Psychometric Defensibility of the Intervention Selection Profile – Social Skills (ISP-SS) with Students at Risk for Behavioral Concerns

Stephen P. Kilgus, Ph.D.

Katie Eklund, Ph.D.

University of Wisconsin-Madison

Nathaniel P. von der Embse, Ph.D.

University of South Florida

Psychology in the Schools

Note: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A180515 to the University of Wisconsin-Madison. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Correspondence concerning this article should be addressed to Stephen P. Kilgus, Department of Educational Psychology, University of Wisconsin-Madison, 1025 W Johnson St, Madison, WI 53706. E-mail: skilgus@wisc.edu

Abstract

The purpose of the current study was to examine the validity and diagnostic accuracy of the *Intervention Selection Profile-Social Skills* (ISP-SS), a brief social skills assessment tool intended for use with students in need of Tier 2 intervention. Participants included 160 elementary and middle school students who had been identified through universal screening as at risk for behavioral concerns. Teacher participants (n = 71) rated each of these students using both the ISP-SS and the *Social Skills Improvement System – Rating Scales* (SSiS-RS), with the latter measure serving as the criterion within validity and diagnostic accuracy analyses. Confirmatory factor analysis supported ISP-SS structural validity, indicating ISP-SS items broadly conformed to a single 'Social Skills' factor. Follow-up analyses suggested ISP-SS broad scale scores demonstrated adequate internal consistency reliability, with hierarchical omega coefficient equal to .86. Correlational analyses supported the concurrent validity of ISP-SS items, finding each ISP-SS item to be moderately or highly related to its corresponding SSiS-RS subscale. Finally, analyses indicated three of the seven ISP-SS items demonstrated sufficient diagnostic accuracy; however, findings suggest additional revisions are needed if the ISP-SS is to be appropriate for use in schools. Implications for practice and future research are discussed.

Keywords: brief behavior rating scales, tier 2, social skills training, social skills assessment

Psychometric Defensibility of the Intervention Selection Profile – Social Skills (ISP-SS) with Students at Risk for Behavioral Concerns

Social skills instruction (SSI) is a frequently used Tier 2 intervention within multi-tiered systems of support, and is defined as systematic instruction of prosocial skills (e.g., cooperation, self-control) that are used to complete a range of social tasks (Sanetti & Simonsen, 2011). SSI is considered particularly appropriate for students evidencing acquisition ("can't do") deficits within their social skills profile, as opposed to those evidencing performance ("won't do") deficits (for whom contingency management interventions might be more appropriate). Extant Tier 2 research has supported the effectiveness of SSI in promoting positive skills and academic engagement, as well as decreasing problem behavior (Bruhn, Lane, & Hirsch, 2014). However, additional studies have indicated that SSI effectiveness is moderated by instructional match; that is, SSI is more effective when instructional content is matched to a student's specific acquisition deficit(s) (Barreras, 2009; Gresham et al., 2006). For example, SSI would be less effective if a student's skill deficit is in the area of *assertion* and the lessons provided address some other skill, such as *self-control*. Such research has therefore established a precedent for the selection of lessons that are aligned to a student's particular skill deficits to increase the likelihood of SSI effectiveness.

Social Skills Assessment

To support instructional content matching, it is necessary for schools to engage in social skills assessment. Such assessment can be conducted with students who have been identified as at risk (e.g., via universal screening) for the purpose of (1) determining if SSI is indeed appropriate for each student (i.e., given the presence of acquisition deficits), and (2) identifying which acquisition deficits should be targeted for instruction. Researchers have developed a range

of social skill assessments, including semi-structured interviews, self-report instruments, sociometric techniques, and rating scales (e.g., *Walker–McConnell Scale of Social Competence and School Adjustment*; Walker & McConnell, 1995). Though promising, many of these methods lack the psychometric evidence required to support use in applied settings (Kilgus, von der Embse, Scott, & Paxton, 2015). Furthermore, many of these methods require a prohibitive amount of funds, time, or effort to complete (Whitcomb & Merrell, 2013). Accordingly, while these methods might be suitable for use with a smaller number of students at Tier 3, there use would likely be challenging at Tier 2.

One of the most commonly researched and used social skills assessments, which has been proposed for use at both Tiers 2 and 3, is the *Social Skills Improvement System – Rating Scales* (SSiS-RS; Gresham & Elliott, 2008). The SSiS-RS represents a suite of assessment tools, which include teacher, parent, and student self-report forms applicable to children ages 3-18. Once completed, the SSiS-RS affords information regarding a student's functioning at both the overall Social Skills level and within each of these seven areas: Communication, Cooperation, Responsibility, Empathy, Assertion, Engagement, and Self-Control. For instance, scale and subscale scores can be contrasted with national norms in determining how a student compares to his or her same-aged peers (e.g., below or above average). Furthermore, item scores can be considered in determining which specific social skills might be considered acquisition or performance deficits. Information regarding these deficits can then support the selection of specific SSI lessons aligned with individual student needs.

