective of the content area at grade level; a variety of sentence lengths of varying linguistic complexity in extended oral or written discourse as required by the specified grade level; oral or written communication in English comparable to proficient English peers
Level 5 Bridging	• The technical language of the content areas; a variety of sentence lengths of varying linguistic complexity in extended oral or written discourse, including stories, essays, or reports; oral or written language approaching comparability to that of English proficient peers when presented with grade-level material
Level 4 Expanding	• Specific and some technical language of the content areas; a variety of sentence lengths of varying linguistic complexity in oral discourse or multiple, related paragraphs; oral or written language with minimal phonological, syntactic, or semantic errors that do not impede the overall meaning of the communication when presented with oral or written connected discourse with occasional visual and graphic support
Level 3 Developing	• General and some specific language of the content areas; expanded sentences in oral interaction or written paragraphs; oral or written language with phonological, syntactic, or semantic errors that may impede the communication but retain much of its meaning when presented with oral or written, narrative or expository descriptions with occasional visual and graphic support
Level 2 Beginning	• General language related to the content areas; phrases or short sentences; oral or written language with phonological, syntactic, or semantic errors that often impede the meaning of the communication when presented with one to multiple-step commands, directions, questions, or a series of statements with visual and graphic support
Level 1 Entering	• Pictorial or graphic representation of the language of the content areas; words, phrases, or chunks of language when presented with one-step commands, directions, WH-questions, or statements with visual and graphic support

Note. Adapted from "Annual technical report no. 5, Vol. 1 of 3: Description, Validity, and Student Results. *Annual technical report for ACCESS for ELLs® English Language Proficiency Test, Series 200, 2008-2009 administration"* WIDA Consortium, 2010, Madison, WI: Author.

asked to provide a list of students they would refer if they otherwise did not have access to a screening guide (see Appendix A). Therefore, teachers were asked to list students they would consider referring. Teachers were asked to consider their classroom as a whole and list each student within the classroom, regardless of their special education status, for whom they had concerns (i.e., poor school performance and would like further evaluation). Teachers were not required to identify a certain number of students. Due to concerns regarding biases associated with referral processes (e.g., Abidin & Robinson, 2002; Podell, & Soodak, 1993; Portes, 2013), this research examined whether students, particularly those who were ELLs, that were referred using the PSG differed from students that teachers referred without having used such a screening guide. Therefore, teacher referrals served as a secondary outcome classification measure to which both the PSG and SSIS-TRS were compared.

Procedure

After receiving consent forms from parents and guardians, teachers completed the PSG or the teacher referral form. Teachers were then prompted to complete the corresponding screening procedure (i.e., the PSG, or teacher referral). The order of the screening procedures was counterbalanced. Teachers then completed the SSIS-TRS for students who were identified through a specific set of criteria (see Appendix B). Lastly, teachers completed a teacher survey (see Appendix C).

The SSIS-TRS is a measure that can be used at the tier 2 or tier 3 level within a RtI framework. In accordance with this framework, no more than 20% of a student population should need to access supports available within these tiers (Jimerson et al., 2007). With the U.S. Department of Education estimating average class sizes to be approximately 25 students (OECD, 2011), teachers in this study were asked to complete the SSIS-TRS for a total of 6 students based

on their scores from the PSG. Limiting the number of students that teachers can further evaluate minimized fatigue. Teachers then completed the SSIS-TRS for three native English-speaking students that received PSG scores that fell in the most critical range of functioning (i.e., highest number of ones in any skill area). If fewer than three students received a score of one in any of the four skill areas, teachers then completed rating scales for students that received the highest number of twos and then if necessary the highest number of threes, until rating scales were completed for three students. If there were multiple students that had the same number of ratings, teachers were prompted to choose the student with the corresponding letter that is closest to A, with Z being the furthest away, until three students were selected. Teachers followed the same procedure for Spanish-speaking ELLs in their classroom. Student demographic information was provided with each SSIS-TRS. Before completing the SSIS-TRS the teacher sent home consent letters with students that met criteria. After the researcher received the letters of consent, teachers were prompted to complete the SSIS-TRS for each student.

Teachers were trained in the use of the PSG and SSIS-TRS through a standard printed script, which included an overview of each measure, highlights from the directions were included on each form, and a simplified description of ways to complete each form was also provided. An overview of the purpose of the study, as well as a discrete set of directions, was provided to each teacher (see Appendix B). Teachers were asked to report the amount of time that they spent completing the PSG and the teacher referral procedure. The researcher was available as needed when teachers required further assistance.

Research Questions and Hypotheses

1) Do the PSG and SSIS-TRS demonstrate appropriate levels of reliability, in Spanish-speaking ELL populations as evidenced through:

a. Internal consistency coefficients; and

b. part-whole reliability coefficients?

Hypothesis 1(a): As previously demonstrated by the results from Felt (2011), the

PSG and the SSIS-TRS are expected to maintain moderate to high levels of internal

consistency in Spanish-speaking ELL populations.

Hypothesis 1(b): The PSG and SSIS-TRS are also expected to maintain moderate to

high levels of part-whole reliability based on previous research (Felt, 2011;

Gresham & Elliott, 2008).

2) Do the PSG and SSIS-TRS demonstrate appropriate levels of validity in Spanish-

speaking ELL populations as evidenced through:

a. Statistically significant correlations across the measures (i.e., PSG, teacher

referral form and SSIS-TRS);

b. moderate to high levels of construct validity as demonstrated by no statistically

significant correlations with grade, gender, percentage of ELLs in the classroom,

ELL status and ELP level;

c. adequate levels of predictive validity when comparing the PSG to outcomes on

the SSIS-TRS, above what is observed with teacher referral;

d. appropriate levels of sensitivity and specificity?

Hypothesis 2(a): Based on previous research (Gresham & Elliott, 2008; Felt,

2011) that suggests that the PSG and the SSIS-TRS measure similar areas of

functioning, the measures will be expected to maintain acceptable levels of

construct validity. Given the limitations in Felt's (2011) sample, when only

Spanish-speaking ELLs are considered, the results from this study are predicted to

demonstrate inconsistent correlational patterns since research suggests that behavior is culturally bound and impacted by linguistic variables (Abedi, 2013; Rivera & Rogers-Atkinson, 1997).

Hypothesis 2(b): The PSG, teacher referral form, and SSIS-TRS will demonstrate moderate to high levels of construct validity as demonstrated by no statistically significant correlations with grade, gender, or classroom. This analysis would suggest that the PSG is not biased toward particular student groups. Such outcomes would indicate that the results of the PSG and SSIS-TRS scales (i.e., social skills, problem behaviors, autism spectrum and academic competence) are not systematically based on other variables, and maintain sufficient levels of construct validity, except for students who are ELLs and for students across ELP levels. When ELL status and ELP levels are taken into account the results of the PSG and SSIS-TRS will systematically vary, and the measures will demonstrate inadequate construct validity. Specifically, Spanish-speaking ELLs will be identified as demonstrating poorer adaptive skills and social skills than native English speaking students.

Hypothesis 2(c): Based on the extant literature on screening tools, the PSG is not predicted to offer an advantage over teacher referral in predicting social skills deficits (Felt, 2011; Miller, 2015). When English language proficiency levels are taken into account, neither the PSG nor the teacher referral form are predicted to demonstrate high levels of predictive validity when examining outcomes on the SSIS-TRS social skills scores.

Hypothesis 2(d): Based on previous research conducted on the SSIS (Felt, 2011) the sensitivity and negative predictive power (NPP) of the PSG are predicted to be technically adequate, while the teacher referral form is predicted to maintain adequate specificity and higher positive predictive power (PPP) when compared to the PSG.

3) Does the PSG function similarly across native English-speaking and Spanish-speaking ELL populations, and demonstrate adequate levels of construct validity, as evidenced by scores that are invariant across across scales (i.e., Math skills, Reading skills, Motivation to Learn, and Prosocial Behavior) on the PSG?

Hypothesis 3: Based on research that suggests that linguistic variables impact teacher perception of academic skills and behaviors (Abedi, 2013; Rhodes, et al., 2005), and that only after controlling for ELL status were the PSG outcome scores predictive of a student's social skills score (Felt, 2011), the PSG is not predicted to be invariant across both Spanish-speaking ELL and native English-speaking students.

4) Does the SSIS-TRS function similarly across native English-speaking and Spanish-speaking ELL populations and demonstrate adequate levels of construct validity as evidenced by scores that are invariant across items, 11 subscales (i.e., Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, Self-control, Externalizing, Bullying, Hyperactivity/Inattention, and Internalizing), and four major scales (i.e., Social Skills, Problem Behaviors, Autism Spectrum, and Academic Competence)?

Hypothesis 4: Based on research that suggests that linguistic variables impact teacher perception of academic skills and behaviors (Abedi, 2013; Rhodes, et al., 2005) the SSIS-TRS is not predicted to be invariant across Spanish-speaking ELL and native English-speaking students.

5) Does the SSIS function similarly across students of varying ELP levels as demonstrated by:

    a. invariance in scores across domain areas (i.e., Math skills, Reading skills, Motivation to Learn, and Prosocial Behavior) on the PSG; and

    b. invariance in scores across items, 11 subscales (i.e., Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, Self-control, Externalizing, Bullying, Hyperactivity/Inattention, and Internalizing), and four major scales (i.e., Social Skills, Problem Behaviors, Autism Spectrum, and Academic Competence)?

Hypothesis 5(a): Since research suggests that behavior is culturally bound and impacted by linguistic variables, statistically significant differences are predicted to occur in scores across domain areas on the PSG across ELP levels (Abedi, 2013; Rivera & Rogers-Atkinson, 1997). Due to a lack of research that has examined the impact of ELP levels and teacher perceptions of student behavior, it is generally unclear whether the PSG will function differently across students with varying ELP levels (Abedi, 2013; Dever & Kamphaus, 2013; Reynolds & Lowe, 2009; Rhodes et al., 2005).

Hypothesis 5(b): Furthermore, given that behavior is culturally bound and impacted by linguistic variables, statistically significant differences are predicted

to occur in scores across items, subscales, and major scales, across students with varying ELP levels.

CHAPTER IV

Results

Analyses and Background

The purpose of this study was to determine the validity and reliability of the PSG and SSIS-TRS across elementary school Spanish-speaking ELLs. Figure 2 below contains a consort diagram that provides an overview of the sample and analyses. A review of research that has applied DIF suggests that such a methodology is well suited for this study, as it allows for a broad investigation into test and item functioning across different groups. A DIF analysis would allow for the examination of the occurrence of different probability distributions, assuming dependency on group membership and performance on the item. Osterlind and Everson (2009) further explain that:

> [If the] condition probability...is not idenfitical for the reference and focal groups...a dependency is present between group membership and performance on the item once we control for [the parameter]. When a test item is scored dichotomously, [the] conditional dependence suggests that the reference and focal group members at the same level [of the parameter] have a different probability of correct response, with the group having the lower conditional probability of correct response being disadvantaged by that particular test item. This line of analysis, then, leads us to conclude that the test item functions differently for the two groups and, consequently, that DIF is extant. (p. 9)

Various applications of DIF were considered. In using item response theory (IRT) a large sample size (i.e., ranging from 500 to 1000) is required across both the focal (i.e., group of interest or Spanish-speaking ELLs) and reference groups (i.e., native English-speaking students; Osterlind & Everson, 2009). However, when high levels of DIF are present, smaller sample sizes

Figure 2: Flowchart: Overview of Study Sample and Analysis

Enrollment

Student and Teacher Sample
(n = 238 [240*] students; 33 [40*] teachers)
Schools (n = 14 [14*] with 3 DLI programs)

Measures
PSG (n = 187 [240*] students; 78 [120*] ELLs)
Teacher referral form (n = 187 [240*] students; 78 [120*] ELLs)
SSIS-TRS (n = 238 [240*] students; 118 [120*] ELLs)

Randomized (n = 238 [240*] students)

Randomization

Completed Teacher Referral then PSG
(n = 118 [120*] students; 17 [120*] teachers)

Completed PSG then Teacher Referral
(n = 120 [120*] students; 16 [20*] teachers)

Total Sample

Analysis

Reliability
PSG (n = 187 [240*] students)
Teacher referral form (n = 187 [240*] students)
SSIS-TRS (n = 238 [240*]

Validity
PSG (n = 187 [240*] students)
Teacher referral form (n = 187 [240*] students)
SSIS-TRS (n = 238 [240*] students)

Internal Consistency

Part-Whole Reliability

Construct Validity

Concurrent Validity

Predictive Validity

Sensitivity, Negative Predictive Power, Specificity, Positive Predictive Power

* asterisk indicates initial projected sample size

can be utilized to determine which items are problematic. For example, Dowdy et al. (2011) applied a DIF analysis to 142 students that were classified as being limited English proficient learners and 110 English proficient learners, and found items within a social-emotional assessment system (i.e., the Behavioral and Emotional Screening System-Teacher and the Behavior Assessment System for Children–Second Edition, Teacher Rating Scale–Child) that were not invariant across the two groups. In addition to applying IRT, DIF can also be detected using the Mantel-Haenszel chi-square test (Holland & Thayer, 1988). Although this method offers less flexibility and precision in the determination of suspected differences between the focal and reference group, the cost of conducting this type of analysis is low, and a significance test is calculated to aid in the interpretation of the final statistic that is reported. Additional limitations associated with the Mantel-Haenszel approach are related to issues of generality. This technique is not designed to detect *nonuniform* item bias and lacks complexity (e.g., comparable to a basic analysis of two item characteristic curves (ICCs) that is based on a one-parameter logistic model). When there is no interaction between the level of social skills functioning and group membership *uniform* DIF can be detected (Swaminathan & Rogers, 1990). *Nonuniform* DIF exists when an interaction is present, which suggests that the difference in the probabilities of a student having strong social skills for the two groups is not the equivalent across levels of ability (e.g., language skills), and is evidenced by nonparallel ICCs. Researchers that have studied the application of the Mantel-Haenszel approach and sought to detect DIF in small sample sizes have reported mixed results (Muniz, Hambleton & Xing, 2001; Parshall & Miller, 1995; Roussos & Stout, 1996).

Logistic regression in DIF is an alternative method. Such a method further allows for the examination of test validity by maximum likelihood (Osterlind & Everson, 2009). Multiple

advantages are noted in utilizing such a method. For example, logistic regression allows for responses to be expressed as a probability and allows for the opportunity for a groups' responses on an item to remain consistent with the measure's score interpretations. Additionally, normality assumptions are relaxed. Therefore multiple real-world contexts can be considered and smaller sample sizes are acceptable. In addition to its status as being more readily applicable, both nonuniform and uniform DIF can be explored when using logistic regression. One limitation is that for "polytomous items, the categorical dependent variable supports either a two-group or three-group solution as a practical limit despite the fact that regression theory allows solutions with more than three categories" (Osterlind & Everson, 2009, p. 60).

Due to the fact that researchers have not attempted to detect DIF in the SSIS-TRS across Spanish-speaking ELLs and native English-speaking students, the uniformity of the results is unknown. Therefore, given a limited sample size, and a lack of knowledge of the amount of DIF that is present in social emotional rating scales, a logistic regression model was applied in the present study. Given the dichotomous nature of DIF analyses, Spanish-speaking ELLs were divided into groups to explore potential item bias across the English language-learning continuum. Additionally, the application of DIF in this study served to further the discussion around the potential for group differences and any decisions must also take into account issues of fairness, as well as test and item bias.

Analyses

Research Question 1: Do the PSG and SSIS-TRS demonstrate appropriate levels of reliability, in Spanish-speaking ELL populations as evidenced through:
       c. Internal consistency coefficients; and
       d. part-whole reliability coefficients?

Hypothesis 1(a): *Supported. The PSG and the SSIS-TRS maintained moderate to high levels of internal consistency in the Spanish-speaking ELL sample.*

When considering both native English speaking students and Spanish-speaking ELLs, the obtained internal consistency of the PSG was .876 ($p < .01$) using Cronbach's alpha. When only examining ELL participants within the sample, the internal consistency of the PSG was .864 ($p < .01$) using Cronbach's alpha. These results suggest that given this sample, the internal consistency of the PSG is adequate and the reliability is high. Results from the examination of the internal consistency of the SSIS-TRS are provided below in Table 9. These findings indicated moderate to high positive correlations across the scales, suggesting that the SSIS-TRS maintains adequate internal consistency. The internal consistency of the scale for Autism was lower than the other scales and subscales both when Spanish-speaking ELL status was (i.e., Cronbach's alpha = .601) and was not (i.e., Cronbach's alpha = .542) considered.

Table 9
Internal Consistency of the SSIS-TRS (Internal consistency for Spanish speaking ELLs are indicated in parentheses)

Scales/Subscales	Cronbach's alpha
Social Skills	.968
	(.967)
Communication	.998
	(.755)
Cooperation	.998
	(.864)
Assertion	.998
	(.772)
Responsibility	.999
	(.835)
Empathy	.827
	(.820)
Engagement	.998
	(.830)
Self−Control	.999
	(.881)
Problem Behaviors	.992
	(.992)
Externalizing	.986
	(.986)
Bullying	.972

	(.973)
Hyperactivity/Inattention	.975
	(.975)
Internalizing	.985
	(.985)
Autism Spectrum	.542
	(.601)
Academic Competence	.950
	(.934)

Hypothesis 1(b): Supported. The PSG and SSIS-TRS maintained moderate to high levels of part-whole reliability.

When considering both native English speaking students and Spanish-speaking ELLs, the obtained part-whole reliability of the PSG was .754. When only examining ELL participants within the sample, the part-whole reliability of the PSG was .738 using Cronbach's alpha. These results further suggest moderate to high levels of part-whole reliability throughout the SSIS-TRS when taking into account English language learning processes. Results from the examination of the part-whole reliability of the SSIS-TRS are provided below in Table 10. These findings indicated moderate to high positive correlations across the scales, suggesting that the SSIS-TRS maintains adequate part-whole reliability both when Spanish-speaking ELL status was and was not considered.

Table 10
Part-Whole Reliability of the SSIS-TRS (Part-whole reliabilities for Spanish speaking ELLs are indicated in parentheses)

Scales/Subscales	Cronbach's Alpha
Social Skills	.753
	(.751)
Communication	.816
	(.742)
Cooperation	.826
	(.778)
Assertion	.816
	(.745)
Responsibility	.826

	(.769)
Empathy	.808
	(.781)
Engagement	.816
	(.760)
Self−Control	.816
	(.775)
Problem Behaviors	.764
	(.764)
Externalizing	.786
	(.764)
Bullying	.833
	(.787)
Hyperactivity/Inattention	.810
	(.833)
Internalizing	.813
	(.810)
Autism Spectrum	.668
	(.813)
Academic Competence	.804
	(.683)
	(.800)

Research Question 2: Do the PSG and SSIS-TRS demonstrate appropriate levels of validity in Spanish-speaking ELL populations, as evidenced through:

    a. statistically significant correlations across the measures (i.e., PSG, teacher referral form and SSIS-TRS);

    b. moderate to high levels of construct validity as demonstrated by no statistically significant correlations with grade, gender, percentage of ELLs in the classroom, ELL status and ELP level;

    c. adequate levels of predictive validity when comparing the PSG to outcomes on the SSIS-TRS, above what is observed with the teacher referral outcomes; and

    d. appropriate levels of sensitivity and specificity?

