Psychological Assessment
2010, Vol. 22, No. 1, 157–166

1040-3590/10/$12.00 DOI: 10.1037/a0018124

Cross-Informant Agreement for Ratings for Social Skill and Problem Behavior Ratings: An Investigation of the Social Skills Improvement System—Rating Scales

Frank M. Gresham
Louisiana State University

Stephen N. Elliott
Vanderbilt University

Clayton R. Cook and Michael J. Vance
Louisiana State University

Ryan Kettler
Vanderbilt University

One of the most consistent findings in rating scale research with children and adolescents is the modest agreement among different informants' ratings. The present study systematically explored patterns of agreement among teachers, parents/caregivers, and students in domains of social skills and problem behaviors using the Social Skills Improvement System—Rating Scales (SSIS–RS; F. M. Gresham & S. N. Elliott, 2008). Two subsamples from the normative sample of the SSIS–RS were used. The first sample of participants consisted of 168 students who had all 3 informants (parent, teacher, and self) complete the SSIS–RS scales, which was necessary to assess agreement across different raters. The second sample consisted of 164 students who had raters in a similar or same role (father–mother, teacher–teacher). The results replicated an extensive literature showing that cross-informant agreements for social skills and problem behaviors are weak to moderate. The current study invoked multitrait–multimethod logic to interpret the correlations among raters derived from different informants and showed that the convergent validity coefficients were consistently stronger than the discriminant validity correlations. Implications for assessment practices and future research are discussed.

Keywords: convergent validity, discriminant validity, social skills

Children and youth having social competence deficits experience substantial difficulties in the development and maintenance of satisfactory interpersonal relationships, exhibition of prosocial behavior patterns, and social acceptance by peers and teachers (Elliott & Gresham, 2008; Gresham, 1997, 1998; Maag, 2005, 2006; Walker, Ramsay, & Gresham, 2004). These social competence deficits lead to short-term, intermediate, and, in many cases, long-term difficulties in areas of educational, psychosocial, and vocational domains of functioning (Kupersmidt, Coie, & Dodge, 1990; Newcomb, Bukowski, & Pattee, 1993; Parker & Asher, 1987). Empirical literature dating back to the 1950s consistently has shown that children and youths experiencing difficulty in interpersonal relationships were at risk for a variety of negative outcomes, such as school dropout, juvenile delinquency, adulthood psychopathology, depression, and suicide (Cowen, Pederson, Babigian, Izzo, & Trost, 1973; Kohn & Clausen, 1955; Kupersmidt, Coie, & Dodge, 1990; Parker & Asher, 1987).

Frank M. Gresham, Clayton R. Cook, and Michael J. Vance, Department of Psychology, Louisiana State University; Stephen N. Elliott and Ryan Kettler, Department of Special Education, Vanderbilt University.

This research was supported by Institute of Educational Sciences, U.S. Department of Education Grants R324A090098 and R324A080113 to Frank M. Gresham.

Correspondence concerning this article should be addressed to Frank M. Gresham, Department of Psychology, Louisiana State University, Baton Rouge, LA 70803. E-mail: gresham@lsu.edu

Social competence deficits also are a major criterion in the current federal definition of emotional disturbance specified in the Individuals with Disabilities Education Improvement Act (2004; IDEIA). In fact, two of the five criteria described in the current definition indicate social competence deficits as part of the disability: (a) "an inability to build or maintain satisfactory interpersonal relationships with peers and teachers" and (b) "the expression of inappropriate behavior under normal circumstances." Social competence or interpersonal difficulties are also part of the diagnostic criteria specified in the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; American Psychiatric Association, 2000). Given these deficits in social competence of children with or at risk for emotional and behavioral disorders, this class of behavior represents an important focus of intervention efforts and appropriate measurement of intervention effects.

An important distinction in the theoretical conceptualization of social behavior is the distinction between the concepts of *social skill* and *social competence* (Gresham, 1998; McFall, 1982). *Social skills* are a specific class of behaviors that an individual exhibits to successfully complete a social task. Social tasks might include such things as peer group entry, initiating and sustaining a conversation, making friends, playing a game with peers, and so forth. It should be noted that social tasks require several interconnected and discrete forms of social skilled behaviors. *Social competence*, in contrast, is an evaluative term based on judgments (given certain criteria) that an individual has performed a social task adequately. These judgments are made by social agents with whom the indi-

vidual interacts in naturalistic environments (e.g., school, home, and community). Given this conceptualization, social skills are specific behaviors exhibited in specific situations that lead to judgments by others that these behaviors were competent or incompetent in accomplishing specific social tasks (Gresham & Elliott, 2008).

Advantages of Behavior Rating Scales

Behavior rating scales are among the most frequently used measures of social behavior in school and clinic settings. Unlike systematic direct observations, behavior rating scales are considered *indirect* measures of behavior, because they are not measuring behavior at the time and place of its actual occurrence. Instead, behavior ratings require the rater to retrospectively rate the occurrence of behavior (Gresham & Lambros, 1998). Behavior rating scales have several advantages, including the following: (a) information is quantifiable and amenable to reliability and validity analyses, (b) they assess a broad range of behavior (e.g., social skills and problem behaviors), (c) multiple raters can be used to assess social behavior from multiple perspectives (teachers, parents, students), and (d) normative data provide a standard for judging the severity of behavior by comparing an individual with representative samples of other individuals (Gresham & Elliott, 2008; McConaughy & Ritter, 2005).

Behavior rating scales enjoy a long history of empirical demonstration of reliability and validity evidence. Some of the most common broad-band behavior ratings scales are well established and have outstanding technical adequacy. These measures include the Behavioral Assessment System for Children—2 (BASC–2, Reynolds & Kamphaus, 2004), Achenbach System of Empirically Based Assessment (ASEBA; Achenbach & Rescorla, 2001), Conners Rating Scales—Revised (Conners, 1997), and Social Skills Rating System (SSRS; Gresham & Elliott, 1990). A review of this literature is far beyond the scope of this article, but it is sufficient to conclude that these measures are technically adequate and widely used by professionals in the field.

