

Gagnon, S. G., Nagle, R. J., & Nickerson, A. B. (2007). Parent and teacher ratings of peer interactive play and social-emotional development of preschool children at risk. *Journal of Early Intervention*. 29(3): 228-42 (Spring 2007).  
Published by SAGE (ISSN: 1053-8151).

## Parent and Teacher Ratings of Peer Interactive Play and Social-Emotional Development of Preschool Children at Risk

Sandra Glover Gagnon, Richard J. Nagle, and Amanda B. Nickerson

### ABSTRACT

Estimates of agreement and consistency between parent and teacher ratings of peer-interactive play and social emotional development were examined in a sample of preschool children at risk for academic difficulties. Eighty-three pairs of parents and teachers of 4-year-old children completed the Penn Interactive Peer Play Scale (PIPPS) and 69 of these pairs also were interviewed with the Vineland Social-Emotional Early Childhood Scales (SEEC). Significant mean differences were found between parent and teacher scores on the PIPPS, but not on the SEEC. No noteworthy correlations were found between raters on the PIPPS, though low to moderate correlations were revealed between parent and teacher responses on the SEEC. Multivariate techniques revealed statistically significant and noteworthy correlations between parent and teacher responses on the SEEC but not on the PIPPS. Results are discussed within the context of conducting multifaceted, ecological assessments with preschool children. Implications for early childhood educators and interventionists are noted regarding the importance of gathering and interpreting multiple sources of information that contribute to an understanding of young children's behavior.

The increased emphasis on prevention and early intervention in the context of the Individuals with Disabilities Education Act (IDEA, 2004) and in other early childhood initiatives such as Good Start, Grow Smart (U.S. Department of Health and Human Services, 2005) has led to a rapid increase in the assessment of preschoolers' social-emotional behavior. The identification of social and emotional problems in young children is critical to intervening early to reduce the likelihood of placement in special education, residential treatment, and incarceration later in life (Squires, Bricker, Heo, & Twombly, 2001). A hallmark of effective assessment is that it is multidimensional, which involves including information from multiple perspectives, instruments, methods, and settings (Bagnato & Neisworth, 1991; Merrell, 1996).

Because preschoolers' behavior is so variable within testing situations (Nagle, 2000), evaluators must rely on input from parents and teachers to obtain complete descriptions of children's functioning across people and contexts. Obtaining information from both parents and teachers also improves validity. For example, a longitudinal study found that the combination of parent and teacher ratings of social-emotional and behavioral functioning resulted in improved predictive power than rating scale data obtained from only parents or only teachers for children's academic, behavioral, and mental health outcomes 6 years later (Verhulst, Koot, & Van der Ende, 1994).

### **Preschool Play-Based Assessment**

A variety of assessment techniques are used to assess preschool social-emotional functioning, including interviews, observations, and third-party rating instruments (Keith & Campbell, 2000; Martin, 1986). To provide the most valid estimates of developmental functioning, Bagnato and Neisworth (1991) advocated for using a multifaceted, ecological approach to examine preschoolers' developmental skills within social, school, and family contexts using multiple methods,

traits, and sources. Two promising methods for use in ecological assessments are play-based assessment and semi-structured interviews.

In addition to serving as an important context during which children learn and develop, play provides an important window through which to view development (Pellegrini, 1998). Play-based assessment provides information on developmental skills, reveals children's patterns of interactions with caregivers and peers, and focuses on play as a primary context for learning and development (Paget, 1999). Play-based assessment is also functional, because the results can be linked immediately to intervention strategies (Lowenthal, 1997). Because play proceeds through a regular developmental sequence during childhood, evaluations of children's play can provide indications of maturation and social competence. Some investigators have examined play within the context of parent-child interactions and peer interactions, while others have developed standardized measures of play to use in the assessment of children's cognitive and language skills (e.g., Fewell, 1986; Linder, 1993; Westby, 1980). The present study used a measure of peer interactive play as an indicator of social competence.