<u>*Hypothesis 2(a):*</u> *Supported. The PSG correlated in a statistically significant way with the SSIS-TRS and teacher referral outcomes. When ELL status was taken into account, the PSG, teacher referral form, and SSIS-TRS largely maintained statistically significant correlations, with noted exceptions as predicted.*

A two-tailed correlation was computed for the PSG, SSIS-TRS, and teacher referral outcomes to estimate the concurrent validity of the measures. Analyses considered all PSG domain scores, as well as the overall social skills classification as determined by whether or not a child was referred for social skills support (i.e., indicated by receiving a score of 1 in at least one domain). A *p*-value of less than .05 was required for significance. The results of the Pearson correlational analysis presented in Table 11 demonstrate that when both native English speaking students and Spanish-speaking ELLs are taken into account, the PSG, teacher referral form, and SSIS-TRS maintain correlations that are statistically significant across each analyzed domain. Outcomes on the teacher referral form were significantly related to PSG prosocial scores, PSG Motivation scores, PSG Reading skills scores, PSG Math skills scores, overall PSG classification, and the SSIS-TRS social skills subscale. However, the values were somewhat weakly associated, suggesting that students with lower scores on the PSG may not actually be as likely to be referred by teachers as students needing social skills intervention. The students' SSIS-TRS social skills subscale score also significantly correlated with the student's PSG prosocial score, PSG motivation score, PSG Reading skills score, PSG Math skills score, and overall PSG

classification. Yet as similarly noted above, weaker correlations were observed when comparing

the SSIS-TRS social skills subscale score and the PSG Reading skills and Math skills scores.

When only the Spanish-speaking ELL sample was considered, the correlations between

the teacher referral form and multiple PSG domain areas (i.e., Motivation score, Reading skills

score, Math skills score, and overall PSG classification) were statistically significantly. However,

this was not observed with the PSG prosocial behavior score.

Table 11
*Correlations Among PSG, Teacher Referral and the SSIS-TRS (n=187) (Correlations that are
solely drawn from Spanish speaking ELLs [n=78] are indicated in parentheses)*

	Teacher Referral	SSIS-TRS Social Skills Subscale Score
PSG Prosocial Behavior Score	r = -.318; $p < 0.001$* (r = -.207; $p = 0.069$)	r = .646; $p < 0.001$* (r = .756; $p < 0.001$*)
PSG Motivation to Learn Score	r = -.417; $p < 0.001$* (r = -.486; $p < 0.001$*)	r = .543; $p < 0.001$* (r = .725; $p < 0.001$*)
PSG Reading skills Score	r = -.293; $p < 0.001$* (r = -.316; $p = 0.005$*)	r = .360; $p < 0.001$* (r = .593; $p < 0.001$*)
PSG Math skills Score	r = -.234; $p = 0.001$* (r = -.258; $p = 0.022$*)	r = .298; $p < 0.001$* (r = .443; $p < 0.001$*)
PSG Classification	r = .261; $p < 0.001$* (r = .348; $p = 0.002$*)	r = -.306; $p < 0.001$* (r = -.447; $p < 0.001$*)
Teacher Referral	N/A	r = -.328; $p < 0.001$* (r = -.274; $p = 0.015$*)

$p < .05$

Finally, each domain of the PSG was correlated with each subscale on the SSIS-TRS

using two-tailed Pearson correlations. Correlations from this study are included in Table 12;

obtained correlations from a similar study, as well as a study cited in the SSIS Rating Scales

Manual (Gresham & Elliott, 2008), are provided in Tables 13 and 14. When considering both

Spanish-speaking ELLs and native English speakers, 52 out of 120 correlations were statistically

significant at the p < .05 level. When only Spanish-speaking ELLs were considered, fewer

statistically significant correlations were observed between the PSG and scales within the SSIS-

TRS. The SSIS-TRS did not significantly correlate with the PSG across the following areas: (a)

PSG prosocial behavior scale and the SSIS-TRS bullying and internalizing scales; (b) PSG

motivation to learn scale and the SSIS-TRS internalizing scale; (c) PSG Reading skills scale and

the SSIS-TRS problem behaviors, externalizing, bullying, and internalizing scales; and (d) PSG

Math skills scale and the SSIS-TRS problem behaviors, externalizing, bullying,

hyperactivity/inattention, internalizing, and autism spectrum scales. In the ELL sample, the PSG

prosocial behavior and motivation to learn scales better correlated with the SSIS-TRS subscales

than did the academic domains across the PSG.

Table 12
*Correlations across the SSIS-TRS Scales and PSG Domain Areas (n=187) (Correlations that are
solely drawn from Spanish speaking ELLs [n=78] are indicated in parentheses)*

	PSG Prosocial Behavior	PSG Motivation to Learn	PSG Reading skills	PSG Math skills
SSIS-TRS Social Skills	r = .646* *p* < 0.001 (r = .756* *p* < 0.001)	r = .543* *p* < 0.001 (r = .725* *p* < 0.001)	r = .360* *p* < 0.001 (r = .593* *p* < 0.001)	r = .298* *p* < 0.001 (r = .443* *p* < 0.001)
SSIS-TRS Communication	r = .569* *p* < 0.001 (r = .733* *p* < 0.001)	r = .463* *p* < 0.001 (r = .619* *p* < 0.001)	r = .284* *p* < 0.001 (r = .457* *p* < 0.001)	r = .215* *p* = 0.003 (r = .332* *p* = .003)
SSIS-TRS Cooperation	r = .572* *p* < 0.001 (r = . 637* *p* < 0.001)	r = .653* *p* < 0.001 (r = .786* *p* < 0.001)	r = .375* *p* < 0.001 (r = .519* *p* < 0.001)	r = .372* *p* < 0.001 (r = .446* *p* < 0.001)
SSIS-TRS Assertion	r = .319* *p* < 0.001 (r = .341* *p* = 0.002)	r = .240* *p* = 0.001 (r = .271* *p* = .016)	r = .252* *p* < 0.001 (r = .386* *p* < 0.000)	r = .260* *p* < 0.001 (r = .426* *p* < 0.000)
SSIS-TRS Responsibility	r = .580* *p* < 0.001 (r = .670* *p* < 0.001)	r = .573* *p* < 0.001 (r = .708* *p* < 0.001)	r = .353* *p* < 0.001 (r = .487* *p* < 0.001)	r = .310* *p* < 0.001 (r = .334* *p* = 003)
SSIS-TRS Empathy	r = .528* *p* < 0.001 (r = .644* *p* < 0.001)	r = .355* *p* < 0.001 (r = .560* *p* < 0.001)	r = .235* *p* = 0.001 (r = .503* *p* < 0.001)	r = .120 *p* = 0.103 (r = .279* *p* = 0.013)

SSIS-TRS Engagement	r = .553* p < 0.001 (r = .666* p < 0.001)	r = .463* p < 0.001 (r = .524* p < 0.001)	r = .310* p < 0.001 (r = .545* p < 0.001)	r = .209* p = 0.004 (r = .451* p < 0.001)
SSIS-TRS Self Control	r = .680* p < 0.001 (r = .710* p < 0.001)	r = .552*; p < 0.001 (r = .577*; p < 0.001)	r = .487* p < 0.001 (r = .456* p < 0.001)	r = .553* p < 0.001 (r = .659* p < 0.001)
SSIS-TRS Problem Behaviors	r = -.499* p < 0.001 (r = -.350* p = 0.002)	r = -.463* p < 0.001 (r = -.304* p = 0.007)	r = -.288* p < 0.001 (r = -.211 p = .064)	r = -.277* p < 0.001 (r = -.166 p = .146)
SSIS-TRS Externalizing	r = -.462* p < 0.001 (r = -.338* p = 0.002)	r = -.450* p < 0.001 (r = -.317* p = 0.005)	r = -.270* p < 0.001 (r = -.178 p = .118)	r = -.223* p = 0.002 (r = -.040 p = .726)
SSIS-TRS Bullying	r = -.335* p < 0.001 (r = -.162 p = 0.157)	r = -.263* p < 0.001 (r = -.027 p = 0.815)	r = -.168* p = 0.022 (r = -.007 p = 0.954)	r = -.118 p = 0.108 (r = -.172 p = 0.133)
SSIS-TRS Hyperactivity/ Inattention	r = -.300* p < 0.001 (r = -.433* p = 0.002)	r = -.233*; p = 0.001 (r = -.422* p = 0.007)	r = -.150* p = 0.040 (r = -.257* p = .023)	r = -.180* p = 0.013 (r = -.191 p = .094)
SSIS-TRS Internalizing	r = -.300* p < 0.001 (r = -.141 p = 0.217)	r = -.233* p = 0.001 (r = -.085 p = 0.458)	r = -.150* p = 0.040 (r = -.111 p = .335)	r = -.180* p = 0.013 (r = -.194 p = .089)
SSIS-TRS Autism Spectrum	r = -.278* p < 0.001 (r = -.376* p = 0.001)	r = -.151* p = 0.039 (r = -.284* p = 0.012)	r = -.087 p = 0.237 (r = -.296* p = .009)	r = -.032 p = 0.666 (r = -.079 p = .494)

	PSG Prosocial Behavior	PSG Motivation to Learn	PSG Reading Skills	PSG Math Skills
Social Skills	.71*	.78*	.57*	.41
Communication	.70*	.68*	.46*	.26
Cooperation	.62*	.74*	.52*	.32
Assertion	.35	.53*	.68*	.48*
Responsibility	.46*	.53*	.37	.24
Empathy	.47*	.44*	.42	.17
Engagement	-.21	-.18	-.16	-.17
Self Control	.38	.64*	.45*	.31
Problem Behaviors	-.64*	-.64*	-.35*	-.31

Externalizing	-.51*	-.60*	-.32	-.29
Bullying	-.40	-.33	-.04	-.03
Hyperactivity/Inattention	-.52*	-.61*	-.35	-.20
Internalizing	-.55*	-.46*	-.25	-.20
Autism Spectrum	-.69*	-.62*	-.42	-.28

*p < .01

Table 14
*Previous Research: Correlations of the SSIS-TRS and PSG Domain Scores
(n=63; Gresham & Elliott, 2008)*

	PSG Prosocial Behavior	PSG Motivation to Learn	PSG Reading Skills	PSG Math Skills
Social Skills	.69	.57	.42	.33
Communication	.65	.52	.35	.23
Cooperation	.68	.64	.39	.31
Assertion	.33	.24	.44	.35
Responsibility	.64	.61	.38	.33
Empathy	.47	.32	.16	.04
Engagement	.64	.51	.54	.44
Self Control	.52	.41	.20	.21
Problem Behaviors	-.55	-.53	-.53	-.23
Externalizing	-.50	-.52	-.17	-.15
Bullying	-.31	-.37	-.04	-.02
Hyperactivity/Inattention	-.54	-.55	-.28	-.28
Internalizing	-.34	-.30	-.32	-.22
Autism Spectrum	-.70	-.57	-.45	-.37

Hypothesis 2(b): *Supported with exceptions. The PSG, teacher referral form, and SSIS-TRS demonstrated moderate to high levels of construct validity as demonstrated by no statistically significant correlations with grade, gender, or classroom. When ELP levels were taken into account, the PSG, teacher referral form, and SSIS-TRS did not demonstrate moderate to high levels of construct validity.*

Two-tailed correlations were computed for the PSG, SSIS-TRS, and teacher referral outcomes. Correlations with gender, ELP status, grade, and proportion of Spanish-speaking ELLs in each classroom were reviewed in order to estimate the construct validity of the measures. Analyses considered all PSG domains, as well as the overall social skills classification as determined by whether or not a child was referred for social skills support (i.e., indicated by

receiving a score of 1 in at least one domain). A *p*-value of less than .05 was required for statistical significance. The results of the Pearson correlational analysis are presented in Table 15.

PSG outcome scores were not significantly related to gender or grade. As predicted, ELL status and the portion of ELLs in the classroom also did not vary in a statistically significant way across the domain and outcome scores on the PSG. ELL status significantly correlated with the PSG Math skills scale, and ELP level (i.e., the ACCESS for ELLs® score) significantly correlated with the various domain areas and outcome scores on the PSG. Additionally, the percentage of ELLs in the classroom significantly correlated with the PSG prosocial behavior scale. Moreover, when only Spanish-speaking ELLs were considered, an even greater number of statistically significant correlations was observed across the PSG scales and various demographic variables (i.e., ACCESS for ELLs® score and percentage of ELLs in the classroom).

In the given sample, the SSIS-TRS scales also were not significantly related to gender, grade, or the percentage of ELLs in the classroom. When both native English-speaking students and Spanish-speaking ELLs were taken into account, correlations among the student's ACCESS for ELLs® score and the SSIS-TRS social skills, autism spectrum, and academic competence scales were statistically significant. Similarly, ELL status significantly correlated with the SSIS-TRS social skills, autism spectrum, and academic competence scales. When only Spanish-speaking ELLs were considered, the social skills, autism spectrum, and academic competence scales also significantly correlated with the ACCESS for ELLs® score and ELL status. The percentage of ELLs in the classroom was statistically significant and inversely related to the PSG prosocial behavior scale. When only Spanish-speaking ELLs were considered, the SSIS-TRS social skills and autism spectrum scales were also statistically significant and inversely related to percentage of ELLs in the classroom.

Table 15

Current Study: Correlations of the PSG and SSIS-TRS Prosocial Behavior Scale Across Demographic Variables (n=187; Correlations that are solely drawn from Spanish speaking ELLs are indicated in parentheses, n=78)

	Gender	Grade	ACCESS for ELLs® Score	ELL Status	Classroom ELL %
PSG Prosocial Behavior score	r = .008 p = .916 (r = .281 p = 0.063)	r = .031 p = .677 (r = -.238* p = 0.036)	r = .177* p = .015 (r = .320* p = 0.004)	r = .114 p = .122	r = -.180* p = .014 (r = -.329* p = 0.003)
PSG Motivation to Learn score	r = .083 p = .259 (r = .269 p = 0.067)	r = .082 p = .263 (r = -.077 p = 0.500)	r = .191* p = .009 (r = .416* p < 0.001)	r = .108 p = .142	r = -.047 p = .521 (r = -.206 p = 0.070)
PSG Reading skills score	r = .018 p = .806 (r = -.567 p = 0.07)	r = -.047 p = .523 (r = .160 p = 0.162)	r = .178* p = .015 (r = .307* p = 0.006)	r = .118 p = .107	r = -.065 p = .374 (r = -.198 p = 0.082)
PSG Math skills score	r = -.051 p = .486 (r = -.238 p = 0.066)	r = .064 p = .381 (r = .002 p = 0.984)	r = .277* p < .001 (r = .362* p = 0.036)	r = .214* p = .003	r = -.050 p = .496 (r = -.023 p = 0.842)
PSG Classification	r = .009 p = .907 (r = -.049 p = 0.668)	r = -.090 p = .222 (r = .177 p = .120)	r = -.080 p = .276 (r = -.264* p = 0.020)	r = .023 p = .757	r = .070 p = .338 (r = .112 p = 0.331)
Teacher Referral	r = -.065 p = .377 (r = -.147 p = 0.198)	r = -.101 p = .171 (r = -.134 p = 0.244)	r = -.131 p = .074 (r = .051 p = 0.659)	r = .155* p = .034	r = .003 p = .966 (r = -.190 p = 0.095)
SSIS-TRS Social Skills Scale score	r = .117 p = 0.071 (r = -.219 p = 0.055)	r = -.158 p = 0.058 (r = -.303 p = 0.061)	r = .352* p < 0.001 (r = .317* p < 0.001)	r = .293* p < 0.001	r = -.117 p = 0.071 (r = -.260* p = 0.004)
SSIS-TRS Problem Behaviors Scale score	r = -.043 p = 0.557 (r = -.219 p = 0.055)	r = .028 p = 0.717 (r = .191 p = 0.109)	r = -.059 p = 0.422 (r = -.030 p = 0.795)	r =.056 p = 0.447	r =.042 p = 0.567 (r = .158 p = 0.168)
SSIS-TRS Autism Spectrum Scale score	r = .029 p = 0.698 (r = .269 p = 0.053)	r = -.160 p = 0.055 (r = -.303* p = 0.001)	r = -.242* p = 0.001 (r = -.317* p = 0.000)	r = -.193* p = 0.008	r = -.234 p = 0.051 (r = -.260* p = 0.004)

SSIS-TRS	r = -.011	r = -.026	r = .235*	r = .167*	r = -.044
Academic	p = 0.879	p = 0.721	p = 0.001	p = 0.022	p = 0.551
Competence Scale	(r = .105	(r = -.209	(r = .399*		(r = .000
score	p = 0.358)	p = 0.066)	p < 0.001)		p = 0.997)

*$p < .05$

Hypothesis 2(c): Not supported with exceptions. The PSG demonstrated moderate to high levels of predictive validity when examining outcomes on the SSIS-TRS social skills scale, but not above what was observed with teacher referral outcomes. Neither the PSG nor the teacher referral form exhibited high levels of predictive validity when examining outcomes on the SSIS-TRS social skills scale, when ELP levels were taken into account.

Multiple regression analyses were conducted on the SSIS-TRS to determine if the PSG and teacher referral adequately predicted students' social skills functioning levels. The first analysis included PSG classification (i.e., indicated by receiving a score of 1 in at least one domain) as the predictor while the second analysis additionally included teacher referral. The regression equation with the PSG classification as the predictor was significant, $R^2 = .094$, adjusted $R^2 = .089$, $F(1, 186) = 19.09$, $p < .001$. The regression equation with teacher referral as the predictor was also significant, $R^2 = .39$, adjusted $R^2 = .106$, $F(1, 186) = 21.99$, $p < .001$. Based on these results, the PSG and teacher referral both appear to be good predictors of a student's social skills score on the SSIS-TRS. A multiple regression analysis was conducted with both PSG classification and teacher referral as predictors to determine which screening measure best predicted the SSIS-TRS social skills score. The linear combination of the two measures was also significantly related to the SSIS-TRS social skills score, $R^2 = .163$, adjusted $R^2 = .153$, $F(2, 186) = 17.86$, $p < .000$. Teacher referral predicted SSIS-TRS social skills score significantly over and above PSG classification, R^2 change $= .16$, $F(1, 184) = 14.47$, $p < .01$, but PSG classification did not predict SSIS-TRS social skills scores significantly over and above teacher referral, R^2 change $= .022$, $F(1, 184) = 11.407$, $p = .054$.

A hierarchical regression analyses was then conducted to see if the measures predicted SSIS-TRS social skills scores after controlling for ELL student ELP level (see Table 16). The results of the first analysis (i.e., step 1) indicated that ELP levels accounted for a significant amount of the student's SSIS-TRS score, $R^2 = .117$, adjusted $R^2 = .117$, $F(1, 186) = 24.449$, $p < .001$, suggesting that ELLs with lower ELP scores tended to have lower social skills scores according to the SSIS-TRS. After controlling for ELP level, PSG classification accounted for

Table 16
Summary of Hierarchical Regression Analysis Predicting TRS Social Skills Score (n=187)

	Model 1			Model 2			Model 3		
	B	*SE B*	*β*	*B*	*SE B*	*B*	*B*	*SE B*	*β*
ELP level	3.711	.751	.73*	3.468	.721	.319*	3.197	.703	.295*
PSG				-18.25	4.32	-.28			
Teacher referral							-11.86	3.31	-.24
Adj R²		.117			.195			.235	
F for ΔR²		24.45*			17.847*			12.850*	

*$p < .01$

7.8% of the SSIS-TRS score, R^2 change $= .078$, $F(2, 186) = 24.45$, $p < .001$, and teacher referral accounted for 5.3% of the SSIS-TRS score, R^2 change $= .053$, $F(3, 186) = 20.080$, $p < .001$. Due to statiscally sigficant results, the PSG and teacher referral appear to predict social skills on the SSIS-TRS, after controlling for ELP levels, but do not appear to be clinically significant.

Hypothesis 2(d): Supported with exceptions. The sensitivity and negative predictive power (NPP) of the PSG were technically adequate, while the specificity and positive predictive power (PPP) were not. The teacher referral form maintained adequate specificity and higher positive predictive power (PPP) when compared to the PSG.

Table 17 shows classification rates on the PSG and by teacher referral broken down into true and false positives and negatives, using the student's classification by the SSIS-TRS as the gold standard. PSG referrals include students that are in need of remedial instruction or coaching (i.e., a score of a 1 in any of the domain areas).