Behavior Rating Scales and Cross-Informant Agreement

One of the most consistent findings in rating scale research with children and adolescents is the modest agreement among different informants' ratings. In a classic meta-analysis of different informants' ratings of social, emotional, and behavioral problems, Achenbach, McConaughy, and Howell (1987) showed the average Pearson *r* among parents, teachers, children, and mental health workers was about .20. Since this meta-analysis, numerous studies have documented that low interrater agreement among informants is a robust finding across various domains of child and adolescent psychopathology (Kraemer et al., 2003; Offord et al., 1996; Reynolds & Kamphaus, 2004; Youngstrom, Loeber, & Stouthamer-Loeber, 2000), adaptive behavior (Harrison & Oakland, 2003; Sparrow, Cichetti, & Balla, 2005), and social competence (Gresham & Elliott, 1990; Renk & Phares, 2004).

Informant discrepancies in rating scales have substantial practical implications for psychologists and other professionals conducting assessments of social–behavioral functioning. Several considerations should be entertained when using rating scales data collected from different informants. First, there is no gold standard against which to compare divergent ratings from various informants (De Los Reyes & Kazdin, 2005; Richters, 1992). Second, strategies to reduce discrepancies using confrontational tactics to "force" concordant ratings have met with limited success (Angold et al., 1987; Nguyen et al., 1994). Third, methods to reduce discrepancies by averaging informants' ratings create substantial statistical and interpretive problems (Bird, Gould, & Staghezza, 1992; Braaten et al., 2001).

Purpose of the Present Study

The purpose of the present study was to systematically explore patterns of agreement among teachers, parents/caregivers, and students in domains of social skills and problem behaviors. This investigation used the revision of the widely used SSRS (Gresham & Elliott, 1990), now known as the Social Skills Improvement System—Rating Scales (SSIS–RS; Gresham & Elliott, 2008). The SSIS–RS is a multirater (teacher, parent/caregiver, and student) series of rating scales that documents the frequency of social skills and competing problem behaviors (see details in the Instrumentation section below). The following hypotheses were specifically tested in this investigation:

1. Pairs of informants (teacher–parent, teacher–student, and parent–student) will exhibit greater than chance levels of agreements as indexed by significant interrater Pearson *r* correlations. These correlations were expected to be greater than those reported by Achenbach et al.'s (1987) meta-analysis and Gresham and Elliott's (1990) SSRS data based on a greater number of common items across the three rater forms of the SSIS–RS.

2. Teacher–parent informants will show higher Pearson *r* correlations than teacher–student and parent–student pairs across social skills and problem behavior domains.

3. Pairs of similar informants (teacher–teacher and parent–parent) will show higher correlations than pairs of dissimilar informants (teacher–parent).

Method

Participants

Two subsamples from the standardization sample of the SSIS–RS were used. The first sample of participants consisted of 168 students who had all three informants (parent, teacher, and self) complete the SSIS–RS scales, which was necessary to assess agreement across different raters. Table 1 depicts domains measured by the SSIS–RS along with sample items and reliability estimates. Table 2 depicts the demographic information for these 168 participants. The average age of the participants was 11.9, and the majority of participants were male (63%). Analyses were performed to examine whether these 168 participants were significantly different than the other participants included in the norm-referenced sample. Results indicated that the sample for this study was representative of the national sample used to develop the norms for the SSIS–RS forms; in other words, there were no

Table 1
Descriptions of SSIS Subscale Items and Related Statistics

Scale/subscale (number of items)	Example items	α[a]		
		Teacher	Parent	Student
Social Skills		.97	.96	.95
Communication (7)	Takes turn in conversations. Makes eye contact when talking.	.86	.77	.79
Cooperation (6)	Follows your directions. Participates appropriately in class.	.90	.85	.81
Assertion (7)	Ask for help from adults. Expresses feelings when wronged.	.86	.78	.75
Responsibility (6)	Is well-behaved when unsupervised. Takes care when using other people's things.	.90	.86	.77
Empathy (6)	Shows concern for others. Feels bad when others are sad.	.91	.87	.82
Engagement (7)	Interacts well with other children. Invites others to join in activities.	.88	.85	.76
Self-Control (7)	Stays calm when teased. Respond appropriately when pushed or hit.	.91	.84	.81
Problem Behaviors		.95	.94	.93
Externalizing (12)	Disobeys rules or requests. Fights with others.	.93	.90	.88
Bullying (5)	Forces others to act against their will. Keeps others out of social circles.	.76	.80	.81
Hyperactivity/Inattention (7)	Fidgets or moves around too much. Gets distracted easily.	.90	.85	.83
Internalizing (7)	Withdraws from others. Acts anxious with others.	.86	.85	.85
Autism Spectrum (7[b])	Is preoccupied with object parts. Repeats the same thing over and over.	.88	.85	

Note. SSIS = Social Skills Improvement System (Gresham & Elliott, 2008).
[a] Mean coefficient alpha across all age levels. [b] Note that there are also eight Social Skill items that contribute to the total Autism Spectrum subscale.

significant statistical differences in terms of gender, grade, or age between this subsample and the entire normative sample.

The second sample of participants consisted of 164 students who had similar raters complete the SSIS–RS (i.e., teacher–teacher and parent–parent), which was necessary to assess the agreement between similar raters. Table 3 depicts the demographic information for these participants. The average age of the participants was 9 years, and the majority of participants were male and of White racial background. Similar to the other subsample, results indicated that the sample for this study was representative of the national sample used to develop the norms for the SSIS–RS forms.