The SSiS-RS possesses multiple strengths, including its strong psychometric support and its alignment with a specific SSI curriculum. Yet, limitations are also noted, including the time and effort required to complete the tool. The SSiS-RS teacher rating scale includes 46 Social

Skills-specific items, which take approximately 5-12 minutes to complete per student (Gresham & Elliott, 2008). Though not prohibitive for any individual student, the need to collect such data at scale would like strain teacher time and available resources. To illustrate, within any given classroom, it can be expected that 15-20% of students will be at risk for social-emotional and behavioral concerns (Schanding & Nowell, 2013). Within a classroom of 20 students, this expectation translates to three or four students requiring Tier 2 intervention. If the SSiS-RS Social Skills items were completed for all of these students, teachers would be required to dedicate 15-48 minutes of their time. When this time is combined with that dedicated to other forms of social-emotional and behavioral assessment (e.g., universal screening and progress monitoring), the SSiS-RS is likely to tax the time available to teachers outside of their other instructional and academic assessment activities.

Development of Novel Social Skills Assessments

Although the SSiS-RS authors suggest the measure is suitable for use at both Tier 2 and 3, recognition of its limited feasibility suggest the measure might often be restricted to use at Tier 3 with a smaller number of more challenging students. Accordingly, recent work has focused on the development of more brief and efficient social skills assessment tools appropriate for use with students at Tier 2. For instance, Kilgus et al. (2015) developed and initially validated the *Intervention Selection Profile – Social Skills* (ISP-SS), a novel social skills assessment tool intended for use across the K-12 spectrum. Like the SSiS-RS, the ISP-SS was designed to assess broad social skills functioning, as well as abilities within seven narrow social skill areas (e.g., communication, engagement). To enhance the tool's efficiency, the ISP-SS was not designed to afford more molecular information regarding specific skill deficits within these seven areas (as is possible via the SSiS-RS). Rather, the measure was intended to identify in which of seven areas a

student is likely to have one or more specific skill deficits. Though broader in nature, such information could still be used to inform SSI matching through the identification of lesson units relevant to student needs.[1]

In constructing the ISP-SS, and in the interest of maximizing brevity and efficiency, Kilgus et al. (2015) endeavored to identify a set of just seven items. Once identified, it was intended that (a) the seven items would collectively correspond to broader social skills functioning, and that (b) each item would correspond to a particular skill area, such that item level scores would speak to the presence or absence of skill deficits within a specific skill area. ISP-SS items were drawn from the initial 21-item pre-factor analyzed version of the *Social and Academic Behavior Risk Screener* (SABRS; Kilgus, Chafouleas, & Riley-Tillman, 2013). SABRS items were chosen for consideration in recognition of their relevance to social functioning within the school settings. Within the Kilgus et al. study, 54 elementary teachers (grades K-5) assessed 243 students randomly selected from the school population (approximately five students per classroom). Teachers rated each student using both the SABRS and the SSiS-RS. ISP-SS development followed a clinimetric approach, through which SABRS items were selected for ISP-SS inclusion based upon their clinical utility (Feinstein, 1987). In the context of social skills assessment, utility was defined as an item's capacity to accurately predict the presence of skill deficits within a narrow social skill area. This capacity was examined via a series of agreement and conditional probability statistics. Each possible SABRS item was compared to each SSiS-RS subscale. A SABRS item was selected as corresponding to a

[1] Although the ISP-SS is not aligned with any particular SSI curriculum, it was designed to assess the social skill areas commonly targeted via existing SSI curricula (e.g., the aforementioned *Social Skills Intervention Guide*).

particular SSiS-RS subscale (and thus included in the ISP-SS) if it yielded some combination of at least moderate kappa (>.40), acceptable sensitivity (>.80), or acceptable specificity (>.70).

In accordance with an argument-based approach to validation (Kane, 2013), Kilgus et al. (2015) findings yielded support for both the interpretation and use of ISP-SS scores. First, support for ISP-SS *use* was found via the aforementioned clinimetric analyses, which identified six items that might be used to predict the presence of skill deficits within six of the seven social skill domains assessed on the SSiS-RS. Though ISP-SS items did not necessarily always appear conceptually relevant to their corresponding domain (e.g., ISP-SS *temper outbursts* and SSiS-RS *empathy*), each item nevertheless exhibited acceptable levels of agreement (mean kappa = .52) and conditional probability statistics (mean sensitivity = .80, mean specificity = .86) relative to its corresponding SSiS-RS domain. More specifically, ISP-SS items were found to agree with their SSiS-RS domains when the former were dichotomized, such that Likert ratings of 0-1 (Never-Sometimes) represented the presence of a skill deficit, and ratings of 2-3 (Often-Almost Always) did not. Second, support for ISP-SS *interpretation* was documented via reliability analyses, which suggested that the ISP-SS items might be collectively interpreted as indicators of a broader Social Skills construct (Cronbach's alpha = .89).

Though promising, three notable limitations to the Kilgus et al. (2015) investigation should be noted. First, the Kilgus et al. study was conducted with a single sample. Cross-validation with new samples is now required to support applied ISP-SS use. Second, the prior investigation did not yield an ISP-SS item that could be used to predict deficits in the social skill area of Communication. As such, further item development and testing remains necessary if the ISP-SS is to predict a broad range of social skill difficulties. Third, the Kilgus et al. study was conducted within a broader sample inclusive of students at risk for behavioral concerns, as well

as students who were not at risk (approximately 65% of that sample, per the SSiS-RS). Yet, the ISP-SS is intended for use with students who are at risk alone for the purposes of (a) determining if an student who is at risk possesses social skill deficits (thereby making SSI appropriate), and (b) differentiating between deficits to inform SSI programming. Previous research suggests that when conducting research regarding a measure intended for use with students who are at risk, the use of a broader and more general sample may result in misestimation of the measure's psychometric properties (Briesch, Swaminathan, Welsh, & Chafouleas, 2014). Specifically, findings may reflect an overestimation of the accuracy with which the ISP-SS identifies and differentiates between skill deficits. As such, research with a sample of only students who are at risk is required to yield a more realistic depiction of ISP-SS validity and diagnostic accuracy.