The Penn Interactive Peer Play Scale (PIPPS; Fantuzzo, Coolahan, Mendez, McDermott, & Sutton-Smith, 1998; Fantuzzo, Mendez, & Tighe, 1998) was developed in response to the paucity of psychometrically sound, developmentally appropriate, and culturally sensitive behavior rating scales for preschoolers. The PIPPS utilizes parent and teacher ratings of skills observed during play to distinguish children who demonstrate effective peer interactions from those who display less positive behaviors.

Semi-structured interviews are also a viable method to gather parents' and teachers' perspectives about preschoolers' social-emotional functioning. Advantages of an interview format include the opportunity to establish rapport, engage in discussion, and obtain a more elaborate description of

children's behaviors. The Vineland Social-Emotional Early Childhood Scales (SEEC; Sparrow, Balla, & Cicchetti, 1998) uses the semi-structured interview format, relying on caregiver responses to provide an overview of the child's social and emotional functioning in the areas of attentional skills, social interactions, construction and observation of relationships, self-regulation, and understanding emotions.

### **Parent-Teacher Convergence on Measures of Social-Emotional and Behavioral Functioning**

Although gathering information from multiple sources is critical in conducting ecological and comprehensive assessments (Neisworth & Bagnato, 2005), different raters often present evaluators with contradictory findings. Studies of behavior rating scales with school-age children have found wide variability in interrater reliability score coefficients and a consistent tendency for these to be lower than test-retest and internal consistency score reliability coefficients (Achenbach, McConaughy, & Howell, 1987; Elliott, Busse, & Gresham, 1993). Different status informants, such as parents and teachers, tend to give discrepant ratings of behavioral and emotional functioning, with a meta-analysis revealing a mean reliability coefficient of .28 (Achenbach et al., 1987). Score reliability coefficients generally are more consistent when informants are rating externalizing, as opposed to internalizing, behaviors (Kolko & Kazdin, 1993; Nickerson & Nagle, 2001) and correlations tend to be higher for children ages 6 through 11 than for adolescent samples (Achenbach et al., 1987).

Although the majority of research in this area has been conducted with school-age children and adolescents, studies focused on preschoolers have found similar trends. Walker and Bracken (1996) investigated the consistency between parental reports of preschool behaviors in a sample of 64 parents (32 pairs) of 4- and 5-year-old children, the majority of whom were Caucasian. Their

results revealed large variations in consistency between mothers and fathers on four preschool behavior rating scales ( $r$  ranging from .09 to .79). Higher levels of consistency were found for clinical symptoms on behavior rating scales than for indicators of child temperament on a temperament scale. The consistency between mother and father ratings also was stronger when rating positive rather than negative behaviors.

In their sample of 134 preschoolers (mean age = 57.4 months) of ethnic minority backgrounds attending pre-kindergarten Head Start programs, Fagan and Fantuzzo (1999) found statistically significant but relatively modest correlations (ranging from .17 to .54) for 6 of the 16 correlations between mothers' and fathers' ratings of social skills. In contrast, no statistically significant correlations emerged between mothers' and teachers' ratings and although 6 of the 20 correlations between fathers' and teachers' ratings were statistically significant, the magnitude of these correlations was only .17 to .25. Overall, these findings demonstrated a relative lack of consistency between parents' and teachers' ratings of preschoolers' social skills.

Findings of low convergence between parents and preschool teachers also have been found on ratings of temperament (Northam, Prior, Sansom, & Oberklaid, 1987) and behavior problems (Hinshaw, Han, Erhardt, & Huber, 1992). Northam et al. compared mothers' and child care providers' ratings of temperament in a sample of 43 children between 12 and 23 months and 60 children ages 24 to 39 months. One of the nine correlations yielded in the analyses with the younger group of children was statistically significant and moderate in strength ( $r = .49$ ,  $p < .01$ ), while three moderately strong, statistically significant correlations ( $r = .36$  to  $.48$ ) emerged for the older group. In the Hinshaw et al. study, parent and teacher ratings of 45 children, primarily of Caucasian descent and ranging in age from 3.5 to 6.8 years, essentially were uncorrelated for internalizing behaviors ( $r = .13$ ) and

modestly associated for externalizing behaviors ( $r = .32$ ).