Instrumentation

The SSIS–RS is a multirater series of rating scales that includes ratings from teachers, parents, and students. The SSIS–RS assesses three domains of social skills, problem behaviors, and academic competence. All forms include common social skills in the following subdomains: communication, cooperation, assertion, responsibility, empathy, engagement, and self-control. The Problem Behavior subscales include the subdomains of: externalizing, bullying, hyperactivity/inattention, internalizing, and autism spectrum. The teacher form of the SSIS–RS includes an Academic Competence scale, on which teachers rate student performance in reading, math, motivation, parental support, and general cognitive functioning.

Teachers and parents indicate the frequency with which the student exhibits each social skill and problem behavior on a 4-point scale of *never, seldom, often,* and *almost always*. Students indicate how true a statement is about each social skill and problem behavior for them using a 4-point scale of *not true, a little*

Table 2
Demographic Characteristics for Subsample With Dissimilar Raters

Variable	Frequency	%
Gender		
Male	106	63
Female	62	37
Grade		
3rd	28	17
4th	19	11
5th	27	16
6th	27	16
7th	16	10
8th	12	7
9th	7	4
10th	5	5
11th	14	8
12th	13	8
Total	168	100

Table 3
Demographic Characteristics of Interrater Reliability Subsample With Similar Raters

Characteristic	Teacher form	Parent form
N	54	110
Age		
M (years:months)	8:7	9:0
SD (months)	40.1	49.6
Gender		
Female	32	44
Male	22	66
Race/ethnicity		
African American	11	10
Hispanic	12	8
White	25	87
Other	6	5
Mother's education		
Grade 11 or less	7	5
Grade 12 or GED	24	28
1–3 years of college	13	46
4 or more years of college	10	31

Note. GED = General Equivalency Diploma.

true, a lot true, and *very true*. In addition, teachers, parents, and students indicate the importance of each social skill to the students' development and classroom success using a 3-point scale of *not important, important,* and *critical*. These importance ratings were not part of the current investigation. Teachers also indicate the student's level of academic competence in reading, math, motivation, parental support, and general cognitive functioning using a 5-point scale of *lowest 10%, next lowest 20%, middle 40%, next highest 20%,* and *highest 10%*.

Data Analytic Strategies

Test–retest reliability estimates (over 42 to 66 days) were .82, .84, and .81 for the Total Social Skills scores for teachers, parents, and students, respectively. Test–retest reliability estimates for the Total Problem behavior scores were .92, .86, and .77 for teachers, parents, and students, respectively. Test-reliability estimates for the Social Skills and Problem Behavior subscales were mostly in the .80s, with the median stability coefficient of around .84 (Gresham & Elliott, 2008).

Validity evidence for the scores obtained from the SSIS–RS has been demonstrated by correlational studies with other widely used instruments such as the Behavioral Assessment System (2nd ed.; BASC–2; Reynolds & Kamphaus, 2004), the SSRS (Gresham & Elliott, 1990), and the Vineland Adaptive Behavior Scale (2nd ed.; Vineland II; Sparrow et al., 2005). Overall, the SSIS–RS show moderate to high correlations (depending on the scale and subscale) with each of these instruments (see Gresham & Elliott, 2008, for more detail). Finally, the SSIS–RS has been shown to differentiate members of special populations such as attention-deficit/hyperactivity disorder, autism spectrum disorder, developmental delay, emotional/behavioral disturbance, intellectual disability, and speech/language impairment (see Gresham & Elliott, 2008).

The present study examined interrater agreement among the three rater dyads (teacher–parent, parent–student, and teacher–student) with two indices. The first method employed to assess interrater agreement was the calculation of *q* correlations among pairs of raters (teacher–parent, teacher–student, and parent–student). The *q* correlations are Pearson correlations between sets of common items provided by different raters or informants. Specifically, bivariate *q* correlations were calculated across raters to examine the convergent and divergent relationships for evidence of interrater reliability and are expressed as Pearson *r*. Convergent correlations were the interrater reliability estimates and were those that represented the correlation between raters' ratings of the same subscale (e.g., Total Social Skills—Parent with Total Social Skills—Teacher), whereas divergent correlations were those that represented correlations between raters on different subscales (e.g., Total Social Skills—Parent with Total Problem Behaviors—Teacher). Consistent with multitrait–multimethod matrix (Campbell & Fiske, 1959), one would anticipate the convergent correlations to be higher in magnitude than the divergent correlations.

Given the fact that multiple correlations were calculated, the *p* value was set at 0.01 to maintain the Type I error rate at an acceptable level. This alpha level was selected because given the number of possible comparisons; a Bonferroni correction would be far too conservative and overcorrect the appropriate alpha level thereby creating an inordinately high Type II error rate (see Rosenthal & Rubin, 1984, for a discussion). Therefore, to consider a correlation statistically significant it had to be associated with a $p < .01$.

The second method consisted of calculating standardized mean difference effect sizes representing the discrepancy between raters in their ratings of social skills and problem behaviors. Effect sizes (ES) were calculated by dividing the absolute value of the *difference score* between raters' standard scores by the pooled standard deviation.

$$\mathrm{ES} = \frac{D}{\left(\frac{SD_1 + SD_2}{2}\right)}. \quad (1)$$

The effect size captures the extent the dyad agreed about the overall level of social skills or problem behaviors in standard deviation units. Effect sizes were interpreted using Cohen's (1988) conventions for small ($d = 0.20$), medium ($d = 0.50$), and large ($d = 0.80$). Therefore, effect sizes approaching less than 0.20 indicated high agreement between raters, whereas effect sizes reaching 0.80 indicated high disagreement. Effect sizes are amenable to arbitrarily assigning the sign or directionality of the effect depending on an interpretation that is consistent with the finding. We assigned positive values to effect sizes to indicate that the adult (parent or teacher) provided less favorable ratings of social skills or provided less favorable ratings of problem behaviors than did the student. For the teacher–parent dyad, positive values were used to indicate that teachers provided less favorable ratings on social skills and problem behaviors than parents. For example, an effect size of 0.15 representing the agreement between teachers and students on the Social Skills Total Score would indicate that teachers provided slightly less favorable ratings of social skills than did students; however, there was high agreement between their ratings.