Sensitivity refers to the ability of a measure to minimize false negatives, or correctly identify individuals that should be considered, and specificity refers to the ability of an instrument to minimize false positives, or correctly eliminate individuals that should not be identified. According to research at the first tier of a multi-gated assessment system, casting a wide net to identify any individual that might be at risk is critical (Glover & Albers, 2007; Levitt et al., 2007). Higher levels of sensitivity are therefore acceptable and considered to be appropriate, even though such levels can be associated with more false positives (Levitt, et al., 2007). Glover and Albers (2007) suggest that sensitivity, specificity, PPP, and NPP levels should not fall below 75%, because false positives create reduced opportunities for learning, lead to inappropriate resource allocation, and increase stress for students, families, and support staff. These measures are predicted to result in outcomes that fall at or above 75% across native English-speaking students and Spanish-speaking ELLs.

Sensitivity was measured by determining the ratio of true positive identifications to true positives and false negatives (a/a+c). PSG sensitivity was calculated to be 89.7% and teacher referral sensitivity was 53.8%. Specificity of the PSG was measured by determining the ratio of true negative identifications to true negative and false positive identifications (d/b+d). Specificity of the PSG was calculated to be 37.2% and teacher referral specificity was 75.7%. When PSG referrals included students in need of remedial instruction and coaching (i.e., a score of a 1 in any of the domain areas) or additional instruction (i.e., a score of a 2 or 3 in any of the domain areas) the sensitivity of the PSG was 90% and the specificity was 29.9%. Positive predictive power (PPP) was measured by determining the ratio of true positive identifications to all positive identifications (a/a+b). It was calculated to be 27.3% for PSG and 36.8% for teacher referral. Finally, negative predictive power (NPP) was measured by determining the ratio of true negative

identifications to all negative identifications (d/c+d). NPP for the PSG was 93.2% and 86.2% for

teacher referral. Base rates across the problem behavior areas (i.e., students identified as having

Table 17

True and False Positive and Negative Classification Rates by the SSIS PSG (n=187)

		SSIS Teacher Rating Scale Results	
		+	-
PSG	+	(a) True Positive 35/187 18.7%	(b) False Positive 93/187 49.7%
	-	(c) False Negative 4/187 2.1%	(d) True Negative 55/187 29.4%
		SSIS Teacher Rating Scale Results	
		+	-
Teacher Referral Form	+	(a) True Positive 21/187 11.2%	(b) False Positive 36/187 19.3%
	-	(c) False Negative 18/187 9.6%	(d) True Negative 112/187 59.9%

concerns in the following areas) were 13.3% for Problem Behaviors overall, 17.2% for

Externalizing problems, 17.2% for Bullying, 15.1% for Hyperactivity/Inattention, 13.4% for

Inattention, and 15.1% for Autism Spectrum.

Research question 3: Does the PSG function similarly across native English-speaking and Spanish-speaking ELL populations, and demonstrate adequate levels of construct validity, as evidenced by scores that are invariant across scales (i.e., Math skills, Reading skills, Motivation to Learn, and Prosocial Behavior) on the PSG?

Hypothesis 3: Partially Supported. The PSG did not function similarly across native English-speaking and Spanish-speaking ELL populations, and demonstrated inadequate levels of construct validity in the Math skills domain area.

Table 18 shows the results from a DIF analysis across the four content areas within the PSG (i.e., Math skills, Reading skills, Motivation to Learn, and Prosocial Behavior). Subscale scores were evaluated based on PSG total scores that included students in need of remedial instruction and coaching (i.e., a score of a 1 in any of the domain areas) and ELL status. In this sample, Spanish-speaking ELLs scored .468 units lower in the PSG Math skills domain. Given that this result is statistically significant, the scores are not invariant and suggest that this item exhibited DIF. The other domain areas (i.e., Reading skills, motivation to learn, and prosocial

Table 18

Summary of Logistic Regression Analysis and DIFAnalysis for PSG Domain Outcomes across ELL and Native English Speakers (n=187)

Domain Areas	**B**	**SE B**	**t-score**	**p-value**
Math Skills	-.468	.135	-3.458	.001*
Reading Skills	-.239	.133	-1.799	.074
Motivation to Learn	-.214	.139	-1.541	.125
Prosocial Behavior	-.210	.130	-1.615	.108

*p < .05

behavior) demonstrated invariance and did not indicate DIF. Results also indicated that Spanish-speaking ELLs scored .239, .214 and .210 points lower then their native English-speaking counterparts across the PSG Reading skills, Motivation to Learn, and Prosocial Behavior domains respectively.

Research question 4: Does the SSIS-TRS function similarly across native English speaking and Spanish-speaking ELLs, and demonstrate adequate levels of construct validity as evidenced by scores that are invariant across items, 11 subscales (i.e., Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, Self-control, Externalizing, Bullying, Hyperactivity/Inattention, and Internalizing), and four major scales (i.e., Social Skills, Problem Behaviors, Autism Spectrum, and Academic Competence)?

Hypothesis 4: Supported. The SSIS-TRS did not consistently function similarly across native English-speaking and Spanish-speaking ELL populations, and demonstrated inadequate levels of construct validity, as evidenced by scores that are not invariant across every item, 11 subscales (i.e., Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, Self-control, Externalizing, Bullying, Hyperactivity/Inattention, and Internalizing Communication,

Cooperation, Assertion, Responsibility, Empathy, Engagement, Self-control, Externalizing, Bullying, Hyperactivity/Inattention, and Internalizing), and four major scales (i.e., Social Skills, Problem Behaviors, Autism Spectrum and Academic Competence).

Table 19 shows the results from a DIF analysis across the Social Skills and Problem Behaviors subscales. Subscale scores were evaluated based on major scale outcome scores and ELL status. The results revealed that the Assertion subscale exhibited DIF in favor of native English-speaking students, with native English-speaking students scoring 1.401 points above Spanish-speaking ELLs. Additionally, the Empathy subscale exhibited DIF in favor of Spanish-speaking ELLs, with Spanish-speaking ELLs scoring .763 points higher than their Native-English speaking counterparts. Other subscales were invariant, which suggested adequate function across this sample. A further examination of the items within each scale is presented in subsequent tables.

Table 19

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Subscale Scores across ELL and Native English Speakers (n=238)

Major Scales/Subscales	B	SE B	t-score	p-value
Social Skills Scale				
Communication Subscale	.150	.197	.764	.445
Cooperation Subscale	.231	.319	.724	.470
Assertion Subscale	-1.401	.432	-3.241	.001*
Responsibility Subscale	.076	.269	.282	.778
Empathy Subscale	.763	.279	2.730	.007*
Engagement	-.290	.304	-.953	.342
Self−Control Subscale	.471	.339	1.389	.166
Problem Behaviors Scale				
Externalizing Subscale	.232	.345	.672	.502
Bullying Subscale	-.189	.195	-.967	.335
Hyperactivity/ Inattention Subscale	.183	.248	.739	.461
Internalizing Subscale	-.110	.329	-.335	.738

$*p < .05$

Table 20 shows the results from a DIF analysis of the Social Skills major scale and

individual items. Item scores were evaluated based on the total Social Skills major scale outcome

score and ELL status. Based on these results, the following items exhibited DIF: item 4 ("Says

'please.'"), item 5 ("Questions rules that may be unfair."), item 8 ("Forgives others."), item 11

("Stands up for herself/himself when treated unfairly."), item 15 ("Says when there is a

problem."), item 20 ("Takes turns in conversations."), item 25 ("Expressed feelings when

wronged."), item 28 ("Is nice to others when they are feeling bad."), item 33 ("Participates in

games or group activities."), item 34 ("Uses appropriate language when upset."), item 35

("Stands up for others who are treated unfairly."), item 39 ("Starts conversations with peers."),

item 43 ("Introduces herself/himself to others.") and item 45 ("Says nice things about

herself/himself without bragging."). For this sample the scores across certain items (i.e., 4, 5, 11,

15, 25, 35, 39, and 43) exhibited DIF against Spanish-speaking ELLs, while other items (i.e., 8,

20, 28, 33, 34, and 45) exhibited DIF in favor of Spanish-speaking ELLs.

Table 20
*Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Social Skills Scale
Score and Item Scores across ELL and Native English Speakers (n=238)*

Social Skills Scale	B	SE B	t-score	p-value
Item 1	-.014	.092	-.147	.883
Item 2	-.003	.067	-.050	.960
Item 3	.061	.083	.733	.464
Item 4	-.193	.083	-2.316	.021*
Item 5	-.501	.124	-4.021	.000*
Item 6	.031	.090	.348	.728
Item 7	.058	.089	.646	.519
Item 8	.239	.075	3.179	.002*
Item 9	.042	.084	.494	.622
Item 10	.075	.066	1.146	.253
Item 11	-.346	.106	-3.263	.001*
Item 12	.046	.071	.649	.517

Item 13	.030	.082	.363	.717
Item 14	.174	.096	1.816	.071
Item 15	-.217	.091	-2.397	.017*
Item 16	.071	.090	.794	.428
Item 17	.090	.076	1.190	.235
Item 18	.135	.072	1.871	.063
Item 19	.099	.067	1.484	.139
Item 20	.148	.073	2.047	.042*
Item 21	.127	.087	1.471	.143
Item 22	.029	.068	.429	.669
Item 23	-.001	.075	-.019	.985
Item 24	-.118	.080	-1.483	.139
Item 25	-.273	.100	-2.733	.007*
Item 26	-.110	.067	-1.641	.102
Item 27	-.029	.092	-.316	.752
Item 28	.198	.064	3.083	.002*
Item 29	-.001	.076	-.011	.991
Item 30	.076	.077	.985	.326
Item 31	.149	.082	1.818	.070
Item 32	.035	.067	.518	.605
Item 33	.152	.070	2.185	.030*
Item 34	.167	.073	2.293	.023*
Item 35	-.288	.091	-3.174	.002*
Item 36	.009	.079	.110	.912
Item 37	.070	.072	.964	.336
Item 38	.100	.067	1.496	.136
Item 39	-.168	.079	-2.130	.034*
Item 40	-.012	.072	-.164	.870
Item 41	.079	.086	.920	.358
Item 42	.020	.071	.278	.782
Item 43	-.412	.093	-4.442	<.001*
Item 44	.008	.068	.113	.910
Item 45	.238	.096	2.472	.014*
Item 46	-.068	.076	-.892	.373

*$p < .05$

Table 21 shows the results from a DIF analysis of the Communication subscale and

individual items. Item scores were evaluated based on the total Communication subscale

outcome score and ELL status. Based on these results, the following items exhibited DIF: item 4

("Says 'please.'") and item 14 ("Speaks in appropriate tone of voice."). For this sample, item 4

exhibited DIF against Spanish-speaking ELLs, while item 14 exhibited DIF in favor of Spanish-

speaking ELLs.

Table 21

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Communication
Subscale Score and Item Scores across ELL and Native English Speakers (n=238)

Communication Subscale	B	SE B	t-score	p-value
Item 4	-.202	.077	-2.611	.010*
Item 10	.043	.062	.695	.488
Item 14	.177	.086	2.062	.040*
Item 20	.107	.068	1.563	.119
Item 24	-.124	.071	-1.738	.083
Item 30	.063	.068	.927	.355
Item 40	-.064	.071	-.903	.368

*p < .05

Table 22 shows the results from a DIF analysis of the Cooperation subscale and

individual items. Item scores were evaluated based on the total Cooperation subscale outcome

score and ELL status. Based on these results none of the items within this subscale exhibited DIF.

Table 22

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Cooperation Subscale
Score and Item Scores across ELL and Native English Speakers (n=238)

Cooperation Subscale	B	SE B	t-score	p-value
Item 2	-.054	.049	-1.106	.270
Item 7	.041	.059	.692	.490
Item 12	-.040	.061	-.646	.519
Item 17	.051	.052	.978	.329
Item 27	-.026	.065	-.400	.690
Item 37	.028	.050	.552	.581

*p < .05

Table 23 shows the results from a DIF analysis of the Assertion subscale and individual

items. Item scores were evaluated based on the total Assertion subscale outcome score and ELL

status. Based on these results, the following items exhibited DIF: item 35 ("Stands up for others

who are treated unfairly.") and item 45 ("Says nice things about herself/himself without

bragging."). For this sample, item 35 exhibited DIF against Spanish-speaking ELLs, whereas

item 45 exhibited DIF in favor of Spanish-speaking ELLs.

Table 23

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Assertion Subscale Score and Item Scores across ELL and Native English Speakers (n=238)

Assertion Subscale	B	SE B	t-score	p-value
Item 1	.124	.080	1.539	.125
Item 5	-.111	.089	-1.243	.215
Item 11	-.030	.075	-.397	.692
Item 15	.004	.064	.060	.952
Item 25	-.011	.070	-.161	.872
Item 35	-.200	.087	-2.289	.023*
Item 45	.224	.108	2.083	.038*

*$p < .05$

Table 24 shows the results from a DIF analysis of the Responsibility subscale and

individual items. Item scores were evaluated based on the total Responsibility subscale outcome

score and ELL status. Based on these results, item 26 ("Takes care when using other people's

things.") exhibited DIF. For this sample, item 26 exhibited DIF in favor of Spanish-speaking

ELLs.

Table 24

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Responsibility Subscale Score and Item Scores across ELL and Native English Speakers (n=238)

Responsibility Subscale	B	SE B	t-score	p-value
Item 6	.045	.067	.671	.503
Item 16	.073	.069	1.060	.290
Item 22	.017	.050	.335	.738
Item 26	.123	.053	-2.327	.021*
Item 32	.029	.054	.536	.593
Item 42	-.041	.067	-.620	.536

*$p < .05$

Table 25 shows the results from a DIF analysis of the Empathy subscale and individual items. Item scores were evaluated based on the total Empathy subscale outcome score and ELL status. Based on these results, none of the items within this subscale exhibited DIF.

Table 25

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Empathy Subscale Score and Item Scores across ELL and Native English Speakers (n=238)

Empathy Subscale	B	SE B	t-score	p-value
Item 3	-.076	.060	-1.264	.207
Item 8	.113	.074	1.542	.124
Item 13	-.093	.055	-1.703	.090
Item 18	.000	.049	.006	.995
Item 28	.074	.042	1.776	.077
Item 38	-.019	.044	-.423	.673

*p < .05

Table 26 shows the results from a DIF analysis of the Engagement subscale and individual items. Item scores were evaluated based on the total Engagement subscale outcome score and ELL status. Based on these results, the following items exhibited DIF: item 33 ("Participates in games or group activities.") and item 43 ("Introduces herself/himself to others."). For this sample, item 43 exhibited DIF against Spanish-speaking ELLs while item 33 exhibited DIF in favor of Spanish-speaking ELLs.

Table 26

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Engagement Subscale Score and Item Scores across ELL and Native English Speakers (n=238)

Engagement Subscale	B	SE B	t-score	p-value
Item 9	.111	.063	1.777	.077
Item 19	.068	.069	.985	.325
Item 23	.057	.059	.977	.329
Item 29	.025	.061	.408	.683
Item 33	.177	.059	2.999	.003*
Item 39	-.083	.061	-1.357	.176
Item 43	-.356	.092	-.213	<.001*

p < .05

Table 27 shows the results from a DIF analysis of the Self-Control subscale and

individual items. Item scores were evaluated based on the total Self-Control subscale outcome

score and ELL status. Based on these results none of the items within this subscale exhibited DIF.

Table 27
*Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Self-Control Subscale
Score and Item Scores across ELL and Native English Speakers (n=238)*

Self-Control Subscale	B	SE B	t-score	p-value
Item 21	.089	.062	1.435	.153
Item 31	.095	.060	1.567	.118
Item 34	.068	.064	1.063	.289
Item 36	-.065	.059	-1.102	.272
Item 41	.018	.068	.258	.797
Item 44	-.083	.058	-1.430	.154
Item 46	-.101	.052	-2.338	.056

p < .05

Table 28 shows the results from a DIF analysis of the Problem Behavior major scale and

individual items. Item scores were evaluated based on the total Problem Behavior major scale

outcome score and ELL status. Based on these results, the following items exhibited DIF: item

54 ("Has stereotyped motor behaviors."), item 55 ("Forces others to act against their will."), item

68 ("Says bad things about self."), item 69 ("Disobeys rules or requests."), and item 71 ("Gets

distracted easily."). For this sample items 54, 55, and 68 exhibited DIF against Spanish-speaking

ELLs, whereas items 69 and 71 exhibited DIF in favor of Spanish-speaking ELLs.

Table 28
*Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Problem Behavior
Scale Score and Item Scores across ELL and Native English Speakers (n=238)*

Problem Behaviors Scale	B	SE B	t-score	p-value
Item 47	.056	.076	.741	.459
Item 48	-.104	.083	-1.246	.214

Item 49	.077	.063	1.219	.224
Item 50	-.017	.071	-.245	.806
Item 51	.016	.078	.207	.836
Item 52	-.025	.050	-.498	.619
Item 53	-.060	.088	-.689	.492
Item 54	-.124	.056	-2.202	.029*
Item 55	-.151	.050	-3.001	.003*
Item 56	.070	.078	.896	.371
Item 57	-.038	.065	-.586	.559
Item 58	-.112	.080	-1.402	.162
Item 59	-.086	.061	-1.415	.158
Item 60	-.035	.055	-.630	.529
Item 61	.021	.058	.370	.712
Item 62	.077	.100	.764	.445
Item 63	.036	.061	.592	.554
Item 64	.006	.073	.084	.933
Item 65	.123	.075	1.643	.102
Item 66	-.011	.062	-.180	.858
Item 67	.082	.069	1.194	.234
Item 68	-.165	.064	-2.582	.010*
Item 69	.294	.077	3.797	<.001*
Item 70	.040	.077	.526	.599
Item 71	.171	.088	1.938	.054*
Item 72	.098	.078	1.257	.210
Item 73	-.014	.071	-.193	.847
Item 74	-.086	.067	-1.284	.200
Item 75	-.089	.070	-1.267	.206
Item 76	-.053	.072	-.734	.464

*$p < .05$

Table 29 shows the results from a DIF analysis of the Externalizing subscale and individual items. Item scores were evaluated based on the total Externalizing subscale outcome score and ELL status. Based on these results, the following items exhibited DIF: item 55 ("Forces others to act against their will."), item 69 ("Disobeys rules or requests."), and item 75 ("Lies or does not tell the truth."). For this sample item 69 exhibited DIF in favor of Spanish-speaking ELLs while items 55 and 75 exhibited DIF against Spanish-speaking ELLs.

Table 29

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Externalizing Subscale Score and Item Scores across ELL and Native English Speakers (n=238)

Externalizing Subscale	B	SE B	t-score	p-value
Item 47	.031	.066	.464	.643
Item 49	.058	.057	1.010	.314
Item 51	-.007	.070	-.106	.916
Item 53	-.064	.090	-.703	.482
Item 55	-.158	.049	-3.213	<.001*
Item 57	-.044	.-65	-.668	.504
Item 61	.008	.054	.151	.880
Item 63	.012	.051	.237	.813
Item 67	.063	.063	1.005	.316
Item 69	.261	.064	4.087	<.001*
Item 73	-.047	.056	-.826	.410
Item 75	-.114	.060	-1.901	.059*

*$p < .05$

Table 30 shows the results from a DIF analysis of the Bullying subscale and individual items. Item scores were evaluated based on the total Bullying subscale outcome score and ELL status. Based on these results, item 49 ("Bullies others.") and item 55 ("Forces others to act against their will.") exhibited DIF. For this sample, items 55 and 49 exhibited DIF against Spanish-speaking ELLs.