Summary and Purpose

A broader purpose of the current investigation was to address each of the shortcomings to the Kilgus et al. (2015) investigation. Three specific research questions were of interest. First, to what extent is the ISP-SS a structurally valid indicator of broader student social skills? Kilgus et al. diagnostic accuracy findings suggested that while each ISP-SS item corresponded to a particular skill domain (thus supporting its potential to inform SSI programming), internal consistency reliability evidence further suggested the items might collectively correspond to a single *Social Skills* factor (in a manner consistent with the SSiS-RS subscale-to-total scale hierarchical structure). Thus, the extent to which ISP-SS items were a structurally valid indicator of *Social Skills* was evaluated via (a) confirmatory factor analysis, which examined the fit of a unidimensional factor structure, and (b) the calculation of hierarchical omega coefficient (Hancock & Mueller, 2001), which corresponded to the reliability of items within the single factor. Unidimensional model fit was also compared to that of a theory-driven multidimensional

model, wherein each ISP-SS item loaded on either an *Interpersonal Skills* or *Social-Emotional Regulation* factor. It was hypothesized that the unidimensional model would yield acceptable fit that was similar to that of the less parsimonious multidimensional model. Furthermore, it was expected that ISP-SS items would demonstrate adequate internal consistency within the single *Social Skills* factor, as evidenced by hierarchical omega coefficients approximating one.

Second, to what extent is the ISP-SS a concurrently valid predictor of the SSiS-RS? This was evaluated via bivariate correlational analyses, examining the relation between ISP-SS items and SSiS-RS subscales, as well as the ISP-SS total scale and the SSiS-RS Social Skills total scale. It was hypothesized ISP-SS items would be moderately to highly correlated with SSiS-RS subscales, with the highest correlations being between each ISP-SS item and its corresponding SSiS-RS subscale. It was further expected that the ISP-SS total scale would be moderately to highly associated with the SSiS-RS Social Skills total scale. Third, to what extent are ISP-SS item cut scores diagnostically accurate predictors of their corresponding social skill domains, represented by dichotomously scored SSiS-RS subscales? This was evaluated via the calculation of conditional probability statistics (e.g., sensitivity, specificity), which indicated the extent to which ISP-SS item cut scores reliably differentiated between those with and without social skill difficulties within each of the seven skill domains assessed via the SSiS-RS. It was hypothesized that at least one cut score associated with acceptable sensitivity and specificity would be identified within each ISP-SS item, supporting that item's use in identifying social skill deficits.

Method

Participants

This study was conducted in one elementary school and two middle schools within the southwestern United States. Participants included 160 elementary and middle school students (K-

8; 76.7% male), as well as the 33 elementary and 38 middle school teachers who rated student behavior. As part of a separate investigation, students in three schools were universally screened for behavioral risk using the *BASC-2 Behavioral and Emotional Screening System* (BESS; Kamphaus & Reynolds, 2007). Students from those schools were considered eligible for the current study if their BESS *T* score fell in the *elevated* (>1 standard deviation) or *extremely elevated* (>2 standard deviations) range for behavioral risk. Participants included 67 elementary students and 93 middle school students (age $M = 11.06$ years, $SD = 2.64$). Regarding ethnicity, 62.9% of students were White, 22.0% Hispanic/Latino(a), 5.7% Black, 1.3% Asian-American, and 3.1% identified with more than one ethnic category. Ethnicity data were only available for 158 of 160 student participants. Across the three schools, the average percentage of students who qualified for free/reduced lunch was equal to 33% (range 32-35%).

Measures

ISP-SS. Teachers completed two behavior rating scales for each student. The first of these was the ISP-SS (Kilgus et al., 2015). The ISP-SS is structured such that each of its seven items corresponds to a skill domain within which a student might have social skill deficits. These seven items (and corresponding domains) include (1) Engages in appropriate conversations with adults and peers (Communication); (2) Compliance with adult directions (Cooperation); (3) Initiation of peer interactions (Assertion); (4) Cooperation with peers (Responsibility); (5) Temper outbursts (Empathy; reverse scored); (6) Joining of peer activities (Engagement); and (7) Polite and socially appropriate responses toward others (Self-Control). While the latter six items were examined by Kilgus et al. (2015), the first item was not. Rather, the first author developed this new item for the purposes of this study given that the previous ISP-SS investigation did not result in the identification of an item corresponding to the *Communication* domain. The new ISP-

SS item was specifically designed to represent the extent to which an individual participates in socially appropriate interactions with others (Whitcomb & Merrell, 2013). Each ISP-SS item was rated on a 4-point Likert Scale (0 = *Never*, 1 = *Sometimes*, 2 = *Often*, and 3 = *Almost Always*), with ratings corresponding to teacher perceptions of the frequency with which students engaged in each behavior during the previous month. The ISP-SS total score represented the sum of all items, with the *temper outbursts* item first reverse coded (0=3 and 1=2).