For the purposes of comparing correlations across raters, scales, and type (e.g., convergent vs. divergent) of correlations, averages were calculated. Because distances between correlations are not equally detectable (Rosenthal & Rosnow, 2008), prior to summation and averaging, the correlations were transformed and placed on a common metric. The Fisher's *Zr* formula was used to transform the correlations and compute an average. The following formula was used:

$$r' = (0.5)\log_e\left|\frac{1 + r}{1 - r}\right|. \quad (2)$$

Once the average *Zr* was calculated, we used the Fisher inverse formula to transform the *Zr* back to *r*.

$$r = \frac{e^{2ES_{Zr}} - 1}{e^{2ES_{Zr}} + 1}. \quad (3)$$

Results

Hypotheses 1 and 2

Parent–teacher agreement.

Social skills. The convergent and divergent correlations for teachers' and parents' ratings on the Social Skill and Problem Behavior Scales are displayed in Tables 4 and 5, respectively. The

Table 4
Teacher and Parent Correlations for Social Skill Subscales and Total Score

	Teacher							
Parent	1	2	3	4	5	6	7	8
1. Communication	**0.28***	0.22*	0.18*	0.27*	0.27*	0.36*	0.23*	0.33*
2. Cooperation	0.26*	**0.28***	0.12	0.33*	0.23*	0.25*	0.24*	0.31*
3. Assertion	0.04	−0.02	**0.15**	0.04	0.11	0.17	−0.10	0.07
4. Responsibility	0.29*	0.29*	0.08	**0.38***	0.32*	0.24*	0.27*	0.34*
5. Empathy	0.17	0.09	0.13	0.15	**0.26***	0.23*	0.10	0.20*
6. Engagement	0.12	0.03	0.18*	0.07	0.16	**0.34***	0.01	0.16
7. Self-control	0.25*	0.28*	0.15	0.33*	0.24*	0.28*	**0.23***	0.32*
8. Total SS	0.24*	0.20*	0.18*	0.26*	0.28*	0.33*	0.17	**0.30***

Note. SS = social skills. Boldface indicates convergent validity coefficients.
* $p < .01$.

convergent correlations representing the convergent validity indices for the Social Skill Subscales are located on the diagonal of the correlation matrix in Table 4. These correlations ranged from a minimum of 0.15 to a maximum of 0.38. The only subscale with a nonsignificant correlation was the Assertion scale. The results indicated that parents and teachers agreed the most on ratings of the Responsibility and Engagement subscales and least on the subscales of Assertion and Self-Control. Comparison of the median convergent correlation ($r = .285$) to the average divergent correlation ($r = .20$) indicated that the convergent correlations were generally stronger in magnitude than the divergent correlations. This difference provides modest support for the convergent validity of the SSIS–RS in that ratings of the same social skills constructs by different raters were found to have somewhat stronger associations than ratings of different social skills constructs by different raters. The parents' and teachers' Social Skills Total Scores correlated significantly with one another ($r = .30$) and with all but one of the subscales (Assertion) from the other rater.

The standardized mean difference effect size for the parent–teacher dyad was 0.24, indicating that teachers provided slightly less favorable ratings than parents (see Table 6). However, according to Cohen's guidelines, the effect size was small indicating modest agreement between parents and teachers on their ratings of social skills.

Problem behaviors. For the Problem Behavior Subscales, the convergent validity estimates were all statistically significant and ranged from a minimum of 0.18 to a maximum of 0.39 (see Table 5). Results indicated that parents and teachers agreed the most on the Externalizing and Hyperactivity scales and least on the Internalizing and Bullying scales. The average convergent validity estimate ($r = .33$) was larger than the average divergent correlation ($r = .25$). This finding indicated that raters agreed more when rating the same problem behavior constructs than different constructs. The parents' and teachers' Problem Behavior Total Score correlated significantly with one another ($r = .31$) and with all of the subscales from the other rater. Comparison of the average convergent correlations for the Social Skills Total Score ($r = .30$) and Problem Behavior Total Score ($r = .33$) indicated that parents and teachers tended to agree slightly more on their ratings of problem behaviors than social skills.

The standardized mean difference effect size for the Problem Behavior Total Score was 0.03, indicating that teachers and parents provided approximately equal overall ratings of problem behaviors (see Table 6). This effect size also indicated that neither teachers nor parents provided less favorable ratings of problem behaviors than the other.

Parent–student agreement.

Social skills. The convergent and divergent correlations for parents' and students' ratings on the Social Skill and Problem Behavior Scales are displayed in Tables 7 and 8, respectively. For the Social Skills Subscales, the convergent validity indices on the diagonal of the correlation matrix ranged from a minimum of 0.02 to a maximum of 0.34. These correlations were weaker than the interrater reliability estimates derived from the parent–teacher

Table 5
Teacher and Parent Correlations for Problem Behavior Subscales and Total Score

	Teacher					
Parent	1	2	3	4	5	6
1. Externalizing	**0.39***	0.29*	0.33*	0.22*	0.37*	0.36*
2. Bullying	0.37*	**0.33***	0.27*	0.19*	0.31	0.33*
3. Hyperactivity	0.34*	0.20*	**0.39***	0.22*	0.36*	0.35*
4. Internalizing	0.12	0.03	0.15	**0.18***	0.23*	0.17
5. Autism	0.23*	0.15	0.25*	0.20*	**0.38***	0.25*
6. Total PB	0.30*	0.20*	0.29*	0.21*	0.34*	**0.31***

Note. PB = problem behavior. Boldface indicates convergent validity coefficients.
* $p < .01$.