Table 30

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Bullying Subscale Score and Item Scores across ELL and Native English Speakers (n=238)

Bullying Subscale	B	SE B	t-score	p-value
Item 49	-.124	.047	2.655	.008*
Item 52	.015	.038	.398	.691
Item 55	-.125	.042	-2.987	.003*
Item 58	-.100	.058	-1.736	.084
Item 61	.086	.048	1.801	.073

*$p < .05$

Table 31 shows the results from a DIF analysis of the Hyperactivity/Inattention subscale and individual items. Item scores were evaluated based on the total Hyperactivity/Inattention subscale outcome score and ELL status. Based on these results none of the items within this subscale exhibited DIF.

Table 31

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS
Hyperactivity/Inattention Subscale Score and Item Scores across ELL and Native English
Speakers (n=238)

Hyperactivity/ Inattention Subscale	B	SE B	t-score	p-value
Item 47	.022	.062	.365	.716
Item 51	-.010	.070	-.145	.885
Item 53	-.101	.071	-1.427	.155
Item 57	-.032	.071	-.449	.654
Item 59	-.088	.062	-1.423	.156
Item 65	.087	.060	1.461	.145
Item 71	.122	.066	1.856	.065

*p < .05

Table 32 shows the results from a DIF analysis of the Internalizing subscale and

individual items. Item scores were evaluated based on the total Internalizing subscale outcome

score and ELL status. Based on these results none of the items within this subscale exhibited DIF.

Table 32

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Internalizing Subscale
Score and Item Scores across ELL and Native English Speakers (n=238)

Internalizing Subscale	B	SE B	t-score	p-value
Item 56	.090	.060	1.485	.139
Item 62	.072	.083	.868	.386
Item 64	.018	.048	.368	.713
Item 68	-.110	.062	-1.782	.076
Item 70	.047	.068	.683	.495
Item 74	-.072	.044	-1.644	.102
Item 76	-.044	.049	-.892	.373

*p < .05

Table 33 shows the results from a DIF analysis of the Autism Spectrum major scale and

individual items. Item scores were evaluated based on the total Autism Spectrum major scale

outcome score and ELL status. Based on these results, the following items exhibited DIF: item

29 ("Invites others to join in activities."), item 39 ("Starts conversations with peers."), item 40

("Uses gestures or body appropriately with others."), item 56 ("Withdraws from others."), item

60 ("Repeats the same things over and over."), item 66 ("Has nonfunctional routines or rituals.")

and item 72 ("Uses odd physical gestures in interactions."). For this sample, items 29, 39, and 40

exhibited DIF against Spanish-speaking ELLs, whereas items 56, 60, 66, and 72 exhibited DIF in

favor of Spanish-speaking ELLs.

Table 33

*Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Autism Spectrum Scale
Score and Item Scores across ELL and Native English Speakers (n=238)*

Autism Spectrum Scale	B	SE B	t-score	p-value
Item 10	-.066	.074	-.895	.372
Item 19	-.126	.089	-1.414	.159
Item 20	-.066	.092	-.717	.474
Item 29	-.192	.088	-2.181	.030*
Item 30	-.073	.084	-.862	.389
Item 38	-.106	.086	-1.234	.218
Item 39	-.240	.076	-3.157	.002*
Item 40	-.239	.092	-2.591	.010*
Item 48	.099	.112	.885	.377
Item 50	.162	.092	1.756	.080
Item 54	.013	.068	.190	.850
Item 56	.214	.091	2.355	.019*
Item 60	.156	.079	1.981	.049*
Item 66	.145	.073	1.971	.050*
Item 72	.318	.098	3.253	<.001*

*$p < .05$

Table 34 shows the results from a DIF analysis of the Academic Competence major scale

and individual items. Item scores were evaluated based on the total Academic Competence major

scale outcome score and ELL status. Based on these results, the following items exhibited DIF:

item 79 ("In mathematics, how does this student compare with other students?") and item 81 ("In

terms of grade-level expectations, this student's skills in mathematics are...?"). Ratings were

based on a 5-point likert-type scale with a "1" indicating a student who functions in the lowest

10th percentile and a "5" indicating the highest 10th percentile. For this sample, items 79 and 81

exhibited DIF against Spanish-speaking ELLs.

Table 34

Summary of Logistic Regression and DIF Analysis for SSIS-TRS Academic Competence Scale Score and Item Scores across ELL and Native English Speakers (n=238)

Academic Competence Scale	B	SE B	t-score	p-value
Item 77	-.074	.046	-1.592	.113
Item 78	.165	.072	2.298	.072
Item 79	-.176	.058	-3.009	.003*
Item 80	.121	.067	1.805	.072
Item 81	-.127	.061	-2.068	.040*
Item 82	.052	.107	.483	.630
Item 83	.039	.054	.718	.474

*p < .05

Research question 5: Does the SSIS-TRS function similarly across students of varying English language proficiency levels as demonstrated by:
  a. *invariance in scores across domain areas (i.e., Math skills, Reading skills, Motivation to Learn, and Prosocial Behavior) on the PSG; and*
  b. *invariance in scores across items, 11 subscales (i.e., Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, Self-Control, Externalizing, Bullying, Hyperactivity/Inattention and Internalizing), and four major scales (i.e., Social Skills, Problem Behaviors, Autism Spectrum and Academic Competence)?*

<u>*Hypothesis 5(a):*</u> *Supported. The PSG did not function similarly acrosts ELP levels and demonstrated inadequate levels of construct validity, as evidenced by scores that were not invariant across domain areas.*

Given the sample size of this population, a DIF analysis was conducted comparing two groups of students at varying ELP levels (i.e., with Group 1 consisting of students at ELP levels 1, 2, and 3 and Group 2 consisting of students at ELP levels 4, 5, and 6). This sample included a total of 78 students, with 59 students in Group 1. Table 35 shows the results from a DIF analysis across the four content areas within the PSG (i.e., Math skills, Reading skills, Motivation to Learn, and Prosocial Behavior). Subscale scores were evaluated based on PSG total scores that included students in need of remedial instruction and coaching (i.e., a score of a 1 in any of the domain areas) and ELP levels. In this sample, students with higher ELP levels (i.e., 4, 5, and 6)

scored .641 points higher in Math skills, .679 points higher in their Motivation to Learn, and .522

points higher in their prosocial behavior than ELL students in Group 1. Given that these results

are statistically significant, the scores are not considered to be invariant and suggest that these

item domains exhibited DIF in favor of students with higher ELP levels.

Table 35

Summary of Logistic Regression Analysis and DIF for PSG Domain Outcomes across ELP levels in Spanish-speaking ELLs (n=78)

Domain Areas	B	SE B	t-score	p-value
Math skills	.641	.235	2.724	.008*
Reading skills	.385	.253	1.525	.131
Motivation to Learn	.679	.243	2.796	.007*
Prosocial Behavior	.522	.245	2.134	.036*

*p < .05

<u>Hypothesis 5(b):</u> Supported. SSIS-TRS did not function similarly across ELP levels as demonstrated by mean level differences across items, 11 subscales (i.e., Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, Self-Control, Externalizing, Bullying, Hyperactivity/Inattention and Internalizing), and four major scales (i.e., Social Skills, Problem Behaviors, Autism Spectrum and Academic Competence).

Given the sample size of this population, a DIF analysis was conducted comparing two

groups of students at varying ELP levels (i.e., with Group 1 consisting of students at ELP levels

1, 2, and 3 and Group 2 consisting of students at ELP levels 4, 5, and 6). This sample included a

total of 118 students with 87 students in Group 1. Table 36 shows the results from a DIF analysis

across the social skills and problem behaviors subscales. Subscale scores were evaluated based

on major scale outcome scores and ELL status. The results reveal that the Assertion,

Externalizing, and Bullying subscales exhibited DIF in favor of students with higher ELP levels.

All other subscales were invariant. A further examination of the items within each scale is
presented in subsequent tables.

Table 36

Summary of Logistic Regression Analysis and DIF for SSIS-TRS Subscale Scores across ELP Levels in Spanish-Speaking ELLs (n=118)

Major Scales/Subscales	B	SE B	t-score	p-value
Social Skills Scale				
Communication Subscale	.258	.326	-.790	.431
Cooperation Subscale	.346	.461	.752	.454
Assertion Subscale	1.412	.681	2.072	.041*
Responsibility Subscale	-.365	.409	-.892	.374
Empathy Subscale	-.530	.488	-1.086	.280
Engagement Subscale	-.044	.448	-.097	.923
Self−Control Subscale	-.562	.487	-1.153	.251
Problem Behaviors Scale				
Externalizing Subscale	1.116	.547	2.041	.044*
Bullying Subscale	.827	.290	2.852	.005*
Hyperactivity/ Inattention Subscale	-.659	.410	-1.610	.110
Internalizing Subscale	-.365	.543	-.672	.503

*$p < .05$

Table 37 shows the results from a DIF analysis of the Social Skills major scale and
individual items. Item scores were evaluated based on the total Social Skills major scale outcome
score and ELP levels. Based on these results, the following items exhibited DIF: item 4 ("Says
'please.'"), item 5 ("Questions rules that may be unfair."), item 7 ("Completes tasks without
bothering others."), item 28 ("Is nice to others when they are feeling bad."), item 36 ("Resolves
disagreements with you calmly."), and item 40 ("Uses gestures or body appropriately with
others."). For this sample, certain items (i.e., 4, 5, and 7) exhibited DIF in favor of students with
higher ELP levels while other items (i.e., 28, 36, and 40) exhibited DIF against students with
higher ELP levels.

Table 37

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Social Skills Scale Score and Item Scores across ELP Levels in Spanish-Speaking ELLs (n=118)

Social Skills Scale	B	SE B	t-score	p-value
Item 1	.209	.144	1.454	.149
Item 2	-.013	.099	-.128	.898
Item 3	-.124	.141	-.883	.379
Item 4	.265	.135	1.959	.053*
Item 5	.618	.192	3.214	.002*
Item 6	-.115	.127	-.905	.367
Item 7	.295	.139	2.126	.036*
Item 8	.186	.123	1.506	.135
Item 9	.190	.120	1.587	.115
Item 10	-.050	.103	-.487	.627
Item 11	.339	.183	1.853	.066
Item 12	.131	.121	1.090	.278
Item 13	-.063	.143	-.443	.659
Item 14	-.113	.145	-.778	.438
Item 15	.193	.143	1.343	.182
Item 16	-.060	.138	-.433	.666
Item 17	.045	.111	.407	.685
Item 18	-.175	.129	-1.352	.179
Item 19	-.177	.105	-1.680	.096
Item 20	.039	.108	.365	.716
Item 21	.071	.136	.524	.601
Item 22	-.009	.108	-.087	.931
Item 23	-.017	.120	-.139	.890
Item 24	-.128	.131	-.979	.329
Item 25	.128	.158	.810	.420
Item 26	-.151	.106	-1.424	.157
Item 27	-.023	.134	-.174	.862
Item 28	-.306	.102	-2.998	.003*
Item 29	-.044	.121	-.365	.716
Item 30	-.016	.118	-.134	.984
Item 31	.004	.120	.030	.976
Item 32	-.149	.101	-1.470	.144
Item 33	-.008	.109	-.072	.942
Item 34	-.181	.113	-1.610	.110
Item 35	.073	.147	.496	.621
Item 36	-.237	.109	-2.182	.031*
Item 37	-.089	.104	-.858	.393
Item 38	-.047	.124	-.379	.706
Item 39	-.071	.121	-586	.559

Item 40	-.254	.103	-2.478	.015*
Item 41	-.022	.118	-.183	.855
Item 42	.118	.106	1.116	.267
Item 43	.082	.149	.553	.581
Item 44	.020	.101	.200	.842
Item 45	-.148	.161	-.918	.360
Item 46	-.217	.115	-1.887	.062

*p < .05

Table 38 shows the results from a DIF analysis of the Communication subscale and

individual items. Item scores were evaluated based on the total Communication subscale

outcome score and ELP levels. Based on these results, item 4 ("Says 'please.'") exhibited DIF,

with the item exhibiting DIF in favor of students with higher ELP levels.

Table 38

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Communication
Subscale Score and Item Scores across ELP Levels in Spanish-Speaking ELLs (n=118)

Communication Subscale	B	SE B	t-score	p-value
Item 4	.249	.119	2.088	p<.001*
Item 10	-.013	.091	-.149	.882
Item 14	-.099	.131	-.755	.452
Item 20	.110	.104	1.062	.291
Item 24	-.120	.111	-1.075	.285
Item 30	.027	.109	.251	.802
Item 40	-.154	.109	-1.413	.160

*p < .05

Table 39 shows the results from a DIF analysis of the Cooperation subscale and

individual items. Item scores were evaluated based on the total Cooperation subscale outcome

score and ELP levels. Based on these results, item 7 ("Completes tasks without bothering

others.'") exhibited DIF in favor of students with higher ELP levels.

Table 39

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Cooperation Subscale
Score and Item Scores across ELP Levels in Spanish-Speaking ELLs (n=118)

Cooperation Subscale	B	SE B	t-score	p-value

Item 2	-.035	.084	-.419	.676
Item 7	.185	.092	2.009	.047*
Item 12	.097	.102	.956	.341
Item 17	-.005	.085	-.061	.952
Item 27	-.124	.094	-1.319	.190
Item 37	-.118	.084	-1.408	.162

*p < .05

Table 40 shows the results from a DIF analysis of the Assertion subscale and individual

items. Item scores were evaluated based on the total Assertion subscale outcome score and ELL

status. Based on these results none of the items exhibited DIF.

Table 40
Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Assertion Subscale Score and Item Scores across ELP Levels in Spanish-Speaking ELLs (n=118)

Assertion Subscale	B	SE B	t-score	p-value
Item 1	.024	.114	.215	.830
Item 5	.171	.144	1.192	.236
Item 11	-.042	.134	-.310	.757
Item 15	-.010	.109	-.090	.928
Item 25	-.095	.120	-.795	.428
Item 35	.034	.143	.235	.814
Item 45	-.083	.179	-.462	.645

*p < .05

Table 41 shows the results from a DIF analysis of the Responsibility subscale and

individual items. Item scores were evaluated based on the total Responsibility subscale outcome

score and ELP levels. Based on these results, item 42 ("Takes responsibility for part of a group

activity.") exhibited DIF in favor of students with higher ELP levels.

Table 41
Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Responsibility Subscale Score and Item Scores across ELP Levels in Spanish-Speaking ELLs (n=118)

Responsibility Subscale	B	SE B	t-score	p-value
Item 6	-.035	.101	-.348	.729
Item 16	-.037	.098	-.373	.710
Item 22	.055	.084	.658	.512

Item 26	-.077	.089	-.864	.389
Item 32	-.103	.080	-1.297	.197
Item 42	.197	.088	2.224	.028*

p < .05

Table 42 shows the results from a DIF analysis of the Empathy subscale and individual

items. Item scores were evaluated based on the total Empathy subscale outcome score and ELP

levels. Based on these results, item 8 ("Forgives others.") and item 28 ("Is nice to others when

they are feeling bad.") exhibited DIF, with item 8 exhibiting DIF in favor of students with higher

ELP levels, whereas item 28 exhibited DIF against students with higher ELP levels.

Table 42

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Empathy Subscale
Score and Item Scores across ELP Levels in Spanish-Speaking ELLs (n=118)

Empathy Subscale	B	SE B	t-score	p-value
Item 3	-.047	.087	-.536	.593
Item 8	.292	.124	2.344	.021*
Item 13	.021	.095	.219	.827
Item 18	-.073	.092	-.791	.430
Item 28	-.205	.066	-3.092	.002*
Item 38	.012	.081	.148	.882

p < .05

Table 43 shows the results from a DIF analysis of the Engagement subscale and

individual items. Item scores were evaluated based on the total Engagement subscale outcome

score and ELP levels. Based on these results, item 9 ("Makes friends easily.") exhibited DIF in

favor of students with higher ELP levels.

Table 43

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Engagement Subscale
Score and Item Scores across ELP Levels in Spanish-Speaking ELLs (n=118)

Engagement Subscale	B	SE B	t-score	p-value
Item 9	.180	.086	2.093	.039*
Item 19	-.067	.115	-.578	.565
Item 23	-.035	.095	-.365	.716
Item 29	.006	.106	.057	.955
Item 33	.021	.095	.224	.823

	-.132	.097	-1.353	.179
Item 39	-.132	.097	-1.353	.179
Item 43	.026	.129	.201	.841

p < .05

Table 44 shows the results from a DIF analysis of the Self-Control subscale and

individual items. Item scores were evaluated based on the total Self-Control subscale outcome

score and ELP levels. Based on these results none of the items exhibited DIF.

Table 44

*Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Self-Control Subscale
Score and Item Scores across ELP Levels in Spanish-Speaking ELLs (n=118)*

Self-Control Subscale	B	SE B	t-score	p-value
Item 21	.101	.106	.954	.342
Item 31	.039	.086	.456	.649
Item 34	-.061	.106	-.573	.568
Item 36	-.107	.085	-1.262	.210
Item 41	.047	.079	.600	.550
Item 44	.118	.091	1.307	.194
Item 46	-.138	.082	-1.677	.096

p < .05

Table 45 shows the results from a DIF analysis of the Problem Behavior major scale and

individual items. Item scores were evaluated based on the total Problem Behavior major scale

outcome score and ELP levels. Based on these results, the following items exhibited DIF: item

48 ("Is preoccupied with object parts."), item 52 ("Does things to make others feel scared."),

item 57 ("Has temper tantrums."), item 58 ("Keeps others out of social circles."), item 67

("Fights with others."), item 71 ("Gets distracted easily."), and item 73 ("Talks back to adults.").

For this sample, items 57, 67, and 73 exhibited DIF in favor of students with higher ELP levels,

whereas items 48, 52, 58, and 71 exhibited DIF against students with higher ELP levels.

Table 45

*Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Problem Behavior
Scale Score and Item Scores across ELP Levels in Spanish-Speaking ELLs (n=118)*

Problem Behaviors Scale	B	SE B	t-score	p-value

Item 47	-.105	.130	-.809	.420
Item 48	-.309	.129	-2.399	.018*
Item 49	-.051	.109	-.466	.642
Item 50	-.168	.115	-1.465	.146
Item 51	-.032	.128	-.253	.801
Item 52	-.281	.078	3.618	p<.001*
Item 53	-.178	.151	-1.182	.240
Item 54	-.101	.068	-1.483	.141
Item 55	.081	.071	1.150	.253
Item 56	-.210	.128	-1.636	.104
Item 57	.230	.101	2.272	.025*
Item 58	-.381	.121	3.153	.002*
Item 59	-.059	.090	-.657	.512
Item 60	.108	.088	1.223	.224
Item 61	.135	.099	1.364	.175
Item 62	.104	.170	.616	.539
Item 63	.182	.100	1.824	.071
Item 64	-.151	.120	-1.258	.211
Item 65	-.139	.126	-1.105	.272
Item 66	-.156	.102	-1.539	.126
Item 67	.373	.116	3.223	.002*
Item 68	.049	.104	.470	.639
Item 69	.016	.152	.106	.916
Item 70	-.201	.130	-1.546	.125
Item 71	-.376	.143	-2.632	.010*
Item 72	-.212	.147	-1.428	.153
Item 73	.249	.115	2.164	.033*
Item 74	.138	.110	1.253	.213
Item 75	.216	.109	1.989	.049
Item 76	-.095	.119	-.803	.424

*$p < .05$

Table 46 shows the results from a DIF analysis of the Externalizing subscale and

individual items. Item scores were evaluated based on the total Externalizing subscale outcome

score and ELP levels. Based on these results, item 47 ("Acts without thinking.") and item 67

("Fights with others.") exhibited DIF. For this sample, item 67 exhibited DIF in favor of students

with higher ELP levels and item 47 exhibited DIF against students with higher ELP levels.