SSiS-RS. Teachers also completed items from the Social Skills scale of the SSiS-RS (Gresham & Elliott, 2008). The Social Skills scale is comprised of 46 items, which the teacher rates in terms of both frequency (0 = *Never*, 1 = *Seldom*, 2 = *Often*, and 3 = *Almost Always*) and perceived importance for achieving success in the classroom (0 = *Not important,* 1 = *Important*, or 2 = *Critical*). Each SSiS-RS item corresponds to one of seven subscales, including *Communication* (polite and courteous social responses), *Cooperation* (helping and sharing; compliance with rules and directions), *Assertion* (initiating, establishing, and sustaining relationships), *Responsibility* (regard for property, work, and one's actions), *Empathy* (understanding and appreciating others' emotions), *Engagement* (responding to others; joining or initiating activities), and *Self-Control* (appropriate responding in or out of conflict situations). Multiple studies have supported SSiS-RS technical adequacy, with results indicative of SSiS-RS internal consistency, test–retest and inter-rater reliability, and concurrent validity (e.g., Gresham, Elliott, Cook, Vance, & Kettler, 2010; Gresham, Elliott, Vance, & Cook, 2011).

Two sets of SSiS-RS scores were calculated. The first was an overall Social Skills norm-referenced standard score (M = 100, SD = 15). The second set of scores corresponded to subscale scores, which represented the sum of scores within each particular skill domain. Subscale scores were compared to national norms in converting scores into behavior levels, which include *below*

average (summed score greater than one *SD* below the mean), *average* (within one *SD* above or below the mean), and *above average* (greater than 1 *SD* above the mean). Students were considered to possess skill difficulties within each subscale area if their summed score fell within the below average range. These dichotomous behavior level scores (0 = no skill difficulties, 1 = skill difficulties) served as criteria within diagnostic accuracy analyses.

Procedures

The current study was conducted within the context of a broader grant-funded investigation, the purpose of which was to support the development of a social-emotional and behavioral assessment system. The first study within this broader investigation corresponded to the validation of the *Social, Academic, and Emotional Behavior Risk Screener* (SAEBRS; Kilgus & von der Embse, 2014), a universal screening tool. As part of that study, the SAEBRS was compared to the BESS, with the latter serving as a criterion measure in evaluating the former's validity and diagnostic accuracy. All students within the three participating schools (for whom a signed parental opt-out form was not received) were screened using both the BESS and the SAEBRS. Students identified as at risk for social-emotional and behavioral concerns were then eligible for participation in this second study, pertaining to ISP-SS validation.

Prior to the study, researchers met with teachers to recruit those interested in participating in the broader investigation inclusive of both studies. Consenting teachers were provided a brief overview of study purposes and informed that their participation would involve both (1) the universal screening of all students in their classroom and (2) the completion of problem analyses measures (i.e., the ISP-SS and SSiS-RS) for all students identified as at risk via the BESS. Teachers viewed each of the study measures and were given detailed instructions and procedures for completing each measure. Researchers were available to answer any questions. Following

this meeting, teachers were provided parental opt-out forms, which were to be sent home to the parents of each student within their classroom. Students were considered eligible for the broader investigation if no signed opt-out form was returned within two weeks. If an opt-out form was received at any point within the investigation, the student was removed from the study and his or her data were destroyed. Following this initial two week period, each teacher screened all participating students in their classroom. Researchers then prepared reports for each school, indicative of which students were identified as at risk for social-emotional and behavioral concerns via the BESS.

Approximately one month later, teachers were provided hyperlinks to online survey software, which was used to access and complete the ISP-SS and SSiS-RS for students identified through the BESS. (To note, one paper-and-pencil SSiS-RS form was purchased for every SSiS-RS administered via the online survey software. This paper form was then destroyed to ensure each form was administered only once and to account for its corresponding online usage.) The order in which teachers completed these two measures for each student was randomly determined, permitting counterbalancing of measure administration. All research procedures were approved by the school district and the university Institutional Review Board.

Data Analysis Plan

Research question 1. ISP-SS structural validity was first evaluated via CFA, which was conducted using Mplus v.7.11 (Muthén & Muthén, 1998-2012). A confirmatory approach to factor structure examination was chosen over a more exploratory approach given prior reliability findings (Kilgus et al., 2015), which informed reasonable expectations for the ISP-SS factor structure (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Thompson, 2004). Nevertheless, the

current confirmatory approach involved model comparisons, permitting examination of whether varying factor structures provided better fit to ISP-SS data.

In recognition of theory and prior ISP-SS research (Kilgus et al., 2015), the primary factor structure of interest represented a unidimensional model, wherein each ISP-SS item loaded on a single factor labeled *Social Skills*. The fit of this particular model was compared to that of a multidimensional model that specified two factors, including *Interpersonal Skills* (ISP-SS items 1, 2, 3, and 6) and *Social-Emotional Regulation* (ISP-SS items 4, 5, and 7). This latter model was founded in research suggesting social functioning is related to one's ability to manage his or her emotions while also engaging with others in a constructive manner (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011). More complicated multidimensional factor models were not considered given the small number of ISP-SS items and the need for any factor structure considered to be identified with at least three items per factor.

Within each CFA, the loading for the first item in each factor was fixed to one to identify the model. In accordance with prior research and theory, the residual terms associated with two ISP-SS items were permitted to covary given their mutual correspondence to student participation in peer activities (i.e., *joining of peer activities* and *initiation of peer interactions*). Factors were also permitted to covary within the multidimensional model. A robust weighted least squares (WLSMV) estimator was used to account for the non-normality of certain item distributions and the ordinal nature of the Likert scale data. Four fit statistics were collectively evaluated in determining model fit, including the chi-square test of model fit (χ^2), comparative fit index (CFI), Tucker-Lewis index (TLI), and root mean square error of approximation (RMSEA). Resulting values were compared to the following cutoffs in evaluating model fit: χ^2 $p > .05$, RMSEA < .08, and CFI/TLI > .90 (Little, 2013). Pattern coefficients were also calculated,

representative of the relationship between each item and factor after accounting for all other factors.