Table 6
Standardized Mean Difference Effect Size for Different Between Raters' Score on Total Social Skills and Total Problem Behavior

Rater dyad	M SS$_1$	SD_1	M SS$_2$	SD_2	/D/	ES
Parent–Teacher						
Social Skills	99	18	94	19	5	0.24
Problem Behaviors	100	15	99	13	1	0.03
Teacher–Student						
Social Skills	94	19	97	21	3	0.15
Problem Behaviors	99	13	99	14	0	0.00
Parent–Student						
Social Skills	99	18	97	21	2	−0.10
Problem Behaviors	100	15	99	14	1	−0.07

Note. SS = social skills; |D| = absolute difference score; ES = effect size.
[1] Mean standard score and standard deviation for first rater. [2] Mean standard score and standard deviation for second rater.

ratings. The Social Skills Subscales with the highest interrater reliability estimates were Responsibility and Engagement. The Social Skills Subscales with the lowest convergent validity correlations were Communication and Engagement. The average of the convergent correlations (r = .19) was larger in magnitude than the average of the divergent correlations (r = .13), indicating that parents and students provider better agreement when rating the same social skills construct than different constructs. Although the overall convergent validity estimates were weak, on average, they were still stronger than the divergent correlations.

The standardized mean difference effect size for the parent–student dyad was −0.10, indicating that students provided slightly less favorable ratings of their social skills than parents (see Table 6). According to Cohen's guidelines, the effect size was small, indicating relatively high agreement between parents and students on their ratings of social skills.

Problem behaviors. With regard to the Problem Behavior Subscales, the convergent validity estimates ranged from 0.21 to 0.36 (see Table 8). Results revealed that the subscales with the strongest estimates were Externalizing Behavior and Bullying, whereas the subscales with the weakest estimates were Internalizing behavior and Hyperactivity. Comparison of the average of the convergent correlations (r = .29) to the average of the divergent correlations (r = .22) indicated that raters tended to agree more when providing ratings of the same problem behavior constructs than different constructs. Parents' and students' Total Problem Behavior Scores correlated significantly with one another and with all of the subscales from the other rater. The average convergent correlations for the Social Skills and Problem Behavior Total Scores were 0.19 and 0.29, respectively. Comparison of these correlations indicated that there was greater agreement between parents and students on their ratings of problem behaviors than social skills.

The standardized mean difference effect size for the Problem Behavior Total Score was −0.07, indicating that students provided slightly less favorable ratings of their problem behaviors than teachers (see Table 6). According to Cohen's guidelines, the effect size is small and indicates that although students provided less favorable ratings of themselves, there was high agreement between their ratings of problem behaviors and those from parents.

Teacher–student agreement.

Social skills. The correlations for teachers' and students' ratings of Social Skills and Problem Behavior Scales are displayed in Tables 9 and 10, respectively. The convergent validity indices for the Social Skill Subscales ranged in magnitude from 0.12 to 0.31. The subscales of Communication and Empathy were associated with nonsignificant correlations, indicating weak convergent validity between teachers and students on these dimensions. The strongest correlations were noted between teachers' and students' ratings for Responsibility and Engagement subscales. The average convergent correlation (r = .21) was stronger than the average divergent correlation (r = .14). This difference indicated that there

Table 7
Parent and Student Correlations for Social Skill Subscales and Total Score

	Student							
Parent	1	2	3	4	5	6	7	8
1. Communication	**0.19***	0.22*	0.11	0.25*	0.16	0.09	0.17	0.21*
2. Cooperation	0.17	**0.34***	0.08	0.27*	0.19*	0.09	0.25*	0.24*
3. Assertion	0.06	0.04	**0.02**	0.09	−0.03	0.01	0.06	0.04
4. Responsibility	0.21*	0.30*	0.08	**0.26***	0.19*	0.09	0.23*	0.24*
5. Empathy	0.10	0.10	0.02	0.16	**0.08**	0.01	0.10	0.10
6. Engagement	0.13	0.07	0.05	0.16	0.06	**0.16**	0.08	0.12
7. Self-control	0.16	0.30*	0.08	0.22*	0.18*	0.11	**0.25***	0.22*
8. Total SS	0.18*	0.25*	0.08	0.25*	0.15	0.10	0.21*	**0.21***

Note. SS = social skills. Boldface indicates convergent validity coefficients.
* $p < .01$.

Table 8
Parent and Student Correlations for Problem Behavior Subscales and Total Score

	Student[a]				
Parent	1	2	3	4	6
1. Externalizing	**0.34***	0.25*	0.34*	0.23*	0.32*
2. Bullying	0.29*	**0.24***	0.24*	0.16	0.27*
3. Hyperactivity	0.28*	0.19*	**0.36***	0.23*	0.28*
4. Internalizing	0.13	0.04	0.20*	**0.21***	0.18*
5. Autism	0.19*	0.16	0.26*	0.22*	0.25*
6. Total PB	0.26*	0.16	0.29*	0.25*	**0.30***

Note. PB = problem behavior. Boldface indicates convergent validity coefficients.
[a] There is no autism subscale on the self-report version of the Social Skills Improvement System—Rating Scales (Gresham & Elliott, 2008).
* $p < .01$.

Table 10
Teacher and Student Correlations for Problem Behavior Subscales and Total Score

	Student self-report[a]				
Teacher report	1	2	3	4	6
1. Externalizing	**0.37***	0.36*	0.36*	0.23*	0.37*
2. Bullying	0.40*	**0.40***	0.34*	0.22*	0.39*
3. Hyperactivity	0.31*	0.30*	**0.34***	0.18*	0.30*
4. Internalizing	0.22*	0.19*	0.22*	**0.14**	0.22*
5. Autism	0.29*	0.29*	0.29*	0.15	0.28*
6. Total PB	0.33*	0.31*	0.31*	0.20*	**0.33***

Note. PB = problem behavior. Boldface indicates convergent validity coefficients.
[a] There is no autism subscale on the self-report version of the Social Skills Improvement System—Rating Scales (Gresham & Elliott, 2008).
* $p < .01$.

was slightly greater agreement between teachers and students on their ratings of the same social skills constructs than different constructs.