Table 46

*Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Externalizing Subscale
Score and Item Scores across ELP Levels in Spanish-Speaking ELLs (n=118)*

Externalizing Subscale	B	SE B	t-score	p-value
Item 47	-.223	.108	-2.062	.041*
Item 49	-.124	.100	-1.246	.215
Item 51	-.127	.113	-1.131	.261
Item 53	-.284	.159	-1.790	.076
Item 55	.070	.071	.991	.324
Item 57	.142	.101	1.405	.163
Item 61	.039	.097	.399	.691
Item 63	.088	.087	1.013	.313
Item 67	.270	.106	2.545	.012*
Item 69	-.106	.121	-.876	.383
Item 73	.140	.094	1.491	.139
Item 75	.115	.099	1.158	.249

*$p < .05$

Table 47 shows the results from a DIF analysis of the Bullying subscale and individual

items. Item scores were evaluated based on the total Bullying subscale outcome score and ELP

levels. Based on these results, the following items exhibited DIF: item 49 ("Bullies others."), and

item 58 ("Keeps others out of social circles."). For this sample, item 49 exhibited DIF in favor of

students with higher ELP levels while item 58 exhibited DIF against students with higher ELP

levels.

Table 47

*Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Bullying Subscale
Score and Item Scores across ELP Levels in Spanish-Speaking ELLs (n=118)*

Bullying Subscale	B	SE B	t-score	p-value
Item 49	.254	.090	-2.819	.006*
Item 52	.105	.060	1.744	.084
Item 55	.028	.067	.418	.677
Item 58	-.241	.098	2.445	.016*
Item 61	-.120	.084	-1.429	.156

*$p < .05$

Table 48 shows the results from a DIF analysis of the Hyperactivity/Inattention subscale

and individual items. Item scores were evaluated based on the total Hyperactivity/Inattention

subscale outcome score and ELP levels. Based on these results, the following items exhibited

DIF: item 57 ("Has temper tantrums.") and item 71 ("Gets distracted easily."). For this sample,

item 57 exhibited DIF in favor of students with higher ELP levels while item 71 exhibited DIF

against students with higher ELP levels.

Table 48

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Hyperactivity/Inattention Subscale Score and Item Scores across ELP Levels in Spanish-Speaking ELLs (n=118)

Hyperactivity/ Inattention Subscale	B	SE B	t-score	p-value
Item 47	.005	.104	.052	.959
Item 51	.057	.109	.524	.601
Item 53	-.063	.126	-.496	.621
Item 57	.295	.114	2.593	.011*
Item 59	-.001	.096	-.010	.992
Item 65	-.042	.102	-.411	.682
Item 71	-.252	.107	-2.348	.021*

*p < .05

Table 49 shows the results from a DIF analysis of the Internalizing subscale and

individual items. Item scores were evaluated based on the total Internalizing subscale outcome

score and ELP levels. Based on these results, item 74 ("Acts sad or depressed.") exhibited DIF.

For this sample, item 74 exhibited DIF in favor of students with higher ELP levels.

Table 49

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Internalizing Subscale Score and Item Scores across ELP Levels in Spanish-Speaking ELLs (n=118)

Internalizing Subscale	B	SE B	t-score	p-value
Item 56	-.148	.099	-1.500	.136
Item 62	.157	.146	1.078	.283
Item 64	-.092	.082	-1.124	.263
Item 68	.098	.092	1.058	.292
Item 70	-.166	.122	-1.355	.178
Item 74	.194	.068	2.835	.005*
Item 76	-.043	.084	-.504	.615

*p < .05

Table 50 shows the results from a DIF analysis of the Autism Spectrum major scale and individual items. Item scores were evaluated based on the total Autism Spectrum major scale outcome score and ELP levels. Based on these results, item 20 ("Takes turns in conversations.") and item 48 ("Is preoccupied with object parts.") exhibited DIF. For this sample, item 20 exhibited DIF in favor of students with higher ELP levels, whereas item 48 exhibited DIF against students with higher ELP levels.

Table 50

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Autism Spectrum Score and Item Scores across ELP Levels in Spanish-Speaking ELLs (n=118)

Autism Spectrum Scale	B	SE B	t-score	p-value
Item 10	.198	.115	1.723	.088
Item 19	.207	.157	1.320	.190
Item 20	.361	.142	2.547	.012*
Item 29	.242	.143	1.685	.095
Item 30	.145	.125	1.160	.248
Item 38	.227	.151	1.501	.136
Item 39	.076	.117	.649	.518
Item 40	.042	.147	.287	.775
Item 48	-.380	.165	-2.298	.023*
Item 50	-.221	.144	-1.538	.127
Item 54	-.140	.077	-1.826	.070
Item 56	-.282	.145	-1.951	.054
Item 60	.051	.136	.374	.709
Item 66	-.219	.122	-1.790	.076
Item 72	-.305	.183	-1.663	.099

*p < .05

Table 51 shows the results from a DIF analysis of the Academic Competence major scale and individual items. Item scores were evaluated based on the total Academic Competence major scale outcome score and ELP levels. Based on these results, none of the items exhibited DIF.

Table 51

Summary of Logistic Regression Analysis and DIF Analysis for SSIS-TRS Academic Competence Score and Item Scores across ELP Levels in Spanish-Speaking ELLs (n=118)

Academic Competence Scale	B	SE B	t-score	p-value

Item 77	-.116	.088	-1.321	.189
Item 78	-.152	.125	-1.219	.225
Item 79	.124	.115	1.077	.284
Item 80	-.152	.118	-1.290	.200
Item 81	.027	.121	.225	.822
Item 82	.271	.204	1.333	.185
Item 83	-.003	.105	-.024	.981

*p < .05

CHAPTER V

Discussion

The central goal of this study was to establish evidence for the reliability and validity of

the SSIS as both a screening (i.e., PSG) and assessment (i.e., SSIS-TRS) tool for social skills and

behavioral functioning within Spanish-speaking ELLs. Throughout this study multiple pieces of

evidence support the hypothesis that the PSG and SSIS-TRS demonstrates adequate reliability,

but questionable validity in Spanish-speaking ELL populations across the English language-

learning continuum. This investigation sought to achieve multiple goals. First, to further examine

the psychometric properties of the SSIS-PSG in a Spanish-speaking ELL population following

an initial review of the measure in 2011. Second, this study served as a preliminary review of the

SSIS-TRS in its entirety in a Spanish-speaking ELL population. Last, results were sought to

further inform the extent to which ELP levels impact the interpretation of a standardized measure

of social skills and behavioral functioning (i.e., SSIS) in the classroom setting (i.e., as measured

through teacher report). Findings across the research questions are discussed below. Limitations

and implications of the research, as well as future directions for research regarding social skills

screening and assessment across ELL populations, are also reviewed.

Primary Findings

Reliability

Research question 1: *Do the PSG and SSIS-TRS demonstrate appropriate levels of
reliability, in Spanish-speaking ELL populations as evidenced through internal consistency
coefficients and part-whole reliability coefficients?*

As previously demonstrated by Felt (2011), the reliability of the PSG, as estimated in this

study, is excellent. Additionally, this study suggested adequate reliability in the SSIS-TRS. This

is evidenced by high Cronbach's alphas both between scale items and between items and

outcomes across each measure. These results suggest that the items on the PSG and SSIS-TRS

are indicative of similar factors and that all items relate to final screening and assessment

outcomes, when used with Spanish-speaking ELLs. As recommended by Glover & Albers

(2007) when considering requirements for appropriate screeners, this evidence of high reliability

provides support for one area of technical adequacy of the measures.

Validity

Research question 2: *Do the PSG and SSIS-TRS demonstrate appropriate levels of validity in Spanish-speaking ELL populations, as evidenced through: (a) positive correlations across the measures (i.e., PSG, teacher referral form and SSIS-TRS); (b) moderate to high levels of construct validity as demonstrated by no statistically significant correlations with grade, gender, percentage of ELLs in the classroom, ELL status, and ELP level; (c) adequate levels of predictive validity when comparing the PSG to outcomes on the SSIS-TRS, above what is observed with the teacher referral; and (d) appropriate levels of sensitivity and specificity?*

As a critical component in the development and review of universal screening measures,

the technical adequacy of the PSG and SSIS-TRS was explored through validity testing (Glover

& Albers, 2007). Previously demonstrated by Felt (2011) and Gresham and Elliot (2008), the

SSIS-TRS and PSG similarly identified students with suspected social skills deficits. This study

provided further evidence for strong to moderate correlations among relationships between the

PSG Prosocial Behavior scale, the PSG Motivation to Learn Scale, and the SSIS-TRS Social

Skills scale score. Similar findings were noted when the measures were used in a predominantly

native Spanish-speaking sample and when ELL status was taken into account. The present study

also replicated findings that indicated that the PSG Math skills and Reading skills scales do not

correlate as strongly with the SSIS-TRS Social Skills scale score. Given that the effects of social

skills functioning and achievement are bi-directional, and that research has recently indicated

that effects of students' achievement on their later social skills is stronger than the effects of

social skills on achievement, it may still be important to consider academic variables in social-

emotional and behavioral screening measures for all students (Caemmerer & Keith, 2014). Lastly,

although statistically significant, the teacher referral form weakly correlated with the PSG and

SSIS-TRS. Similarly, Felt (2011) whose study was limited by a small sample size and the need to control alpha across multiple correlations, found that teacher referral was not significantly related to whether a student was classified as in need of social skills intervention by the PSG. These results suggest that at the individual domain levels, the PSG identifies the same students as teacher referral, but does not do so at an overall scale level.

In the present study, when only Spanish-speaking ELLs were considered, teacher referral outcomes did not significantly correlate with the PSG Prosocial Behavior domain, suggesting that the PSG may identify Spanish-speaking ELLs differently from teachers themselves. Overall the PSG and SSIS-TRS correlated with each other more strongly, suggesting the PSG is a valid screener for social skills outcomes on the SSIS-TRS for Spanish-speaking ELLs. However, given that the limitations of this study (as described below) and that the correlations trend in the expected direction, it is possible that the PSG may be more highly correlated with the SSIS-TRS than a teacher referral form.

The authors of the SSIS (Gresham & Elliott, 2008) correlated scores across the four PSG domains to the 14 subscales on the TRS using a mainly Caucasian and native English-speaking sample. They found moderate to strong correlations across the PSG domains and weaker relationships when comparing the academic domains on the PSG to the social skills subscales on the SSIS-TRS. Felt (2011) additionally examined correlations across the two measures in a Spanish-speaking ELL sample and found similar results. In the present study, results further indicated moderate to strong, positive correlations with the Social Skills scale on the SSIS-TRS and both the Prosocial Behavior and Motivation to Learn scales on the PSG. Negative correlations were found when the PSG was compared to the SSIS-TRS problem behavior scales. Weaker correlations were observed when comparing the SSIS-TRS to academic domains across

the PSG. Comparable results were found for Spanish-speaking ELLs, suggesting that the PSG and the SSIS-TRS measure similar concepts and can be useful when used in conjunction with one another, even when ELL status is taken into account.

This study also examined the construct validity of the PSG and SSIS-TRS. Although generally determined to be adequate, the use of the measures in Spanish-speaking ELL student populations may be questionable. Similar to Felt (2011), the present study found that the PSG domains and outcome classification and SSIS-TRS scales did not significantly correlate with gender or grade, even when only Spanish-speaking ELLs were considered. However, it is notable that when only Spanish-speaking ELLs were considered, grade level was statistically significant and inversely related to the PSG Prosocial Behavior scale. This could have been accounted for by the fact that the majority of students in the study were in first through third grade, which impacted the results.

When ELP levels were considered, multiple domains across the PSG and SSIS-TRS significantly correlated with the ACCESS for ELLs® scores and ELL status. The measures may therefore reveal social skills difficulties, lower motivation to learn, and additional areas of concern across Spanish-speaking ELL populations. Additionally, higher ELP levels were associated with lower ratings on the SSIS-TRS problem behavior scale. This suggests that native Spanish-speaking students with lower English proficiency levels may be rated as having more behavioral problems. An inverse relationship was noted when the SSIS-TRS Autism Spectrum scale was compared to ELL status as well as ELP levels. Symptoms associated with autism spectrum disorders may therefore be observed at higher ELP levels in Spanish-speaking ELL populations when using the SSIS-TRS.

Overall, the proportion of Spanish-speaking ELLs in the classroom did not significantly correlate with the PSG and the SSIS-TRS, as was also found by Felt (2011). However, on the PSG, classrooms with higher percentages of ELLs had significantly lower scores on the PSG Prosocial Behavior scale. Additionally, when Spanish-speaking ELLs were considered, classrooms with more Spanish-speaking ELLs had significantly lower scores on the SSIS-TRS social skills and higher scores on the Autism Spectrum scale. Therefore, when the PSG and SSIS-TRS are used in classrooms with more ELLs, the measures may not be consistently appropriate in identifying social skills functioning and autism spectrum behaviors for Spanish-speaking ELLs. Overall these findings could suggest possible bias within the PSG and SSIS-TRS in over-identifying Spanish-speaking ELLs with lower ELP levels. Alternatively, they may also further reflect previous research findings that students with lower ELP levels are more likely to experience social skills difficulties (Spomer & Cowan, 2001). Since the PSG and SSIS-TRS have been established as valid and reliable measures, it is possible that these correlations are indicative of the social skills difficulties typically experienced by Spanish-speaking ELLs (Gresham & Elliott, 2008), at least when interacting with native English-speaking students.

Based on the extant literature on universal screening tools, the PSG is expected to offer an advantage over teacher referral in predicting social skills deficits. As screeners should be time-efficient and cost-effective, the current research sought to determine whether the PSG was more useful in predicting a student's social skills than a teacher referral process, which is a relatively quick and simple measure (Glover & Albers, 2007). Therefore, the predictive validity of the PSG was evaluated. Similar to Felt (2011), the results of this study also suggested that the PSG does not provide additional information over a teacher referral form, as teacher referral accounted for 16% of the variance above the PSG, while the PSG accounted for only 2.2% of the

variance above teacher referral in predicting outcomes on the SSIS-TRS Social Skills scale. However in interpreting these scores it is important to note that both forms are based on teacher ratings but that the PSG provides a structured rating scale to collect teacher input, which can provide teachers with an organized and ready-made process that can be used in schools. It is unsurprising that teacher referral was related to the SSIS-TRS Social Skills scale score, as teacher judgment is critical in predicting outcomes (Elliott et al., 2007). However, it was expected that the PSG would be a significantly better measure of social skills functioning on the SSIS-TRS, as prior research has found structured measures to outperform unstructured judgment (Elliott et al., 2007). Although it is possible that the PSG may not offer a useful service to schools, and rather support a basic teacher referral procedure, caution must be taken. In interpreting these findings as a limited number of students were included in the screening sample and this diminished the number of individuals that teachers could ultimately consider overall. A full screening of an entire classroom is therefore warranted before more conclusive claims can be made about the usefulness of the PSG over and above a teacher referral process.

Additionally, consistent with results presented by Felt (2011), this study suggested that after controlling for a student's ELP level, which was determined to relate to a student's social skills ratings in the previous analyses, neither teacher referral nor the PSG outcome score served as useful predictors of a student's social skills score on the SSIS-TRS. This suggests that it is important to first consider a student's language proficiency level before determining how students will be identified as needing social skills intervention.

Lastly, this research considered the sensitivity and specificity of the PSG in identifying students in need of social skills interventions. The PSG demonstrated very strong sensitivity

(89.7%) but rather weak specificity (37.2%). As suggested by Levitt et al. (2007) the determination of decision points for appropriate sensitivity and specificity levels should

> be influenced by the…desired prevention goal. If the goal is universal prevention…the program intends to cast a wide net…In this case, an instrument with higher sensitivity and [NPP] would be appropriate…although higher sensitivity could be associated with more false positives, this allows the early identification program to be confident that very few children were missed (p. 179).

Glover and Albers (2007) further state that sensitivity is the most important metric in the first gate of a multi-gated assessment system. Therefore, as also suggested by Felt (2011), the PSG would appear to be an appropriate screener within a multi-gated assessment given its specificity and NPP.

Teacher referral, on the other hand, demonstrated weak sensitivity but appropriate specificity. These results suggest that the PSG is a more sensitive measure than teacher referral, but teacher referral maintains more specificity, when considering social skills difficulties. Glover and Albers (2007) recommended that when the specificity and NPP of the screener are too low, the over-identification of students could result in a potential drain on school resources. The teacher referral form did not meet the requirements for sensitivity (i.e., at or above 75%), but did have adequate specificity and NPP. Teacher referral also had higher positive predictive power, while the PSG had higher negative predictive power. Thus, it appears that the PSG could be a better screening measure, as it is more likely to identify all students in need of social skills intervention.

Therefore neither measure would appear to be particularly excellent. However, when considering a multi-tiered system with multiple assessment gates, the PSG may be a better initial

screening measure as it would be less likely to miss students in need of social skills intervention. If schools use such a measure, they must be ready to deploy additional resources to further assess students identified by the PSG in order to prevent problems associated with over-referral. Furthermore, given that the PSG and SSIS-TRS do not clearly demonstrate adequate technical adequacy when utilized for Spanish-speaking ELL students, additional considerations must be made. Therefore, making use of a combination of measures (e.g., teacher referral), a modified version of the PSG, or an additional second tier assessment measure would be useful in order to ensure appropriate sensitivity and specificity levels and that the correct students are provided with the needed supports around their social skills functioning.

Research question 3: Does the PSG function similarly across native English-speaking and Spanish-speaking ELL populations, and demonstrate adequate levels of construct validity, as evidenced by scores that are invariant across scales (i.e., Math skills, Reading skills, Motivation to Learn, and Prosocial Behavior) on the PSG?

The construct validity of the PSG was further examined through a logistic regression DIF analysis. Few items with DIF were identified in the PSG; results indicated that only the Math skills scale demonstrated a lack of invariance, suggesting that the same construct was not being measured across native English-speaking and Spanish-speaking English ELLs. However, the overall analysis indicated that Spanish-speaking ELLs consistently received lower scores across each domain on the PSG when compared to native English-speaking students. Therefore, as similarly noted in previous research on multi-gated assessment systems, the PSG classifies Spanish-speaking ELLs as displaying academic skills and social skills (i.e., Prosocial Behavior and Motivation to Learn) that are weaker than their native English-speaking counterparts (Dowdy et al., 2011; Felt, 2011). As previously noted, this may be due to the technical inadequacy of the measure when used in Spanish-speaking ELLs, or may suggest that Spanish-speaking ELLs have more social-emotional and behavioral difficulties that interfere both with

social skills and academic performance. Previous research has suggested that ELLs have appeared to have attention problems, difficulties following directions, and interpersonal competence (Artiles & Ortiz, 2002; Bialystok, 2001; Edl et al., 2008). Research has further indicated that dual language learners exhibit specific behavioral patterns that differ from monolingual students (e.g., attentiveness to changes in environment, flexibility in switching between activities; Rhodes et al., 2005; Solano-Flores, 2011). ELLs may also appear to progress more slowly through academic material due to dual language learning processes and may demonstrate behavioral patterns that are culturally influenced, thereby making students of similar backgrounds more comfortable with spending time with other such similar students (Rhodes et al., 2005; Solano-Flores, 2011; Solano-Flores & Nelson-Barber, 2001). Therefore, caution should be exercised when utilizing standardized measures that have been developed in native English-speaking populations, as skills and behavioral patterns may be identified as maladaptive but may otherwise serve a different purpose in facilitating learning and achievement for ELL populations. Caution should also be exercised when using a scale that exhibits a history of DIF and the results should be further examined in future research.