As a follow-up to CFA, the reliability of the ISP-SS items within factors associated with the superior factor structure was evaluated via the hierarchical omega coefficient (*H*). The *H* coefficient is considered appropriate when considering items that may be multidimensional in nature, but for which a single general factor is still expected (Widhiarso & Ravand, 2014). This is contrasted with coefficient alpha, which presumes item unidimensionality. *H* coefficient values range between 0 and 1, with higher values corresponding to greater reliability.

Research question 2. Pearson's product-moment correlation coefficients were calculated to examine the concurrent validity of ISP-SS items (Likert ratings) and total scale (summed scores) relative to SSiS-RS subscales (summed scores) and overall scale (standard scores). Each ISP-SS item and scale was compared to all SSiS-RS scales; however, the correlation of particular interest corresponded to the relation between each ISP-SS item/scale and the SSiS-RS scale representative of the item's social skill domain (e.g., ISP-SS *cooperation with peers* and SSiS-RS *Responsibility*). Correlational magnitude was compared to Cohen's (1988) criteria for low (<.10), moderate (>.30), and high (>.50). Correlational analyses were conducted twice, supporting disaggregation of analyses by grade level (Elementary [K-5] and Middle [6-8]).

Research question 3. Receiver operating characteristic (ROC) curve analyses were used to evaluate the diagnostic accuracy of cut scores within each ISP-SS item, with cut scores of interest being Likert ratings of =0 (Never), ≤1 (Sometimes), and ≤2 (Often). The accuracy of each cut score within an ISP-SS item was considered relative to the item's corresponding SSiS-RS subscale (i.e., the criterion), which was scored dichotomously in the manner described in the Measures subsection above. Prior to the examination of individual cut scores, the area under the

curve (AUC) was calculated for each ISP-SS item, providing an indicator of the item's overall diagnostic accuracy. Though no gold standard interpretive criteria have been identified, a common heuristic for AUC interpretation defines AUCs between .50–.70 as indicative of low accuracy, .70 –.90 moderate, and .90 –1.00 high (Streiner & Cairney, 2007). AUC 95% confidence intervals (CIs) were also calculated using the asymptotic exact method.

Next, the diagnostic accuracy associated with individual cut scores was evaluated via four conditional probability statistics. *Sensitivity* (SE) was defined as the proportion of students with social skill difficulties who were correctly identified as such. *Specificity* (SP) was defined as the proportion of students without difficulties who were correctly identified as such. *Positive predictive values* (PPV) corresponded to the proportion of individuals identified as having social skill difficulties who actually had difficulties. *Negative predictive values* (NPV) corresponded to the proportion of individuals identified as not having difficulties who actually did not have difficulties. SE and SP were primarily considered in the interest of identifying which cut score performed best within each ISP-SS item. In accordance with previous research (e.g., Kilgus, Riley-Tillman, Chafouleas, Christ, & Welsh, 2014), SE values of ≥.80 were considered acceptable while values ≥.70 were considered borderline. In addition, SP values of ≥.70 and ≥.60 were considered acceptable and borderline, respectively. For each SE and SP value, 95% CIs were calculated using a bootstrapping method with 2000 stratified bootstrapped replicates.

Interest in PPV and NPV was largely related to the current sample of students who are at risk for behavioral concerns (identified as such via the BESS), wherein the majority of students possessed skill deficits (as defined via the SSiS-RS). Predictive values indicated each ISP-SS item's applied impact given the base rate of skill deficits. Specifically, PPV and NPV values indicated the percentage of students in the sample that would have been incorrectly provided (1-

PPV) or denied (1-NPV) instruction if the ISP-SS had been used to support SSI-related decisions. As with correlational analyses, ROC curve analyses were calculated twice, supporting disaggregation by grade level. All research question 2 and 3 analyses were conducted via R v.3.3.1 packages, including psych and pROC.

Results

Research Question 1

See Table 1 for descriptive statistics associated with each ISP-SS item. Unidimensional model fit was supported by two of four fit statistics, including TLI (.94) and CFI (.96). In contrast, the χ^2 test was found to be statistically significant ($p < .001$) and RMSEA fell slightly above the cutoff value (.15). Pattern coefficients ranged between .38 and .94, with all but one coefficient falling above .50. The pattern of model fit was identical for the multidimensional model, with two of four statistics supporting model fit (χ^2 $p < .001$, RMSEA = .16, CFI = .96, TLI = .94). Such lack of improved fit supported retention of the more parsimonious unidimensional model. Follow-up reliability analyses specific to the unidimensional model supported the reliability of items within the *Social Skills* factor, with *H* coefficient equal to .86. See Figure 1 for a graphical depiction of the unidimensional factor structure, along with corresponding pattern coefficient values.