The standardized mean difference effect size for the parent–teacher dyad was 0.15, indicating that teachers provided slightly less favorable ratings than students (see Table 6). According to Cohen's guidelines, the effect size was small, indicating relatively high agreement between teachers and students on their ratings of social skills.

Problem behaviors. The convergent validity estimates for the Problem Behavior Scales ranged in magnitude from a minimum of 0.14 to a maximum of 0.40. All but one of the convergent indices (Internalizing Behaviors) was statistically significant. The subscales with the strongest convergent validity estimates were Externalizing Behaviors and Bullying. Comparison of the average convergent correlation ($r = .32$) to the average divergent correlation ($r = .29$), indicating that teachers and students agreed more on their ratings of the same problem behavior constructs than different ones. The average convergent correlations for the Social Skills and Problem Behavior Total Scores were 0.21 and 0.32, respectively.

The standardized mean difference effect size for the Problem Behavior Total Score was 0.00, indicating that teachers and students provided equal ratings of overall problem behaviors (see Table 6). The sign of the effect size indicated that neither teachers nor students provided less favorable ratings of problem behaviors than the other.

Hypothesis 3

One hypothesis we had going into this study was that pairs of similar informants (teacher–teacher and parent–parent) would show higher correlations than pairs of dissimilar informants (teacher–parent). Table 11 portrays the results of this comparison. As predicted, teacher–teacher and parent–parent dyads demonstrated stronger agreement than did teacher–parent dyads. Across all Social Skill and Problem Behavior scales and subscales, similar informants showed higher interrater reliability estimates than dissimilar informants. The average interrater correlations for teacher–teacher, parent–parent, and teacher–parent dyads were 0.58, 0.55, and 0.30, respectively. These average correlations were consistent with Hypothesis 3. One can see by looking at Table 11 that similar informants tended to agree more on the Social Skills subscales, whereas dissimilar informants tended to agree more on the Problem Behavior subscales. Table 12 depicts highest convergent validity estimates by domain and rater combination.

Table 9
Teacher and Student Correlations for Social Skill Subscales and Total Score

	Student self-report							
Teacher report	1	2	3	4	5	6	7	8
1. Communication	**0.12**	0.14	0.11	0.20*	0.11	0.14	0.17	0.17
2. Cooperation	0.17	**0.24***	0.17	0.32*	0.13	0.12	0.18*	0.23*
3. Assertion	0.03	−0.01	**0.20***	0.15	0.06	0.19*	−0.01	0.11
4. Responsibility	0.13	0.18*	0.11	**0.26***	0.08	0.03	0.12	0.16
5. Empathy	0.06	0.08	0.11	0.20*	**0.12**	0.14	0.10	0.14
6. Engagement	0.13	0.06	0.14	0.15	0.18*	**0.31***	0.08	0.18*
7. Self-control	0.13	0.16	0.15	0.23*	0.13	0.17	**0.21***	0.20*
8. Total SS	0.13	0.15	0.18*	0.27*	0.14	0.20*	0.15	**0.21***

Note. SS = social skills. Boldface indicates convergent validity coefficients.
* $p < .01$.

Table 11
Interrater Reliability Correlations Between Similar and Dissimilar Informants

Scale and subscale	Rater dyads		
	Teacher–teacher	Parent–parent	Teacher–parent
Social Skills	.70	.62	.30
Communication	.63	.62	.28
Cooperation	.60	.67	.28
Assertion	.38	.35	.15
Responsibility	.54	.70	.38
Empathy	.55	.51	.26
Engagement	.71	.54	.34
Self-Control	.62	.62	.23
Problem Behaviors	.57	.47	.31
Externalizing	.53	.57	.39
Bullying	.46	.37	.33
Hyperactivity/Inattention	.56	.58	.39
Internalizing	.50	.39	.18
Autism Spectrum	.69	.58	.38
Average rater dyad *r*	**.58**	**.55**	**.30**

Discussion

The current investigation replicates an extensive literature showing that cross-informant agreements for problem behaviors and child/adolescent psychopathology are weak to moderate (Achenbach et al., 1987; Bird et al., 1992; De Los Reyes & Kazdin, 2005; Kraemer et al., 2003; Youngstrom et al., 2000). This study extends these previous findings to informant ratings of social skills. Our findings are entirely consistent with the meta-analytic results presented by Renk and Phares (2004). Effect size estimates from this meta-analysis of 74 studies showed an average parent–teacher correlation of .38, a mean self–teacher correlation of .25, and a mean self–parent correlation of .21 Correlation contrasts (using the test for independent correlations) between the Renk and Phares meta-analysis and the current study were statistically equivalent (*Mdn* $p = .425$). Thus, the current data indicate that teachers, parents, and students do not perceive their level of social skills differently, a finding that is highly consistent with the literature on cross-informant agreement on ratings of children and adolescent social skills (Renk & Phares, 2004). Our findings in this study showed a parent–teacher correlation of .30, a student–teacher correlation of .21, and a student–parent correlation of .21. Comparisons of these correlations using a test for dependent correlations show that the parent–teacher ($p = .19$), parent–student ($p = .50$), and the teacher–student ($p = .50$) correlations were not statistically different.