Research question 4: *Does the SSIS-TRS function similarly across native English speaking and Spanish-speaking English language learner students, and demonstrate adequate levels of construct validity as evidenced by scores that are invariant across items, 11 subscales (i.e., Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, Self-control, Externalizing, Bullying, Hyperactivity/Inattention, and Internalizing), and four major scales (i.e., Social Skills, Problem Behaviors, Autism Spectrum and Academic Competence)?*

In this study, results suggested that the second-gate assessment (i.e., SSIS-TRS) maintains a number of subscales and major scales that are invariant across linguistic status. Ultimately, items with DIF were identified indicating that the SSIS-TRS may demonstrate only selected areas of weakness when considering its function and use for Spanish-speaking ELLs. Beginning with the subscales that make up the Social Skills scale, two areas appeared to show

bias when taking into account Spanish-speaking ELL status. Assertion, which Gresham and Elliot (2008) defined as "initiating behaviors, such as asking others for information, introducing oneself, and responding to the actions of others" (p. 1), was found to display measurement bias in favor of native English-speaking students, thus suggesting that this is an area on the SSIS-TRS Social Skills scale that may be misidentifying functioning in Spanish-speaking ELLs. Additionally, Empathy, which Gresham and Elliot (2008) defined as, "showing concern and respect for others' feelings and viewpoints" (p. 2), was found to display measurement bias in favor of Spanish-speaking ELLs, suggesting that this is an area on the SSIS-TRS Social Skills scale that may be misidentifying functioning in native English-speaking students. More specifically, the Assertion scale requires one to be able to express feelings when wronged, ask for help from adults, say when there is a problem, stand up for others who are treated unfairly, question rules that may be unfair, say nice things about herself/himself without bragging, and stand up for herself/himself when treated unfairly. The Empathy scale requires one to be able to try to understand how you feel, try to make others feel better, forgive others, try to understand how others feel, try to comfort others, and show concern for others. In addition to the fact that these two scales may be biased and over identify problem areas for Spanish-speaking ELLs, it is important to remember the impact that a lack of understanding of cultural norms and social rules can have on being able to act assertively and display social skills that are expected in classroom settings. Therefore, as previously mentioned, these results could be interpreted that Spanish-speaking ELLs are more likely to experience social skills difficulties. Additionally, when compared to assertiveness, Empathy is a concept that can be perceived through more passive behaviors. Therefore, for Spanish-speaking ELLs, who are working to learn a new language as well as behavioral expectations and academic material, it is possible that behaviors are

understood as being more of a general willingness to more actively try to understand how others behave and feel. Researchers have also found that Spanish-speaking ELLs often display social skills that emphasize the importance of being respectful (i.e., *respeto)*, demonstrating polite behavior to show one's good upbringing (i.e., *bien educado*), and not questioning figures of authority (Hair et al., 2006).

Although many items were invariant, an item-level analysis revealed that items that were more language-based (e.g., says 'please', questions rules that may be unfair, stands up for self when treated unfairly, says when there is a problem, expresses feelings when wronged, stands up for others, starts conversations with peers, and introduces self to others) demonstrated measurement bias against Spanish-speaking ELLs. Research has further revealed the extent to which standardized assessments act as "cultural artifacts" and that learning a new language in a new culture can place an individual at a disadvantage compared to those who are native to the culture in terms of lack of understanding of social norms, poorer self-advocacy skills, and lower achievement in classroom environments (Barrera et al., 2002; Rhodes et al., 2005; Solano-Flores, 2011). Therefore, an exploration of skills at higher ELP levels may aid in determining if behaviors change as a child becomes more proficient in a language. Moreover, researchers have suggested that Spanish-speaking ELLs are particularly at risk for struggling in social situations and competitive classroom environments that do no allow for opportunities for collaboration and group work (Barrera et al., 2002; Spomer & Cowan, 2001). Other research has also indicated that Spanish-speaking ELLs are more likely to experience (a) increased victimization in kindergarten, (b) difficulty with assertiveness, and (c) social isolation due to perceptions of limited opportunities to interact with peers (Chang et al., 2007; Conchas, 2001).

In the present study, items across the Externalizing, Problem Behavior, and Autism

Spectrum disorder scales (e.g., disobeying rules, getting distracted easily, withdrawing from others, repeating the same thing over and over, having nonfunctional routines or rituals, and using odd physical gestures in interactions) were also biased against Spanish-speaking ELLs, suggesting that Spanish-speaking ELLs experience more behavioral problems than their native English-speaking counterparts. However, items involving forgiveness, turn-taking, kindness towards others when they are feeling badly, taking care of using other people's things, and saying nice things were biased in favor of Spanish-speaking ELLs. Therefore, Spanish-speaking ELLs may not only have access to particular skills that are expected in classrooms settings, but may also display other adaptive behaviors (e.g., code-switching abilities) that are not being identified that ultimately promote social-emotional and academic development. It is possible that skills are also not fully captured by items on the SSIS-TRS. Simultaneously, the conflict and possible racism children from ethnic minority backgrounds face as they work to navigate a majority culture are important to consider (Rhodes et al., 2005). Cultural differences likely influence the outcome of the results on the SSIS-TRS and, therefore as previously mentioned, caution should be exercised when using items that display DIF.

Research question 5: Does the SSIS-TRS function differently across students of varying ELP levels as demonstrated by (a) a lack of invariance in scores across domain areas (i.e., Math skills, Reading skills, Motivation to Learn, and Prosocial Behavior) on the PSG; and (b) a lack of invariance in scores across items, 11 subscales (i.e., Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, Self-control, Externalizing, Bullying, Hyperactivity/Inattention, and Internalizing), and four major scales (i.e., Social Skills, Problem Behaviors, Autism Spectrum and Academic Competence)?

Findings in this study suggested that the PSG and SSIS-TRS function differently at higher and lower ELP levels across Spanish-speaking ELLs. When initially screened, students with lower ELP levels were found to score lower in the areas of Math skills, Prosocial Behavior (i.e., effective communication, cooperative, exhibits self-control, and demonstrates empathy

towards others), and Motivation to Learn (i.e., engagement in instructional activities, ability to stay on task, effort displayed, and attentiveness). Findings therefore suggest that Spanish-speaking ELLs with lower ELP levels receive lower scores on the PSG when compared to Spanish-speaking ELLs with higher ELP levels. Results from an omnibus test revealed that students with higher ELP levels were found to have higher scores across the Communication, Cooperation, and Assertion scales, while also demonstrating fewer problems across the Hyperactivity/Inattention and Internalizing scales as measured by the SSIS-TRS. DIF was also observed across the Assertion, Externalizing, and Bullying scales in favor of Spanish-speaking ELLs at higher ELP levels, thus suggesting that students with greater knowledge of the English language demonstrate such behaviors more readily, or that the scales are biased in these areas. Although DIF was not indicated, students at lower ELP levels were found to score higher on the Responsibility, Empathy, Engagement, and Self-Control scales. They were also found to score higher on the Hyperactive/Inattentive and Internalizing scales. Although ELP levels were not considered along a continuum as ELL students were dichotomized into Group 1 (ELP levels 1, 2, and 3) and Group 2 (ELP levels 4, 5, and 6), Dowdy et al. (2011) and previous researchers have found that ELLs or students with limited ELP levels have more attention problems, difficulty following directions, and demonstrate more learning problems.

An in-depth review of the items showed the presence of DIF in a number of areas. Higher scores were reported for Spanish-speaking ELLs with higher ELP levels across items that directly required linguistic knowledge (e.g., says 'please', questions rules that may be unfair) and a strong understanding of social norms (e.g., takes responsibility for part of a group activity, forgives others, appropriately takes turns, and makes friends easily). Additionally, students with lower ELP levels had more difficulty completing tasks without bothering others. Spanish-

speaking ELLs with lower ELP levels also received scores on items (i.e., such that the items showed DIF) that indicated problems with a preoccupation with object parts, doing things that make others feel scared, keeping others out of social circles, getting distracted easily, and acting without thinking. Conversely, scores for students at lower ELP levels were significant across items that suggested an ability to be nice to others when they are feeling badly, resolve disagreements calmly, and use gestures appropriately with others. Spanish-speaking ELLs with higher ELP levels further scored higher on items that indicated problems associated with temper tantrums, fighting with others, talking back to adults, bullying others, and acting sad or depressed. These exploratory findings further provide evidence that Spanish-speaking ELLs exhibit different patterns of behavior at different ELP levels. Although these results may indicate bias in item functioning, previous research has indicated concerns across various areas of functioning and results can be used to further understand the extent to which areas of strengths and weakness can be supported for Spanish-speaking ELLs so as to promote engagement and academic achievement in classroom settings.

Limitations

One limitation in this study was the sample that was ultimately utilized. Recruited from a mid-western location and given that recruitment was restricted due to constraints implemented by the IRB, it is possible that the results may have varied if a nationally representative sample had been utilized or if additional sites that were more heavily populated by Spanish-speaking ELLs could have been recruited. Regarding the sample itself, a larger number of students were found to be at an ELP level that was at the "developing" level, with fewer students being at either extreme. Few kindergarten, 4th grade, and 5th grade students were recruited for the study, which additionally may have influenced the findings. A larger, random sample could have been

more representative of the actual population. Additionally, a number of the teachers taught at the same school and in the same district, which suggests a similar training background that may have influenced the results. Moreover, ELP levels were based on scores from a measure that was administered months before teachers participated in the present study and therefore ELP levels were likely higher than were indicated.

In addition to the sample's characteristics, it is important to note that the multi-gated assessment system to screen the entire classroom, as is the PSG's intended purpose, was not possible. Given that the study required active consent from parents, it was not possible for teachers to consider every student in their classroom while completing the screeners. Given that the size of the sample was substantial this is not necessarily a limitation but should be considered in interpreting the analyses.

This study measured the use of the PSG and SSIS-TRS as measures of social skills by comparing scores to established measures, but this study did not necessarily further clarify particular strategies or skills that Spanish-speaking ELLs exhibit outside of the SSIS. Ultimately, it is important to make this distinction and not directly make conclusions about actual social skills. This study served to examine whether or not items functioned similarly across Spanish-speaking ELLs, while also considering ELP levels. At the present time, although the SSIS-TRS was used as the measure to compare ratings throughout this multi-gated assessment system, a "gold standard" has not been established due to a narrow amount of the research that has been done on the use of social-emotional and behavioral measures with Spanish-speaking ELLs. In combination with the limited amount of research that has been done with social-emotional and behavioral measures in Spanish-speaking ELL sample, findings from this present study may simply indicate that such measures should be used carefully when assessing functioning in

Spanish-speaking ELLs. Results also heavily relied on teacher perceptions and ratings, and there was no observational or other data to confirm these ratings.

Lastly, given that profiles of Spanish-speaking ELLs across ELP levels have not been thoroughly developed, this study served to explore and establish this knowledge base further. Acculturation, socialization, socio-economic status, and immigration status were not directly measured. Furthermore, according to Osterlind and Everson (2011), when little is known about the nature of a profile it is possible that uniform DIF, or the probability of the occurrence of something is consistently higher for the reference group "over all trait levels", may not be present (p. 11). More specifically, Osterlind and Everson (2011) stated that, "there are instances when the nature of this dependency is contingent on where examinees lay on the underlying ability continuum" (p. 11). Exploration of *nonuniform* DIF may therefore be necessary, but such applications of DIF could not be utilized in the current study due to the small sample size overall and across ELP levels.

Implications and Future Directions

Given the multi-directional relationships across academic achievement, school engagement, and social skills functioning, understanding areas of strength and weakness is critical in order to promote success for every student (Caemmerer & Keith, 2014). Given that students have continued to remain under-referred for social-emotional problems and that social skills have been shown to be an important academic enabler, multi-gated social-emotional and behavioral screening systems are important for developing effective intervention and instructional programs (e.g., effective RTI systems; Elliott et al., 2007; Kratochwill et al., 2007; Whitcomb & Merrell, 2012). Additionally, when compared to their native English-speaking peers, Spanish-speaking ELLs are more likely to experience less school belonging and an

increase in problems associated with social-emotional and behavioral challenges (Kulis, Marsiglia, & Nieri, 2009; Martinez, Degarmo, & Eddy, 2004). Furthermore, culturally and linguistically diverse students historically have been more likely to drop out of high school and tend to be overrepresented in juvenile correctional centers (National Center for Education Statistics, 2011).

Measurement tools must be accurate and sensitive, and the PSG and SSIS-TRS evidence good reliability with Spanish-speaking ELLs. However, it appears from this study and also by previous research (Felt, 2011) that the PSG offers only a small benefit over a simple teacher nomination strategy. More research with a representative sample may be useful in further clarifying this finding.

Analyses of the validity of the use of the PSG and SSIS-TRS in Spanish-speaking ELLs resulted in mixed findings. Multiple items appeared to differentially function (i.e., exhibit DIF) for Spanish-speaking ELLs and across ELP levels, in terms of both under detecting and over detecting problems in social skills development. Research has repeatedly demonstrated that ELLs experience challenges in developing social skills that facilitate classroom behavior and academic engagement as compared to their native English-speaking peers (Chang et al., 2007; Spomer & Cowan, 2001). But it has been unclear what skills different students bring with them as they enter school. Future research could benefit from including multiple measures of student social skills against which to compare screening outcomes. Presently there is not a distinction between whether Spanish-speaking ELLs actually exhibit more problem behaviors, or display discrete adaptive skills that are not being recognized that actually facilitate academic achievement, or that raters or standardized assessment tools are biased against students with limited ELP levels. Consequently, it will be important for future research to establish profiles of

social-emotional and behavioral development across Spanish-speaking ELLs across ELP levels.

However, when more specifically reviewing standardized measures including additional

measures of social skills, such as classroom observations by trained raters, additional information

could be gained regarding behaviors and social skills functioning, which would aid in further

establishing a valid measure. Gathering background information from the student, parent, or rater

on the student's cultural socialization and determining relevant cultural factors are important

issues for consideration, and would further facilitate a better understanding of the extent to which

measures can be validly used in a particular population. Additionally, since the SSIS-TRS

requires teachers to indicate the level of importance of items (i.e., not important, important, and

critical), further analyses will be conducted on the data collected in this study to determine future

directions that research can go in terms of understanding which behaviors are more and less

useful for Spanish-speaking ELLs to display across ELP levels.

 Considering differences in profiles across varying ELP levels will continue to be

important in future research endeavors that examine social skills functioning as well as academic

achievement in ELL populations. Garnering such an understanding will widely facilitate

intervention work and the critical junctures at which educators, parents, and providers should

intervene with students so as to provide treatment that is appropriate and cost-effective. In this

study, Spanish-speaking ELLs elicited scores that suggested potential areas of strength and

weakness across the PSG and SSIS-TRS and that were variable across ELP levels. Although

these scores cannot be used to describe a definitive profile, future research that establishes an

understanding of both the adaptive abilities and problem areas will be important. Undoubtedly,

caution should be taken in using this particular measure in assessing Spanish-speaking ELLs.

Spanish-speaking ELLs demonstrate a unique sets of social skills and social-emotional

development; accurately identifying students for the purpose of information, instruction, and

treatment will continue to be essential.

References

Abedi, J. (2013). Testing of English Language Learner Students. Behavioral, Social, and

Emotional Assessment of Children. In Geisinger, K. (Eds.) *APA Handbook of Testing*

and Assessment in Psychology, Vol. III (pp. 355 – 368). Washington, DC: American

Psychological Association.

Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English

language learners: Implications for policy-based empirical research. *Review of*

Educational Research, 74, 1-28.

Abidin, R. R., & Robinson, L. L. (2002). Stress, biases, or professionalism: What drives teachers'

referral judgments of students with challenging behaviors? *Journal of Emotional and*

Behavioral Disorders, 10, 204-212.

Achenbach, T. M. (1991). *Integrative guide for the 1991 CBCL/6-18, YSR, and TRF Profiles*.

Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M., Bird, H. R., Canino, G., Phares, V., Gould, M. S., & Rubio-Stipec, M. (1990).

Epidemiological comparisons of Puerto Rican and U.S. mainland children: Parent,

teacher, and self-reports. *Journal of the American Academy of Child & Adolescent*

Psychiatry. 29, 84-93.

Achenbach, T.M., & McConaughy, S.H. (1997). *Empirically Based Assessment of Child and*

Adolescent Psychopathology: Practical Applications (2nd edition). Los Angeles,

California: Sage.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms &*

Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, &

Families.

Adelman, H. S., & Taylor, L. (2010). Placing prevention into the context of school improvement. In B. Doll, W. Pfohl, & J.S. Yoon (Eds.), *Handbook of youth prevention science* (pp. 19 – 44). New York, NY: Routledge.

Aisenberg, E. (2001). The effects of exposure to community violence upon Latina mothers and preschool children. *Hispanic Journal of Behavioral Sciences, 23*, 378-398.

Albers, C.A., & Martinez, R. (2015). *Academic success in English: Using response to intervention as best practice for English language learners*. New York: Guilford.

Albers, C.A., Mission, P. L., & Bice-Urbach, B.J. (2013). Considering diverse learner characteristics in problem-solving. In Brown-Chidsey, R., Andren, K. J., & Harrison, P.L. (Eds.) *Assessment for intervention: A problem-solving approach, Second Edition*. (pp. 101 – 124). New York, NY: Guilford.

Albers, C. A., Kenyon, D. M., & Boals, T. J. (2009). Measures for determining English language proficiency and the resulting implications for instructional provision and intervention. *Assessment for Effective Intervention, 34*, 74–85.

Algozzine, B., Wang, C., & Violette, A. S. (2011). Re-examining the relationship between academic achievement and social behavior. *Journal of Positive Behavior Interventions*, *13*, 3-16.

American Educational Research Association, American Psychological Association, & the National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*. Washington, DC: American Psychiatric Association.

American Psychological Association. (2002). *American Psychological Association ethical*

principles of psychologists and code of conduct. Washington, DC: APA.

American Psychological Association Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist*, *61*, 271-285.

Anstrom, K., DiCerbo, P., Butler, F., Katz, A., Millet, J., & Rivera, C. (2010). *A review of the literature on academic English: Implications for K-12 English language learners*. Arlington, VA: The George Washington University Center for Equity and Excellence in Education.

Antshel, K. M., & Remer, R. (2003). Social skills training in children with attention deficit hyperactivity disorder: A randomized-controlled clinical trail. *Journal of Clinical Child and Adolescent Psychology*, *32*, 153–165.

Artiles, A. J., & Ortiz, A. A. (2002). Ethnic disproportionality in special education. In A. J. Artiles & A. A. Ortiz (Eds.). *English language learners with special needs: Identification, assessment, and instruction.* Washington, DC: Center for Applied Linguistics.

Asos, S., Lieb, R., Mayfield, J., Miller, M., & Penucci, A. (2004). *Benefits and costs of prevention and early intervention programs for youth.* Olympia, WA: Washington State Institute for Public Policy.

Bailey, A. L., & Butler, F. A. (2002). *An evidentiary framework for operationalizing academic language for broad application to K-12 education: A design document*. Los Angeles, CA: Center for the Study of Evaluation.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.

Barrera, M., Biglan, A., Taylor, T. K., Gunn, B., Smolkowski, K., Black, C., Ary, D. V., &

Fowler, R. C. (2002). Early elementary school intervention to reduce conduct problems: A randomized trial with Hispanic and non-Hispanic children. *Prevention Science*, *3*, l83– 94.

Barrera, I., Corso, R. M., & Macpherson, D. (2003). *Skilled dialogue: Strategies for responding to cultural diversity in early childhood*. Baltimore, MD: Brookes.

Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. New York, NY: Cambridge University Press.

Bird, H.R., Canino, G., Gould, M.S., Ribera, J., Rubio-Stipec, M., Woodbury, M., Huertas-Goldman, S., & Sesman, M. (1987). Use of the Child Behavior Checklist as a screening instrument for epidemiological research in child psychiatry: Results of a pilot study. *Journal of the American Academy of Child Adolescents Psychiatry, 26*, 207-213.

Bolt, D., Hoyt, W., & Kratochwill, T. R. (2007). Using multilevel modeling to examine the effects of multi-tiered interventions. *Psychology in the Schools, 44*, 503- 256.

Bowman, B., & Moore, E. K. (2005). *School readiness and social-emotional development: Perspectives on cultural diversity*. Washington, DC: National Black Child Development Institute.