Research Question 2

See Table 1 for a summary of bivariate correlational findings indicative of the concurrent validity of each ISP-SS item. At both the elementary and middle levels, correlations between each ISP-SS item and its corresponding SSiS-RS subscale were either in the moderate or high range, as well as statistically significant at the $p < .01$ or .05 level. This was with the exception of the correlation between ISP-SS *cooperation with peers* and SSiS-RS *Responsibility* at the middle

school level, which fell in the low range. Further review indicated that within five and three SSiS-RS subscales at the elementary and middle school levels (respectively), the highest correlation pertained to the ISP-SS item most aligned with the subscale (e.g., *Initiation of peer interactions* [ISP-SS] and Assertion [SSiS]). The ISP-SS total scale was statistically significantly correlated ($p < .01$) with all SSiS-RS scales, with the largest correlation being with the SSiS-RS total scale ($r = .80$).

Research Question 3

See Table 2 for ROC curve findings. AUC values consistently fell in the moderate range across both grade levels. This was with the exception of two values, which fell in or below the low range. Further review of cut score-level results indicated that in accordance with Kilgus et al. (2015) findings, the ≤1 cut consistently performed the best among evaluated cut scores. Thus, in the interest of brevity, we report findings for this cut score alone in Table 2. Six ISP-SS items yielded acceptable or borderline SE at the elementary level, while five items did at the middle school level. All items yielded acceptable or borderline SP at the elementary level, while two did not at the middle school level. It should be noted that 95% CIs were rather large, suggesting the imprecision of SE and SP estimates, likely resulting from the comparatively low sample sizes at each grade level. Finally, PPV values ranged widely (.34-.93), as did NPVs (.00-.92).

Discussion

The primary goal of the current study was to support continued validation of the ISP-SS, while also improving upon shortcomings of prior research. Most notably, the current study considered ISP-SS performance within a sample of students who are at risk for behavioral concerns, thus yielding a more accurate estimate of ISP-SS psychometric defensibility within its intended applied context. Taken together, CFA and reliability findings supported the structural

validity of the ISP-SS, suggesting ISP-SS items might be reasonably interpreted as indicators of a single *Social Skills* factor. ISP-SS validity was further supported by correlational findings, which indicated ISP-SS items were correlated with their corresponding SSiS-RS subscale to a predominantly medium or large and statistically significant ($p < .001$) degree. Further examination of correlational patterns suggested the presence of convergent and discriminant relationships in accordance with expectations for certain items. Specifically, the majority of SSiS-RS subscales were most highly correlated with their corresponding ISP-SS items.

Though promising, diagnostic accuracy findings were indicative of the somewhat inconsistent performance of ISP-SS items in predicting social skill deficits. While six ISP-SS items were associated with borderline or acceptable SE and SP at the elementary level, only three items exhibited such performance at the middle school level. The remaining items did not reach acceptable thresholds for both sensitivity and specificity, with SE values being particularly problematic. When considered relative to prior research, the current findings suggest the ISP-SS might be well-equipped to inform decisions regarding whether a student requires instruction in Communication, Cooperation, or Engagement. In contrast, the ISP-SS appears less well-equipped to inform decisions regarding Assertion, Responsibility, Empathy, or Self-Control.

Findings therefore appear to support the need for revisions to the ISP-SS if the measure is to possess utility beyond the prediction of the broader social skills functioning. Researchers may develop and evaluate novel ISP-SS items to supplement or replace existing items. As noted above, six of the seven ISP-SS items were selected for inclusion within the measure via a clinimetric approach. These items were not specifically developed to correspond to each social skills domain, but rather were selected based upon their ability to predict social skill deficits within a pilot sample. Given this approach to development, it may be expected that some items

might underperform in a new sample. An alternative method of item development might then call for the generation of items more specifically aligned with each social skills domain. The current findings provide support for such an approach, with the newly developed 'Communication' item yielding particularly strong validity and diagnostic accuracy evidence. Such ISP-SS revisions are currently under way as part of a Goal 5: Measurement grant from the Institute of Education Sciences (IES).

Limitations and Future Directions

Certain limitations to the current study should be noted. First, the current findings are susceptible to mono-method and mono-informant biases, as both the ISP-SS and SSiS-RS represent behavior rating scales completed by classroom teachers. Future studies ought to employ alternative outcome measures (e.g., systematic direct observation) or multiple raters (e.g., parents, self-report measures) to provide a more rigorous and less biased test of ISP-SS performance. Second, the current sample was somewhat restricted in size (n = 160). Though the number of participants did not appear to affect the power of analyses (i.e., given that several statistically significant correlations were noted), it does limit the generalizability of findings. With that said, the current sample was relatively large given that it was purposefully restricted to students at risk for behavioral difficulties. Regardless, future research should include larger samples to enhance the external validity of resulting conclusions.

Third, related to the previous point, the restricted sample size precluded disaggregation of structural validity analyses by grade level. As such, we were unable to determine whether the ISP-SS factor structure and internal consistency varied across these two groups. Future studies should employ larger samples to support the use of such large-sample data techniques by grade level. Fifth, the paper considers internal consistency reliability alone. Future research should

consider alternate forms of reliability (e.g., test-retest, inter-rater), as it can be expected that other facets of measurement (e.g., time, raters) may account for variability in ISP-SS scores. Sixth, unfortunately teacher demographic data were not collected. As a result, it is impossible to determine to which population of teachers the current findings might generalize. It is therefore recommended that such information be collected in future investigations.

References

Barreras, R. B. (2009). *An experimental analysis of the treatment validity of the social skills deficit model for at-risk adolescents* (Unpublished doctoral dissertation). University of California, Riverside, CA.

Briesch, A. M., Swaminathan, H., Welsh, M., & Chafouleas, S. M. (2014). Generalizability theory: A practical guide to study design, implementation, and interpretation. *Journal of School Psychology, 52*(1), 13-35.