One important feature of rating scales is the extent to which informants or raters agree with regard to the ratings of the same construct. The concept of convergent validity stipulates that ratings of the same construct by different raters should be stronger in magnitude than ratings of different constructs by different raters. As such, applying multitrait–multimethod logic to examine the correlations among ratings derived from different raters can be used to assess the convergent validity of the SSIS–RS (Campbell & Fiske, 1959). Across different rater pairings, our results revealed that the convergent validity coefficients were consistently stronger than the divergent (discriminant) validity correlations. Overall, these differences, however, were relatively small across raters, scales, and subscales and should not be overinterpreted.

The rater combination and type of subscale both were found to impact cross-informant agreement estimates. With respect to rater pairs, somewhat stronger convergent validity estimates were found for teacher–parent ratings, and the weakest estimates were found for parent–student ratings, particularly for the social skills subscales. As indicated earlier, these convergent validity coefficients were not statistically different. One possible reason for the lower parent–student and teacher–student correlations may be differences in scaling. The student scale used rating anchors indicating self-perceptions, whereas parent and teacher raters used a frequency rating. This should be investigated in future research. The Problem Behavior Scale and subscales were consistently associated with better agreement across rater dyads. We also replicated previous literature showing concordance rates are higher for externalizing than for internalizing behaviors (Achenbach et al., 1987; Comer & Kendall, 2004; Duhig, Renk, Epstein, & Phares, 2000). The magnitude of these correlations for externalizing and internalizing behaviors is statistically similar across rater pairs (.39 and .18 for parent–teacher, .34 and .21 for parent–child, and .37 and .14 for self–parent).

We also showed that when raters share similar environments, the correlations between informants increase dramatically. For example, teacher–teacher correlations for Social Skills and Problem behaviors were .68 and .61, respectively. Parent–parent correlations were of a similar magnitude of .62 and .50, respectively. These findings are consistent with the conclusions of Achenbach et al. (1987) concerning the situational specificity of behavior. The

Table 12
Comparison of Social Skills and Problem Behavior Scales With Strongest and Weakest Agreement

Raters	Social Skills interrater reliability		Problem Behavior interrater reliability	
	Strongest	Weakest	Strongest	Weakest
Teacher–Parent	Responsibility Engagement	Assertion Self-Control	Externalizing Hyperactivity	Internalizing Bullying
Parent–Student	Responsibility Cooperation	Assertion Empathy	Externalizing Hyperactivity	Internalizing Bullying
Teacher–Student	Responsibility Engagement	Communication Empathy	Externalizing Bullying	Internalizing Hyperactivity

current findings are also consistent with those of De Los Reyes and Kazdin (2005), who suggested that context is often not taken into account when rating behaviors. That is, teachers and parents may be rating perceived dispositional qualities rather than the contextualized qualities surrounding specific behaviors. To this end, De Los Reyes and Kazdin involved the notion of the actor–observer phenomenon (Jones & Nisbett, 1972), in which observers of another individual's behavior often attribute the cause of the behavior to the person's disposition (internal qualities) rather than the context or environment surrounding the behavior.

The current findings revealed that raters provided more consistent ratings for certain subscales than for others. For example, different rater combinations showed the strongest agreement estimates for the Externalizing and Bullying Problem Behavior subscales, whereas the weakest estimates were found for the Internalizing subscale. On the Social Skills Scale, parents tended to give the highest ratings, followed by students, and then by teachers. This pattern could be attributed to situational specificity, with students evaluating their skills in both school and home settings, whereas teachers and parents only interact with children in one or the other setting.

One consideration that is not addressed by the current data is the notion that there may be relatively higher levels of agreement among raters when rating behavior at the item level. For example, one might expect higher levels of agreement between teachers and parents when rating the items "fidgets or moves around too much" or "talks back to adults." One might expect lower levels of agreement among raters when rating items such as "acts lonely" or "has low energy." Similar levels of higher and lower agreement might be found when rating social skills items, such as "invites others to join activities" versus "tries to comfort others." Higher levels of agreement should be expected for items that are more directly *observable* than when rating items that require more inference on the part of the rater. This is a fruitful area for future research on cross-informant agreement.

The current study revealed that similar informants (e.g., teacher–teacher and parent–parent) provided higher agreement on the Social Skills subscales, whereas dissimilar informants (teacher–parent) provided higher agreement on the Problem Behavior subscales. A possible explanation for this difference is that social skills demonstrate greater situational specificity than do problem behaviors. That is, social skills are highly context specific, and the social skills necessary to function successfully at home differ from those required at school. This notion has been discussed extensively in the social skills literature (see DuPaul & Eckert, 1994; Elliott & Gresham, 2008; Gresham, 1981, 1998).

The implications of the current results for assessment practice suggest that discrepant ratings can be at least partially resolved by further assessment strategies that contextualize the ratings. Assessment techniques, such as functional assessment interviews with teachers, parents, and students as well as systematic direct observations assessing the antecedents and consequences of social behavior, would be informative. Future research investigating the underlying mechanisms or factors creating informant discrepancies should be pursued. The most fundamental conclusion is that there is no universal gold standard in the assessment of child and adolescent social skills.

References

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

Achenbach, T., & Rescorla, L. (2001). *Manual for the ASEBA school-age forms and profiles.* Burlington, VT: University of Vermont Research Center for Children, Youth, and Families.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Angold, A., Weissman, M., John, K., Merikangas, K., Prusoff, B., Wickramaratne, P., . . . Warner, V. (1987). Parent and child reports of depressive symptoms in children at low and high risk of depression. *Journal of the American Academy of Child & Adolescent Psychiatry, 28,* 901–915.

Bird, H., Gould, M., & Staghezza, B. (1992). Aggregating data from multiple informants in child psychiatry epidemiological research. *Journal of the American Academy of Child & Adolescent Psychiatry, 31,* 78–85.

Braaten, E., Bidereman, J., DiMauro, A., Mick, E., Monuteaux, M., Muehl, K., & Faraone, S. V. (2001). Methodological complexities in the diagnosis of major depression in youth: An analysis of mother and youth self-reports. *Journal of Child and Adolescent Psychopharmacology, 11,* 395–407.