Braccio, J. A. (2009). *Teacher ratings of resilience, social skills, and reading and math achievement in Hispanic male learning disabled students*. Dissertation Abstracts International. *70(*6).

Brennan, R. L. (2006). Perspectives on the Evolution and Future of Educational Measurement. In Brennan, R. L. (Ed.) *Educational Measurement 4th Edition*. (pp.1 – 16). Portsmouth, NH: Greenwood Publishing Group, Inc.

Bronson, M. B. (2000). *Self-regulation in early childhood.* New York, NY: Guildford.

Brown-Chidsey, R., & Andren, K.J. (2013). *Ideal Problem Solving: Introduction*. New York, NY: Guilford.

Brown-Chidsey, R. (2005). Academic skills are basic (to) children's personal wellness. *Trainer's Forum: Periodical of the Trainers of School Psychologists*, *25*(1), 4-10.

Burns, B. J., Costello, E. J., Angold, A., Tweed, D., Stangl, D., Farmer, E., & Erkanli, A. (1995). Children's mental health service use across service sectors. *Health Affairs*, *14*, 147–159.

Caemmerer, J. M., & Keith, T. Z. (2014). Longitudinal, reciprocal effects of social skills and achievement from kindergarten to eighth grade. *Journal of School Psychology, 53,* 265-281.

Campbell, S. B. (2002). *Behavior problems in preschool children: Clinical and developmental issues* (2nd ed.). New York, NY: Guilford.

Caldarella, P. Young, E. L., Richardson, M. J., Young, B. J., & Young, K. R. (2008). Validation of the Systematic Screening for Behavior Disorders in middle and junior high school. *Journal of Emotional and Behavioral Disorders, 16,* 105-117.

California Department of Education. (2012). *California English Language Development Test: Annual technical 2010-11 Edition*. Sacramento, California: Author.

Caprara, G.V., Barbaranelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P.G. (2000). Pro-social foundations of children's academic achievement. *Psychological Science*, *11*, 302 – 306.

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science, 11*, 282 – 298.

CDC. (2005). Racial/ethnic disparities in prevalence, treatment and control of hypertension—United States—1999–2002. *Morbid and Mortality Weekly Report*, 54(01), 7-9.

Chapman, C., Laird, J., & KewalRamani, A. (2010). *Trends in high school dropout and*

completion rates in the United States: 1972–2008. U.S. Department of Education.

Washington, DC: National Center for Education Statistics.

Chang, F., Crawford, G., Early, D., Bryant, D., Howes, C., Burchinal, M., Barbarin, O., Clifford, R., & Pianta, R. (2007). Spanish-speaking children's social and language development in pre- kindergarten classrooms. *Early Education and Development, 18,* 243-269.

Charniak, E. (1991). Bayesian networks without tears. *AI Magazine*, *12*, 50-63.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Lawrence Erlbaum.

Commission on Chronic Illness. (1957). *Chronic illness in the United States* (Vol. 1). Cambridge, MA: Harvard University.

Conchas, G. (2001). Structuring failure and success: Understanding the variability in Latino school engagement. *Harvard Educational Review, 71,* 475-503.

Conoley, J. C., & Impara, J. C. (1995). *Social Skills Rating System. By: Gresham, Frank M., Elliott, & Stephen N.* In Twelfth Mental Measurement Yearbook. Accession number 12191056.

Crijnen, A. A. M., Achenbach, T. M., & Verhults, F. C. (1997). Comparison of problems reported by parents of children in 12 cultures: Total problems, externalizing, and internalizing. *Journal of the American Academy of Child & Adolescent Psychiatry, 36,* 1269-1277.

Crijnen, A. A. M., Achenbach, T. M., & Verhulst, F. C. (1999). Problems reported by parents of children in multiple cultures: the Child Behavior Checklist syndrome constructs. *The American Journal of Psychiatry, 156,* 569-574.

Crosnoe, R. (2007). *Mexican roots, American schools: Helping Mexican immigrant children*

succeed: Palo Alto, CA: Stanford University Press.

Darney, D., Reinke, W. M., Herman, K. C., Stormont, M., & Ialongo, N. S. (2012). Children with co-occurring academic and behavior problems in first grade: Distal outcomes in twelfth grade. *Journal of School Psychology, 51,* 117-128.

de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press.

Dever, B.V., & Kamphaus, R. W. (2013). Behavioral, Social, and Emotional Assessment of Children. In Geisinger, K. (Eds.) *APA Handbook of Testing and Assessment in Psychology, Vol. III* (pp. 129 – 148). Washington, DC: American Psychological Association.

Diperna, J. C. (2005). Academic enablers and student achievement: Implications for assessment and intervention services in the schools. *Psychology in the Schools*, *43*, 7-17.

DiPerna, J. C., & Elliott, S. N. (2000). *The Academic Competence Evaluation Scales*. San Antonio, TX: The Psychological Corporation.

Diperna, J. C., & Volpe, R. J. (2005). Self-report on the social skills rating system: Analysis of reliability and validity for an elementary sample. *Psychology in the Schools*, *42*, 345-354.

DiPerna, J. C., Volpe, R. J., & Elliott, S. N. (2002). A model of academic enablers and elementary reading/language arts achievement. *School Psychology Review*, *31*, 298-312.

DiPerna, J. C., Volpe, R. J., & Elliott, S. N. (2005). A model of academic enablers and mathematics achievement in the elementary grades. *Journal of School Psychology*, *43*, 379-392.

Doll, B., & Yoon, J.S. (2010). The current status of youth prevention science. In B. Doll, W. Pfohl, & J.S. Yoon (Eds.), *Handbook of youth prevention science* (pp. 1 – 18). New York, NY: Routledge.

Dowdy, E., Dever, B. V., DiStefano, C., & Chin, J. K. (2011). Screening for emotional and behavioral risk among students with limited English proficiency. *School Psychology Quarterly, 26*, 14 – 26.

Dulay, H., Burt, M., & Krashen, S. (1982). *Language two.* New York, NY: Oxford University Press.

Edl, H. M., Jones, M. H., & Estell, D. B. (2008). Ethnicity and English proficiency: Teacher perceptions of academic and interpersonal competence in European American and Latino students. *School Psychology Review, 37,* 38–45.

Elliott, S. N., Barnard, J., & Gresham, F. M. (1989). Preschoolers' social behavior: Teachers' and parents' assessments. *Journal of Psychoeducational Assessment, 7,* 223-234.

Elliott, S. N., Huai, N., & Roach, A. T. (2007). Universal and early screening for educational difficulties: Current and future approaches. *Journal of School Psychology, 45,* 137-161.

Esquivel, G., & Keitel, M. (1990). Counseling immigrant children in the schools. *Elementary School Guidance and Counseling, 24,* 213-221.

Evangelista, N., & McLellan, M. J. (2004). The Zero to Three Diagnostic System: A framework for considering emotional and behavioral problems in young children. *School Psychology Review, 33,* 159 – 173.

Feil, E. G., Severson, H. H., & Walker, H. M. (1998). Screening for emotional and behavioral delays: The early screening project. *Journal of Early Intervention, 21,* 252-266.

Feil, E. G., & Walker, H. M. (1995). The early screening project for young children with

behavior problems. *Journal of Emotional and Behavioral Disorders, 3,* 194-203.

Felt, J. (2011). *Social skills measurement in the Latino English language learner*

population: The Social Skills Improvement System Performance Screening Guide.

(Unpublished thesis). University of Wisconsin-Madison, Madison, Wisconsin.

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, 169 – 180.

Fishman, J. A., (1965). Who speaks what language to whom and when? *La Linguistique*, *2*, 67-

88.

Francis, D. J., Rivera, M., Lesaux, N., Kieffer, M., & Rivera, H. (2006). *Practical guidelines for*

the education of English language learners: Research-based recommendations for

instruction and academic interventions. Portsmouth, NH: RMC Research Corporation,

Center on Instruction.

Galindo, C., & Fuller, B. (2010). The social competence of Latino kindergartners and growth in

mathematical understanding. *Developmental Psychology, 46,* 579-592.

García Coll, C., & Marks, A. (2009). *Immigrant stories: Ethnicity and academics in middle*

childhood. New York, NY: Oxford University Press.

Garrison, W.T., & Earls, F. (1985). The Child Behavior Checklist as a screening instrument for

young children. *Journal of the American Academy of Child Psychiatry*, 24, 76-80.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference.*

11.0 update (4th ed.). Boston: Allyn & Bacon.

Gettinger, M., Ball, C., Mulford, L., & Hoffman, A. (2010). Prevention and early intervention for

preschool children at risk for learning and behavior problems. In B. Doll, W. Pfohl, & J.S.

Yoon (Eds.), *Handbook of youth prevention science* (pp. 349 – 374). New York, NY: Routledge.

Glover, T.A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology*, *45*, 117-135.

Goodman, G., & Webb, M. M. (2006). Reading disability referrals: Teacher bias and other factors that impact response to intervention. *Learning Disabilities: A Contemporary Journal, 4*, 59-70.

Gonzalez, L. E. (2000). *The relationship between classroom settings and language models with regard to the development of language and social skills of Spanish-speaking preschoolers with language disorders*. Pace University. (Unpublished doctoral dissertation). New York City and Westchester, NY.

Gottleib, J., Gottlieb, B. W., & Trongue, S. (1991). Parent and teacher referrals for a psychoeducational evaluation. *Journal of Special Education*, *25*, 155-167.

Gottlieb, M., & Hamayan, E. (2007). Assessing oral and written language proficiency: A guide for psychologists and teachers. *Handbook of multicultural school psychology: An interdisciplinary perspective.* Mahwah, NJ: Lawrence Erlbaum, Inc.

Greene, M. (1992). Evaluation and dignity. *Quarterly of the National Writing Project*, *14*, 10-13.

Greenfield, P.M. (1997a). Culture as process: Empirical methods for cultural psychology. In J.W. Berry, Y.H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology (2nd Ed), Vol. 1: Theory and method* (pp. 301 - 346). Needham Heights, MA: Allyn & Bacon.

Greenfield, P.M. (1997b). You can't take it with you: Why ability assessments don't cross cultures. *American Psychologist, 52*, 1115 - 1124.

Greenfield, P.M., Quiroz, B., & Raeff, C. (1998). Cross-cultural conflict and harmony in the

social construction of the child. In S. Harkness, C. Raeff, & C.M. Super (Eds.), *The social construction of the child: Nature and sources of variability. New directions in child psychology*. San Francisco: Jossey-Bass.

Gresham, F. M., & Elliott, S. N. (1990). *Social Skills Rating System*. Minneapolis, MN: NCS Pearson, Inc.

Gresham, F. M., & Elliott, S. N. (2008). *Rating Scales Manual: SSIS – Social Skills Improvement System*. Minneapolis, MN: NCS Pearson, Inc.

Gresham, F. M., & Elliott, S. N. (2008). *Social Skills Intervention Guide (SSIG)*. Minneapolis, MN: NCS Pearson, Inc.

Gross, D., Fogg, L., Young, M., Ridge, A., Cowell, J. M., Richardson, R., & Sivan, A. (2006). The equivalence of the child behavior checklist/1½-5 year olds across parent race/ethnicity, income level and language. *Psychological Assessment, 18*, 313–323.

Hair, E., Halle, T., Terry-Humen, E., Lavelle, B., & Calkins, J. (2006). Children's school readiness in the ECLS-K: Predictions to academic, health, and social outcomes in first grade. *Early Childhood Research Quarterly, 21,* 431-454.

Harry, B. (1994). Behavioral disorders in the context of families. In R. L. Peterson & S. Ishii-Jordan (Eds.), *Multicultural issues in the education of students with behavioral disorders* (pp. 149-161). Cambridge, MA: Brookline Books.

Hebbeler, K., Spiker, D., & Kahn, L. (2012). Individuals with Disabilities Education Act's Early Childhood Programs: Powerful vision and pesky details. *Topics in Early Childhood Special Education*, *31*, 199 – 207.

Heller, K. A., Holtzman, W., & Messick, S. (1982). *Placing children in special education: A strategy of equity*. Washington, DC: National Academy Press.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H.I. Braun (Eds.), *Test Validity* (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Hosp, J. L., & Madyun, N., H. (2007). Addressing disproportionality with response to intervention. In S.R. Jimerson, M.K. Burns, & A.M. VanDerHeyden, (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 172-181). New York, NY: Springer.

Hosterman, S. J., DuPaul, G. J., Jitendra, A. K. (2008). Teacher ratings of ADHD symptoms in ethnic minority students: Bias or behavioral difference? *School Psychology Quarterly*, *23*, 418 - 435.

Huey, S.J., & Polo, A.J. (2008). Evidence-based psychosocial treatments for ethnic minority youth. *Journal of Clinical Child and Adolescent Psychology, 37*, 262-301.

Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: conceptual, methodological, and statistical issues. *Psychological Assessment*, *15*, 446-455.

Institute of Medicine. (2002). *Unequal treatment: Confronting racial and ethnic disparities in health care.* Washington: National Academies Press.

Jenkins, D. C., Sr. (1998). *The use of the Systematic Screening for Behavior Disorders in a culturally diverse school*. Retrieved from Dissertation Abstracts International Section A: Humanities and Social Sciences, Vol. 58(10-A), pp. 3832.

Jimerson, S. R., Burns, M. K., VanDerHeyden, A. M. (2007). Response to intervention at school:

The science and practice of assessment and intervention. In S.R. Jimerson, M.K. Burns, & A.M. VanDerHeyden, (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 3-9). New York, NY: Springer.

Juvonen, J., Espinoza, G., & Knifsend, C. (2012). The role of peer relationships in student academic and extracurricular engagement. In S.L. Christenson, A. L., Reschly, & C. Wulie, (Eds.) *Handbook of research on student engagement* (pp. 387-401). New York, NY: Springer.

Kane, M. T. (2006). Validation. In Brennan, R. L. (Eds.) *Educational measurement* (4th Ed.; pp.17-64). Portsmouth, NH: Greenwood Publishing Group, Inc.

Kang, E. (2005). *Measurement equivalence of the Child Behavior Checklist among White, Hispanic, and Black low-income families.* Dissertation Abstracts International: Section B: The Sciences and Engineering, 65(9). Pennsylvania State University, Philadelphia, Pennsylvania.

Kaplan, R. M. (2000). Two pathways to prevention. *American Psychologist*, *55*, 382-396.

Karabenick, S. A., & Clemens Noda, P. A. (2004). Professional development implications of teachers' beliefs and attitudes toward English language learners. *Bilingual Research Journal: The Journal of the National Association for Bilingual Education*, *28*, 55-75.

Kataoka, S. H., Zhang, L., & Wells, K.B. (2002). Unmet need for mental health care among U.S. children: Variation by ethnicity and insurance status. *American Journal of Psychiatry, 159,* 1548-1555.

Kettler, R.J., Glover, T.A., Albers, C.A., & Feeney-Kettler, K.A. (Eds.) (2014). *Universal screening in educational settings: Evidence-based decision making for schools.* Washington, DC: American Psychological Association.

Kulis, S., Marsiglia, F. F., & Nieri, T. (2009). Perceived ethnic discrimination versus acculturation stress: Influences on substance use among Latino youth in the southwest. *Journal of Health and Social Behavior*, *50*, 443–459.

Lacalle, M., Ezpeleta, L., & Doménech, J. M. (2012). DSM-oriented scales of the Child Behavior Checklist and Youth Self-Report in clinically referred Spanish children. *The Spanish Journal of Psychology*, *15*, 377-387.

Landau, S., & Milich, R. (1990). Assessing children's social status and peer relations. In A. M. La Greca (Ed.) *Through the eyes of the child: Obtaining self-reports from children and adolescents* (pp. 259-291). Boston, MA: Allyn and Bacon.

Landy, S. (2009). *Pathways to competence: Encouraging healthy social and emotional development in young children*. Baltimore, MD: Paul H. Brookes Publishing Co.

Lane, K. L., Richards-Tutor, C., Oakes, W. P., & Connor, K. (2013). Initial evidence for the reliability and validity of the Student Risk Screening Scale with elementary age English learners. *Assessment for Effective Intervention, 37*(2), 99 – 122.

Leadbeater, B. J., & Bishop, S. J. (1994). Predictors of behavior problems in preschool children of inner-city Afro-American and Puerto Rican adolescent mothers. *Child Development*, *65*, 638 – 648.

LeBoeuf, W. A., Fantuzzo, J. W., & Lopez, M. L. (2010). Measurement and population miss-fits: A case study on the importance of using appropriate measures to evaluate early childhood interventions. *Applied Developmental Science, 14*, 45-53.

Lee, I. (1998) Supporting greater autonomy in language learning. *English Language Teaching Journal*, *52*, 282–289.

Lequerica, M., & Hermosa, B. (1995). Maternal reports of behavior problems in preschool Hispanic children: An exploratory study in preventive pediatrics. *Journal of the National Medical Association*, *87*, 861-8.

Levitt, J. M., Saka, N., Romanelli, L. H., & Hoagwood, K. (2007). Early Identification of mental health problems in schools: The status of instrumentation. *Journal of School Psychology*, *45*, 163-191.

Lochman, J.E. (1995). Screening of child behavior problems for prevention programs at school entry. *Journal of Consulting Psychology*, *63*, 549 – 559.

Lopez, M. L., Tarullo, L. B., Forness, S. R., & Boyce, C. A. (2000) Early identification and intervention: Head Start's response to mental health challenges. *Early Education & Development, 11*, 265-282.

Losen, D. J., & Orfield, G. (2002). *Racial inequity in special education.* Cambridge, MA: Harvard Education Press.

Lynch, M. J., & Hanson, E. W. (1992). *Developing cross cultural competence: A guide for working with young children and their families.* Baltimore, MD: Brookes.

Manz, P. H., Fantuzzo, J. W., & McDermott, P. A. (1999). The parent version of the preschool social skills rating scale: An analysis of its use with low-income, African American children. *School Psychology Review, 28*, 493–504.

Martinez, C. R., Degarmo, D. S., & Eddy, J. M. (2004). Promoting academic success among Latino students. *Hispanic Journal of Behavioral Sciences*, *26*, 128–151.

Matson, S. C., & Haglund, K. (2000). Relationship between scholastic and health behaviors and

reading level in adolescent females. *Clinical pediatrics*, *39*(5), 1-9.

McCurdy, M., Coutts, M.J., Sheridan, S.M., & Campbell, L.M. (2013). Ecological variables in school-based assessment and intervention planning. New York, NY: Guilford.

Malecki, C. K., & Elliott, S. N. (2002). Children's social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly*, *17*(1), 1-23.

Miller, F. G., Cohen, D., Chafouleas, S.M., Riley-Tillman, T. Chris, Welsh, M.E. & Fabiano, G.A. (2015). A comparison of measures to screen for social, emotional, and behavioral risk. *School Psychology Quarterly*, *30*, 184-196.

Moss, P. A. (1994). Can there be validity without reliability? *Educational Researcher, 23,* 5-12.

Muniz, J., Hambleton, R. K., & Xing, D. (2001). Small sample studies to detect flaws in item translations. *International Journal of Testing*, *1*, 115-135.

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

National Association of State Directors of Special Education. (2005). *Response to intervention: Policy considerations and implementation*. Alexandria, VA: Author.

National Center for Education Statistics. (2011). *The nation's report card: Reading 2011* (NCES 2012–457). Washington, DC: Institute of Education Sciences, U.S. Department of Education. Retrieved from http://nces.Ed.gov/nationsreportcard/pdf/main2011/2012457.pdf

No Child Left Behind Act of 2001 (Pub. L. No, 107-110), 20 U.S.C. 6301 et seq.

Noam, G. G., & Hermann, C. A. (2002). Where education and mental health meet:

Developmental prevention and early intervention in schools. *Development and Psychopathology*, *14*, 861-875.