Bruhn, A. L., Lane, K. L., & Hirsch, S. E. (2014). A review of tier 2 interventions conducted within multitiered models of behavior prevention. *Journal of Emotional and Behavioral Disorders, 22*(3), 171-189.

Cohen, J. (1988). *Statistical power analysis for the behavior sciences* (2nd ed.). New York, NY: Academic Press.

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

Eklund K., & Dowdy, E. (2014). Screening for behavioral and emotional risk versus traditional school identification methods. *School Mental Health, 6*(1), 40-49.

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

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299.

Feinstein, A. R. (1987). *Clinimetrics*. New Haven, CT: Yale University Press.

Gresham, F. M. (1981). Social skills training with handicapped children: A review. *Review of Educational Research, 51*(1), 139–176.

Gresham, F. M., & Elliott, S. N. (2008). *Social skills improvement system: Rating scales*. Bloomington, MN: Pearson.

Gresham, F. M., Elliott, S. N., Cook, C. R., Vance, M. J., & Kettler, R. (2010). Cross-informant agreement for ratings for social skill and problem behavior ratings: An investigation of the Social Skills Improvement System—Rating Scales. *Psychological assessment*, *22*(1), 157-166.

Gresham, F. M., Elliott, S. N., Vance, M. J., & Cook, C. R. (2011). Comparability of the Social Skills Rating System to the Social Skills Improvement System: Content and psychometric comparisons across elementary and secondary age levels. *School Psychology Quarterly, 26*(1)*,* 27-44.

Gresham, F. M., Van, M., & Cook, C. R. (2006). Social skills training for teaching replacement behaviors: Remediation of acquisition deficits for at-risk children. *Behavioral Disorders, 31*(4), 32–46.

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In S. Cudeck, S. Du Troit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future—A Festschrift in honor of Karl Jöreskog* (pp. 195-216). Lincolnwood, IL: Scientific Sofware International.

Kamphaus, R. W., & Reynolds, C. R. (2007). *BASC-2 Behavioral and Emotional Screening System*. Minneapolis, MN: Pearson.

Kane, M. K. (2013). The argument-based approach to validation. *School Psychology Review, 42*(4), 448-457.

Kilgus, S. P., Chafouleas, S. M., & Riley-Tillman, T. C. (2013). Development and initial validation of the Social and Academic Behavior Risk Screener for elementary grades. *School Psychology Quarterly, 28*(3), 210-226.

Kilgus, S. P., Riley-Tillman, T. C., Chafouleas, S. M., Christ, T. J., & Welsh, M. E. (2014). Direct behavior rating as a school-based behavior universal screener: Replication across sites. *Journal of School Psychology, 52*(1), 63-82.

Kilgus, S. P., von der Embse, N. P., Scott, K., & Paxton, S. (2015). Use of the Intervention Selection Profile–Social Skills (ISP-SS) to identify social skill acquisition deficits: A preliminary validation study. *Assessment for Effective Intervention, 40*(4)*,* 228-239.

Little, P. T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.

Merrell, K. W. (1993). Using behavioral rating scales to assess social skills and antisocial behavior in school settings: Development of the School Social Behavior Scales. *School Psychology Review, 22*(1), 115 133.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Schanding, G. T., & Nowell, K. P. (2013). Universal screening for emotional and behavioral problems: Fitting a population-based model. *Journal of Applied School Psychology*, *29*(1), 104-119.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.

von der Embse, N. P., Pendergast, L., Kilgus, S. P., & Eklund, K. R. (2016). Evaluating the applied use of a mental health screener: Structural validity of the Social, Academic, and Emotional Behavior Risk Screener. *Psychological Assessment, 28*(10)*,* 1265-1275.

Whitcomb, S. A., & Merrell, K. W. (2013). *Behavioral, social, and emotional assessment of children and adolescents* (4th ed.). New York, NY: Routledge.

Widhiarso, W., & Ravand, H. (2014). Estimating reliability coefficient for multidimensional measures: A pedagogical illustration. *Review of Psychology*, *21*(2), 111-121.

Table 1

Bivariate correlation coefficients indicative of the relationship between Intervention Selection Profile – Social Skill (ISP-SS) items and Social Skills Improvement System – Rating Scales (SSiS-RS) scales

	Descriptive Statistics			SSiS-RS Subscale							
ISP-SS Item/Scale	M	SD	BR[1]	Comm	Coop	Assert	Resp	Emp	Eng	SC	Total
Elementary School											
1	1.51	0.79	.45	**.74**	.50	.39	.59	.49	.63	.39	.74
2	1.73	0.90	.55	.59	**.55**	.36	.48	.37	.51	.26	.60
3	1.54	0.64	.45	.48	.11	**.60**	.11	.45	.78	.09	.53
4	0.91	0.95	.54	.55	.46	.04	**.55**	.56	.56	.44	.63
5	1.76	0.78	.22	-.52	-.11	.22	.48	**-.42**	-.28	-.77	-.49
6	1.67	0.77	.42	.50	.10	.37	.20	.47	**.71**	.17	.52
7	1.64	0.77	.45	.74	.42	.09	.69	.63	.60	**.56**	.75
Total	11.94	3.95	N/A	.82	.43	.34	.61	.67	.82	.55	**.85**
Middle School											
1	1.46	0.64	.51	**.67**	.49	.06	.52	.52	.32	.60	.67
2	1.88	0.81	.59	.40	**.62**	-.03	.51	.28	.04	.31	.42
3	1.61	0.63	.34	.27	-.31	**.43**	-.18	.30	.72	.08	.33
4	0.81	0.90	.46	.42	.14	-.05	**.22**	.29	.31	.54	.41
5	1.72	0.77	.24	-.51	-.35	.14	-.41	**-.43**	-.27	-.70	-.54
6	1.47	0.70	.43	.43	-.11	.37	.03	.39	**.74**	.25	.49
7	1.57	0.73	.58	.63	.37	-.07	.45	.47	.28	**.57**	.57
Total	11.91	3.41	N/A	.72	.32	.13	.41	.59	.60	.66	**.74**