Campbell, D., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

Comer, J., & Kendall, P. (2004). A symptom-level examination of parent–child agreement in the diagnosis of anxious youth. *Journal of the American Academy of Child & Adolescent Psychiatry, 43,* 878–886.

Conners, C. K. (1997). *Conners Rating Scales—Revised user's manual.* Toronto, Ontario, Canada: Multi-Health Systems.

Cowen, E., Pederson, A., Babigian, H., Izzo, I., & Trost, M. (1973). Long-term follow-up of early detected vulnerable children. *Journal of Consulting and Clinical Psychology, 41,* 1400–1416.

De Los Reyes, A., & Kazdin, A. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131,* 483–509.

Duhig, A., Renk, K., Epstein, M., & Phares, V. (2000). Interparental agreement on internalizing, externalizing, and total behavior problems: A meta-analysis. *Clinical Psychology: Science and Practice, 7,* 435–453.

DuPaul, G., & Eckert, T. (1994). The effects of social skills curricula: Now you see them, now you don't. *School Psychology Quarterly, 9,* 113–132.

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

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

Gresham, F. M. (1997). Social skills. In G. Bear, K. Minke, & A. Thomas (Eds.), *Children's needs: Psychological perspectives* (2nd ed., pp. 515–526). Bethesda, MD: National Association of School Psychologists.

Gresham, F. M. (1998). Social skills training with children: Social learning and applied behavioral analytic approaches. In T. S. Watson & F. M. Gresham (Eds.), *Handbook of child behavior therapy* (pp. 475–497). New York, NY: Plenum Press.

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

Gresham, F. M., & Elliott, S. N. (2008). *Social Skills Improvement System—Rating Scales.* Minneapolis, MN: Pearson Assessments.

Gresham, F. M., & Lambros, K. M. (1998). Behavioral and functional

assessment. In T. S. Watson & F. M. Gresham (Eds.), *Handbook of child behavior therapy* (pp. 3–22). New York, NY: Plenum Press.

Harrison, P., & Oakland, T. (2003). *Adaptive Behavior Assessment System–II.* San Antonio, TX: The Psychological Corporation.

Individuals with Disabilities Education Improvement Act, Pub. L. No. 108–446, 20 U.S.C. § 1400 *et seq.* (2004).

Jones, E., & Nisbett, R. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. Jones, D. Kanouse, H. Kelly, R. Nisbett, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). Morristown, NJ: General Learning Press.

Kohn, M., & Clausen, J. (1955). Social isolation and schizophrenia. *American Sociological Review, 20,* 265–273.

Kraemer, H., Measelle, J., Ablow, J., Essex, M., Boyce, W., & Kupfer, D. (2003). A new approach for integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry, 160,* 1566–1577.

Kupersmidt, J., Coie, J., & Dodge, K. (1990). The role of peer relationships in the development of disorder. In S. Asher & J. Coie (Eds.), *Peer rejection in childhood* (pp. 274–308). New York, NY: Cambridge University Press.

Maag, J. (2005). Social skills training for students with emotional and behavioral disorders and learning disabilities: Problems, conclusions, and suggestions. *Exceptionality, 13,* 155–172.

Maag, J. (2006). Social skills training for students with emotional and behavioral disorders: A review of reviews. *Behavioral Disorders, 32,* 5–17.

McFall, R. (1982). A review and reformulation of the concept of social skills. *Behavioral Assessment, 4,* 1–35.

Newcomb, A., Bukowski, W., & Pattee, L. (1993). Children's peer relations: A meta-analytic review of popular, rejected, neglected, controversial, and average sociometric status. *Psychological Bulletin, 113,* 306–347.

Nguyen, N., Wittlesey, S., Schimecca, K., Digiamoco, D., Bui, B., Parsons, O., . . . Paddock, D. (1994). Parent–child agreement in prepubertal depression: Findings with a modified assessment method. *Journal of the American Academy of Child & Adolescent Psychiatry, 33,* 1275–1283.

Offord, D., Boyle, M., Racine, Y., Szatmari, P., Fleming, J., Sanford, M., & Lipman, E. L. (1996). Integrating assessment data from multiple informants. *Journal of the American Academy of Child & Adolescent Psychiatry, 35,* 1078–1085.

Parker, J., & Asher, S. (1987). Peer relations and later personal adjustment: Are low-accepted children at risk? *Psychological Bulletin, 102,* 357–389.

Renk, K., & Phares, V. (2004). Cross-informant ratings of social competence in children and adolescents. *Clinical Psychology Review, 24,* 239–254.

Reynolds, C., & Kamphaus, R. (2004). *Behavioral Assessment System for Children* (2nd ed.). Minneapolis, MN: Pearson Assessments.

Richters, J. (1992). Depressed mothers as informants about their children: A critical review of the evidence for distortion. *Psychological Bulletin, 112,* 485–499.

Rosenthal, R., & Rosnow, R. (2008). *Essentials of behavioral research: Methods and data analysis.* New York, NY: McGraw Hill.

Rosenthal, R., & Rubin, D. (1984). Multiple contrasts and ordered Bonferroni procedures. *Journal of Educational Psychology, 76,* 1028–1034.

Sparrow, S., Cichetti, D., & Balla, D. (2005). *Vineland Adaptive Behavior Scales* (2nd ed.). Minneapolis, MN: Pearson Assessments.

Walker, H. M., Ramsay, E., & Gresham, F. M. (2004). *Antisocial behavior at school: Evidence-based practices* (2nd ed.). Belmont, CA: Wadsworth/Thomson Learning.

Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting and Clinical Psychology, 68,* 1038–1050.

Received May 22, 2009
Revision received July 9, 2009
Accepted July 9, 2009 ■