Oakland, T., Wechsler, S. M., & Maree, K. (2013). Test use with children across cultures: A view from three countries. In K. Geisinger (Ed.), *APA handbook of testing and assessment in psychology, Vol. III* (pp. 231 – 257). Washington DC: American Psychological Association.

OECD. (2011). *Education at a glance 2011: Highlights*. Washington DC: Author.

Ollendick, T. H., Oswald, P. D., & Francis, G. (1989). Validity of teacher nomination in identifying aggressive, withdrawn, and popular children. *Journal of Clinical Child Psychology, 18,* 221-229.

Ortiz, S. (2002). Best practices in nondiscriminatory assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 1321–1336). Washington, DC: National Association of School Psychologists.

Ortman, J. M. and Shin, H. B. (2011). Language projections: 2010 to 2020. *American Sociological Association Annual Meeting, NV Online.* Retrieved from http://www.allacademic.com/meta/p508373_index.html.

Osterlind, S. J., & Everson, H.T. (2009). *Differential item functioning* (2nd ed.). New York, NY: SAGE Publications, Inc.

Parshall, C. G., & Miller, T. R. (1995). Exact versus asymptotic Mantel‑Haenszel DIF statistics: A comparison of performance under small-sample conditions. *Journal of Educational Measurement*, *32*, 302-316.

Pfiffner, L. J., & McBurnett, K. (1997). Social skills training with parent generalization: Treatment effects for children with attention deficit disorder. *Journal of Consulting and*

Clinical Psychology, 65, 749–757.

Pianta, R. C., Laparo, K., Payne, C., Cox, M., & Bradley, R. (2002). The relation of kindergarten classroom environment to teacher, family, and school characteristics and child outcomes. *The Elementary School Journal, 102,* 225-238.

Podell, D. M., & Soodak, L. C. (1993). Teacher efficacy and bias in special education referrals. *The Journal of Educational Research*, *86*, 247-253.

Portes, P. R. (2013). Ethnicity and culture in educational psychology. In E. Anderman & L. Corno (Eds.), *Handbook of educational psychology* (pp. 358 -398). New York, NY: Routledge.

Powless, D. L., & Elliott, S. N. (1993). Assessment of social skills of Native American preschoolers: Teachers' and parents' ratings. *Journal of School Psychology, 31,* 293-307.

Ray, C. E., & Elliott, S. N. (2006). Social adjustment and academic achievement: A predictive model for students with diverse academic and behavior competencies. *School Psychology Review*, *35*, 493-501.

Reinhard, P., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In S. L. Christenson, A. L., Reschly, & C. Wulie (Eds.), *Handbook of research on student engagement* (pp. 259-282). New York, NY: Springer.

Reynolds, C.R., & Lowe, P.A. (2009). The problem of bias in psychological assessment. In T.B. Gutkin & C.R. Reynolds (Eds.), *The handbook of school psychology* (pp. 332-374). Hoboken, NJ: Wiley.

Reynolds, A. J., & Walberg, H. J. (1991). A structural model of science achievement. *Journal of Educational Psychology*, *83*, 97-107.

Reynolds, A. J., & Walberg, H. J. (1992). A structural model of science achievement and

attitude: An extension to high school. *Journal of educational Psychology, 84*, 371-382.

Rivera, B. D., & Rogers-Atkinson, D. (1997). Culturally sensitive interventions: Social skills training with children and parents from culturally and linguistically diverse backgrounds. *Intervention in School and Clinic, 33,* 75-80.

Rhodes, R., Ochoa, S. H. & Ortiz, S. O. (2005). *Assessment of culturally and linguistically diverse students: A practical guide*. New York, NY: Guilford Press.

Ruffalo, S. L., & Elliott, S. N. (1997). Teachers' and parents' ratings of children's social skills: A closer look at cross-informant agreements through an item analysis protocol. *School Psychology Review, 26*, 489–501.

Roussos, L. A., & Stout, W. F. (1996). Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel Haenszel Type I Error Performance. *Journal of Educational Measurement*, *33*, 215-230.

Sattler, J. M. (2008). *Assessment of children: Cognitive Foundations* (5th ed.). La Mesa, CA: Jerome M. Sattler, Publisher, Inc.

Saville-Troike, M. (1987). Bilingual discourse: The negotiation of meaning without a common code. *Linguistics, 25*, 81–106.

Severson, H. H., Walker, H. M., Hope-Doolittle, J., Kratochwill, T. R., & Gresham, F. M. (2007). Proactive, early screening to detect behaviorally at-risk students: Issues, approaches, emerging innovations, and professional practices. *Journal of School Psychology, 45,* 103-223.

Sideridis, G. D., Antoniou, F., & Padeliadu, S. (2008). Teacher biases in the identification of learning disabilities: An application of the logistic multilevel model. *Learning Disability Quarterly*, *31*, 199-209.

Sheridan, S. M., Knoche, L. L., Edwards, C. P., Bovaird, J. A., & Kupzyk, K. A. (2010). Parent engagement and school readiness: Effects of the getting ready intervention on preschool children's social-emotional competencies. *Early Education and Development, 21*, 125-156.

Shonkoff, J. P., & Phillips, D. (2000). *From neurons to neighborhoods: The science of early childhood.* Washington, DC: National Academy Press.

Smith, G. T., Fischer, S., & Fister, S. M. (2003). Incremental validity principles in test construction. *Psychological Assessment*, *15*, 467.

Solano-Flores, G. (2011). Assessing the cultural validity of assessment practices: An introduction. In M. del Rosario Basterra, E. Trumbull, & G. Solano-Flores, (Eds.), *Cultural Validity in Assessment* (pp. 3-21). New York, NY: Taylor & Francis.

Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching*, *38*, 553-573.

Spomer, M. L., & Cowen, E. L. (2001). A comparison of the school mental health referrals profile of young ESL and English speaking children. *Journal of Community Psychology*, *29*, 69-82.

Stephens, T. M. (1992). *Social Behavior Assessment Inventory*. Psychological Assessment Resources: Lutz, FL.

Stipek, D. & P. Byler, (2001). Academic achievement and social behaviors associated with age of entry into kindergarten. *Applied Developmental Psychology 22*, 175-189.

Strom, K. (2000). *Profile of state prisoners under age 18, 1985-97*. US Department of Justice, Office of Justice Programs. Washington, DC: Bureau of Justice Statistics.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic

regression procedures. *Journal of Educational Measurement*, *27*, 361-370.

Takanishi, R. (2004). Leveling the playing field: Supporting immigrant children from birth to eight. *The Future of Children*, *14*, 60-79.

Thorp, E. K. (1997). Increasing opportunities for partnership with culturally and linguistically diverse families. *Intervention in School and Clinic*, *32*, 261-269.

Toppelberg, C. O., Medrano, L., Morgens, L. P., & Nieto-Castanon, A. (2002). Bilingual children referred for psychiatric services: Associations of language disorders, language skills, and psychopathology. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*, 712-722.

United States Department of Health and Human Services. (1999). *Mental health: A report of the Surgeon General*. Rockville, MD: United States Department of Health and Human Services, Substance Abuse and Mental Health Administration, Center for Mental Health Services, National Institutes of Health, National Institute of Mental Health.

United States Office of the Surgeon General. (2001). *Mental health: Race, ethnicity, and culture: A supplement to the surgeon general's report on mental health*. Washington: Substance Abuse and Mental Health Administration, U.S. Department of Health and Human Services.

United States Department of Education. (1999, 2004). *Individuals with Disabilities Education Ac*t (IDEA).

Van Horn, M., Atkins-Burnett, S., Karolin, E., Ramey, S. L. & Snyder, S. (2007). Parent ratings of children's social skills: Longitudinal psychometric analyses of the social skills rating system. *School Psychology Quarterly, 22*, 110-119.

Walker, H. M., & Severson, H. H. (1990). *Systematic Screening for Behavior Disorders: User*

Manual. Longmont, CO: Sopres West.

Walker, H. M., & Severson, H. H. (1992). *Systematic Screening for Behavior Disorders (2nd ed.).* Longmont, CO: Sopris West.

Walker, H. M., Severson, H. H., Nicholson, F., & Kehle, T. (1994). Replication of the Systematic Screening of Behavior Disorders (SSBD) procedure for the identification of at-risk children. *Journal of Emotional and Behavioral Disorders, 2,* 66-77.

Walker, H. M., Severson, H. H., Todis, B. J., & Block-Pedagogy, A. E. (1990). Systematic Screening for Behavior Disorders (SSBD): Further validation, replication, and normative data. *Remedial & Special Education, 11,* 32-46.

Walker, A., Shafer, J., & Iiams, M. (2004). Not in my classroom: Teacher attitudes towards English language learners in the mainstream classroom. *NABE Journal of Research and Practice, 2,* 130-160.

Weiss, S. J., Goebel, P., Page, A., Wilson, P., & Warda, M. (1999). The impact of cultural and familial context on behavioral and emotional problems of preschool Latino children. *Child Psychiatry and Human Development, 29,* 287– 301.

Wentzel, K. R., & Watkins, D. E. (2002). Peer relationships and collaborative learning as contexts for academic enablers. *School Psychology Review, 31,* 366-377.

Whitcomb, S.A., & Merrell, K.W. (2012). *Behavioral, social, and emotional assessment of children and adolescents* (4th ed.). New York, NY: Taylor & Francis.

White, J. L., Albers, C. A., DiPerna, J. C., Elliott, S. N., Kratochwill, T. R., & Roach, A. T. (2004). *Cost analysis in educational decision making: Approaches, procedures, and case examples.* Madison, WI: Wisconsin Center for Educational Research.

WIDA Consortium (2010). Annual technical report no. 5, volume 1 of 3: Description, validity, and student results. *Annual technical report for ACCESS for ELLs® English Language Proficiency Test, Series 200, 2008-2009 administration.* Madison, WI: Author.

Willinsky, J. (1990). *The new literacy: Redefining reading and writing in the schools*. New York, NY: Routledge.

Winter, S. M., & Kelley, M. F. (2008). Forty years of school readiness research: What have we learned? *Journal of the Association for Childhood Education International, Annual Theme, 84*, 260-266.

Zins, J. E., & Elias, M. E. (2006). Social and emotional learning. In G. G. Bear & K. M. Minke (Eds.), *Children's Needs III: Development, prevention, and intervention* (p. 1-13). Bethesda, MD: National Association of School Psychologists.

Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry, 39*(4), 561-577.

Appendix A: Teacher Referral Form

Teacher Referral Form

Directions: When considering **ONLY** the students, whose parents signed a consent form, list **ONLY** those students for whom you have concerns about their social skills functioning, emotional development and/or behavior. DO NOT consider the results from the Performance Screening Guide. Please list only the students whom you would refer for the purpose of providing them with additional support.

IMPORTANT: next to their name write <u>SELL</u> if the student is a **Spanish-speaking English language learner** and <u>ELL</u> if the student is **any other type of English language learner**.

Student
1.
2.
3.
4.
5.
6.
7.
8.
9.
10.
11.
12.
13.

Appendix B: Teacher Directions

Dear Teacher,

Thank you so much for your participation in this research. I hope this study will be valuable and benefit teachers and students.

The items that are included in this packet should be:

1) **2 Teacher Consent Forms**
2) **1 Performance Screening Guide Protocol**
3) **1 Teacher Referral Form**
4) **1 Teacher Survey**
5) **Guardian Consent Forms in English and Spanish**
6) **6 Social Skills Improvement System–Teacher Rating Scales**

To thank you for your participation you will receive $50. Your total participation will take approximately 2 hours.

Please read through this letter before completing anything.

1. **First**, please sign the Teacher Consent Form.

2. **Second**, please send consent forms home to all of your students for their guardians to sign and return to you. When you present your class with consent forms explain that they need to take to it home to their parents/guardians and have them sign it and bring it back to you. There is a copy for the parent(s)/guardian(s) to sign and a copy for them to keep. You can use the script below:

   "Ok everyone, I have something to give you. You all have a chance to participate in a project about social skills. I am giving you a letter that you need to take home to your parents/guardians and have them sign. Then I need you to bring it back to me as soon as you can. If your parent(s)/guardian(s) have any questions they can contact the people that are listed here on the back of this letter." (show contact information on the back of one of the consent forms). **Then make sure each student is given two consent forms in English and two consent forms in Spanish (the parents will keep one copy and have their child return the other one).**

   If you are able to meet with guardians (i.e., parents) directly you can use this script to describe the study and collect consent forms directly from them:

   ___(School name)___ has been selected to be a part of an exciting research study, and I just wanted to take a minute to explain the study and ask for your consent. As you may know, social-emotional functioning has a significant impact on students' academic achievement and mental health. Determining a child's social skills functioning is important in order to determine whether or not he or she is in need of services that would prevent academic failure and mental illness. However we do not know how these measures identify social skills in Spanish-speaking English language learners. This research will seek to establish this.

   This research is just requiring that you allow me to fill out a screener and rating scale for your child. And if your child has been given the World-Class Instructional Design and Assessment (WIDA) Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs®) to be able to access his or her scores. Your child's information will be kept completely confidential and I will not have access to your child's final scores on these measures.

   If you consent to this research I am just asking you to sign this consent form. And you may keep this additional copy for your records.

   You will then wait (at least 2 weeks) until you receive forms from all of the parents/guardians of the students in the class. Additionally, please fill out page three of this packet, and the teacher survey on page four of this packet.

STOP…wait until you receive at least 6 parental consent forms

3. **Third**, Complete the <u>Performance Screening Guide Protocol (PSG)</u>.
   Instructions for completing the PSG: First, fill in the information regarding your classroom on the back cover of the form. You may leave the purpose of the evaluation empty. Next, flip the form over, open it and in the far right-hand column under "*Student Name*" write the name of each student whose guardian(s) provided consent.
   *NOTE: If the student is a **Spanish-speaking ELL** write <u>SELL</u> next to the student name. If the student is **any other type of English language learner** write <u>ELL</u> next to the student's name.
   * WARNGING: ONLY go on to complete teacher-rating scales for students who are NATIVE ENGLISH SPEAKERS or SPANISH-SPEAKING Ells.
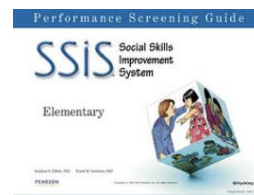
   Instructions on how to complete the form may be found at the top right-hand side on the inside of the front cover. Please be sure to read the definitions and descriptions of each performance level (1-5) *BEFORE* beginning your rankings. Also please be sure to circle the performance level number next to each student's name that matches that student's *CURRENT* level of functioning.

4. **Fourth**, follow the directions on the <u>Teacher Referral Form (TRF)</u> and complete the TRF for the student's whose guardian(s) provided consent.

5. **Fifth**, look at the <u>Performance Screening Guide Protocol (PSG)</u>. Circle *THREE* native English speakers with the most number of "1s". Then circle *THREE* Spanish-speaking English language learners that have the most number of "1s" on the PSG. If there are not enough native English-speaking students or Spanish-speaking English language learners with "1s", choose students that have the highest number of "2s", followed by the highest number of "3s" until you have selected a total of *SIX* students. (If students have an equal number of "1s", "2s", or "3s", alphabetize the letters that are assigned to each student and pick every other student from the list, until you have a total of *THREE* native English-speaking students and *THREE* Spanish-speaking English language learners.)

6. **Sixth,** Complete the <u>Social Skills Improvement System–Teacher Rating Scales (SSIS–TRS)</u> for each of the *SIX* students.

   * WARNGING: ONLY complete teacher-rating scales for students who are NATIVE ENGLISH SPEAKERS or SPANISH-SPEAKING ELLS

   Open the SSIS-TRS and on the side of the form fill in the information regarding the student's full name, date, student's birth date, student's gender, student's grade level, your full name, your position, and how long you have known the student.

   Please be sure to rate students based on their behavior during the *PAST TWO MONTHS*. Also, please be sure to *MARK AN ANSWER FOR EVERY ITEM* (<u>and only one answer per item</u>), even if you feel it does not apply to that student. Just give your best estimate. Instructions on how to complete the form may be found on the top left-hand page.

For Spanish-speaking English language learners please also list the student's ACCESS score in the space provided to you on the inside of the form.

7. **<u>Seventh</u>**, after you complete *SIX* Social Skills Improvement System–Teacher Rating Scale (SSIS–TRS) (i.e., *THREE* Spanish-speaking English language learner students, *THREE* native English-speaking students) place all of the materials in the folder and contact the student researcher (Paige Mission, mission@wisc.edu; (650) 743-6236).

<u>Reminders</u>: Please remember to keep these materials *CONFIDENTIAL* at *ALL* times. Do not leave forms on your desk or in places where other individuals may access them. Respect the privacy and confidentiality of your students.

Remember that these measures have not been validated for use with Spanish-speaking ELLs. Therefore ratings generated by the PSG may not be true indicators of the student's social skills.

Most importantly please remember that you may ask any questions about the research at any time. If you have any questions about the research you may contact the research team:
Principal Investigator: **Craig A. Albers, PhD at caalbers@wisc.edu (608) 262–4586**
Student researcher: **Paige L. Mission, MS at mission@wisc.edu (650) 743-6236**

Your participation is completely voluntary. If you decide not to participate or withdraw from the study, you may do so at any time. Withdrawal from the study will have no effect on your access to the research materials.

Thank you for your participation in this research,

Paige L. Mission
PhD Student, School Psychology
Educational Psychology
University of Wisconsin-Madison

Name_____ Date_____ School_____

Grade level currently teaching_____

Please indicate how the student researcher can best contact you:

Email Address _____

Day-Time Phone Number _____

Night-Time Phone Number _____

I prefer to be contacted by (circle your preference): **Phone** or **Email**

Appendix C: Teacher Survey

Please fill in the circles and/or write your response for the following questions regarding <u>your personal and professional background</u>:

1. What is your gender?
   - ○ Male
   - ○ Female
   - ○ Other

2. What is your country of origin?
   - ○ United States
   - ○ Other

3. Do you consider yourself to be Hispanic/Latino?
   - ○ Yes
   - ○ No

4. How do you describe your racial background?
   (Mark all that apply)
   - ○ American Indian or Alaska Native
   - ○ Asian
   - ○ Black or African American
   - ○ Native Hawaiian or Other Pacific Islander
   - ○ White
   - ○ Other (Please Specify)_____

5. How many years have you been teaching?
   - ○ < 3 years
   - ○ 3-5 years
   - ○ 5-10 years
   - ○ 10-20 years
   - ○ > 20 years

6. What is the highest educational level you have achieved?
   - ○ High school diploma
   - ○ GED
   - ○ Bachelors degree
   - ○ Associates degree
   - ○ Masters degree
   - ○ Doctoral degree
   - ○ Other (Please Specify) _____

7. What language(s) do you speak?

   *(EXAMPLE: Language: **English**___ Circle level of proficiency: Native Speaker / Fluent / (Proficient) / Conversational / Beginner)*

   Language:_____ Circle level of proficiency: Native Speaker / Fluent / Proficient / Conversational / Beginner

   Language:_____ Circle level of proficiency: Native Speaker / Fluent / Proficient / Conversational / Beginner

   Language:_____ Circle level of proficiency: Native Speaker / Fluent / Proficient / Conversational / Beginner

   Language:_____ Circle level of proficiency: Native Speaker / Fluent / Proficient / Conversational / Beginner

   Language:_____ Circle level of proficiency: Native Speaker / Fluent / Proficient / Conversational / Beginner

   Language:_____ Circle level of proficiency: Native Speaker / Fluent / Proficient / Conversational / Beginner

   Language:_____ Circle level of proficiency: Native Speaker / Fluent / Proficient / Conversational / Beginner

Please complete the following questions regarding your classroom:

8. How many students are in your class? (Please Specify) _____

9. Of the students in your classroom, how many students are Spanish speaking English language learners?

   (Please Specify) _____