*$p < .05$, **$p < .01$

1 = Base rates (BR), or the proportion of students within the sample that were found to possess a social skill deficit per each ISP-SS item. A student was scored as possessing a social skill deficit if the corresponding item rating was ≤ 1 (on a 0-3 four-point Likert scale).

Note. Bolded values represent the correlation between each ISP-SS item and its corresponding SSiS-RS domain. Comm = Communication, Coop = Cooperation, Assert = Assertion, Resp = Responsibility, Emp = Empathy, Eng = Engagement, SC = Self-Control, and Total = Total SSiS-RS scale score ($M = 100$, $SD = 15$).

Table 2

Conditional probability statistics indicative of the diagnostic accuracy of Intervention Selection Profile – Social Skill (ISP-SS) item cut scores relative to Social Skills Improvement System – Rating Scales (SSiS-RS) subscales

ISP-SS Item	SSiS-RS Subscale	AUC (95% CI)	Cut	SE (95% CI)	SP (95% CI)	PPV	NPV
Elementary School							
1	Communication	.86 (.78-.95)	≤ 1	.81 (.66-.94)	.89 (.77-.97)	.87	.84
2	Cooperation	.75 (.63-.85)	≤ 1	.73 (.60-.85)	.70 (.52-.85)	.78	.63
3	Assertion	.82 (.72-.91)	≤ 1	.84 (.68-1.00)	.71 (.58-.83)	.53	.92
4	Responsibility	.77 (.67-.87)	≤ 1	.83 (.69-.97)	.68 (.53-.84)	.67	.84
5	Empathy	.76 (.64-.87)	≤ 1	.44 (.24-.64)	.91 (.81-.98)	.73	.73
6	Engagement	.88 (.81-.96)	≤ 1	.80 (.63-.93)	.89 (.78-.97)	.86	.85
7	Self-Control	.77 (.67-.88)	≤ 1	.72 (.55-.86)	.76 (.63-.89)	.70	.78
Middle School							
1	Communication	.82 (.74-.90)	≤ 1	.83 (.72-.93)	.81 (.68-.91)	.81	.83
2	Cooperation	.81 (.72-.90)	≤ 1	.74 (.64-.84)	.83 (.67-.96)	.93	.53
3	Assertion	.67 (.54-.79)	≤ 1	.52 (.29-.71)	.71 (.60-.80)	.34	.84
4	Responsibility	.48 (.37-.59)	≤ 1	1.00 (1.00-1.00)	.00 (.00-.00)	.69	.00
5	Empathy	.71 (.61-.81)	≤ 1	.40 (.27-.53)	.92 (.83-.98)	.82	.62
6	Engagement	.86 (.79-.93)	≤ 1	.85 (.74-.97)	.81 (.71-.92)	.73	.91
7	Self-Control	.73 (.64-.82)	≤ 1	.82 (.68-.92)	.58 (.45-.71)	.57	.82

Note. SE = sensitivity, SP = specificity, PPV = positive predictive value, NPV = negative predictive value, and CI = confidence interval

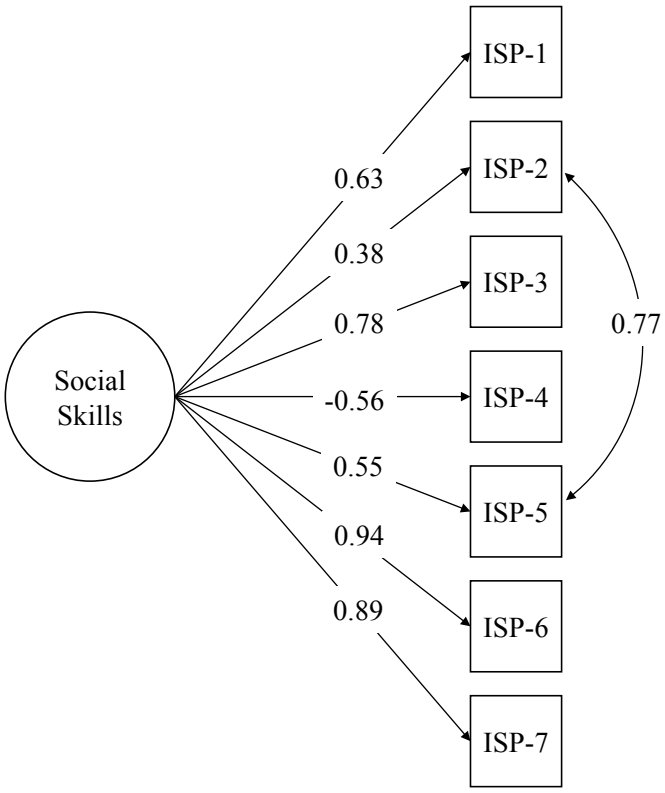

Figure 1. Unidimensional factor model with standardized pattern coefficients.