Several possible reasons might account for the low convergence between parents' and teachers' ratings. Situation specificity is a commonly cited explanation that focuses on contextual issues, based on the assumption that behavior varies as a function of the setting in which it occurs (Kazdin, 1979). In addition, parents and teachers have different opportunities to observe behavior, with parents tending to have greater familiarity with their children's pattern of speech and nonverbal cues (Diamond & Squires, 1993; Kaufman, Swan, & Wood, 1980). Rater biases of informants, including response set, social desirability, and frame of reference, also have been found to affect convergence (Nickerson & Nagle, 2001; Youngstrom, Loeber, & Stouthamer-Loeber, 2000). Child variables, such as socio-economic and racial-ethnic differences, also have been correlated with low rater convergence (Fagan & Fantuzzo, 1999; Kaufman et al., 1980; Youngstrom et al., 2000). For preschool children, differences in ratings also might be due to transitional periods in development, when emerging competencies are not yet displayed consistently across settings (Diamond & Squires, 1993).

Existing research frequently has conceptualized inter-rater reliability from a unitary perspective by focusing primarily on the consistency of two ratings by calculating correlation coefficients. Stemler (2004) asserted that this common practice of describing inter-rater reliability as a single, unitary concept is imprecise. Instead, he differentiated between three different categories of inter-rater reliability: (a) consensus estimates, which measure agreement; (b) consistency estimates; and (c) measurement estimates, which assess the extent to which raters assess underlying common constructs.

The large majority of published studies have used behavior rating scales to assess convergence between parents' and professionals' perspectives about children's social

and behavioral competence; however, there is evidence that convergence varies based on the content assessed (e.g., specific skills vs. global performance) and the format of the assessment (e.g., recognition vs. recall; Diamond & Squires, 1993). Play-based assessment and semi-structured interviews of preschool social-emotional functioning, therefore, might yield different findings of rater consistency as compared to previous results lacking convergence.

Szatmari, Archer, Fisman, and Streiner (1994) found consistency between parent and teacher ratings of 83 children ages 4 to 6 years with developmental delays ( $r = .60$ ) on the socialization domain of the Vineland Adaptive Behavior Scales. In terms of peer interactive play, Fantuzzo et al. (1998) found statistically significant but moderately low ( $r = .18$  to  $.36$ ) correlations between corresponding domains on the parent and teacher versions of the PIPPS in a sample of 297 primarily African American children ages 37 to 64 months attending an urban Head Start program. Taken together, these findings support the premise that the type and content of assessment tools might influence rater convergence positively, but further study is needed.

The purpose of the present study was to examine three aspects of convergence between parent and teacher ratings on two different measures of behavioral and social-emotional functioning in a sample of preschool children at risk for academic difficulties. Specific research questions were: (a) Is there agreement (consensus) between parent and teacher ratings on the PIPPS and SEEC? (b) How consistent are parent and teacher ratings on the PIPPS and SEEC? and (c) Do parent and teacher ratings on the PIPPS and SEEC reflect measurement of common underlying constructs?

## **METHOD**

### **Participants**

This study was conducted in a public school district in the southeastern United States serving over 15,000 students in 17 schools. Data were collected from students enrolled in an early childhood program for 4-year-old children considered to be at risk for school problems. Children referred to the program by parents, pediatricians, child care personnel, or other community agencies were screened by the program's teachers using the Developmental Indicators Assessment of Learning - Third Edition (DIAL-3; Mardell-Czudnowski & Goldenberg, 1998). Eligibility was based on DIAL-3 scores, with children who scored in the lowest percentiles considered "at risk" and admitted to the program.

Parents of all 300 children in the program were contacted for participation in the study, with the exception of those children with diagnosed disabilities and children for whom English was not their primary language. For this study, researchers were interested in evaluating the convergence of ratings of parents and teachers of children at risk. Additionally, scores on measures not developed for use with participants for whom English is not their first language are compromised because of possible misinterpretations of the written and oral questions presented. Parents with minimal competence in English might not have been able to respond to the written or oral questions and their inclusion might have compromised the validity of the findings.



*Table 1*  
*Parent Characteristics*

	Mothers		Fathers	
	<i>n</i>	%	<i>n</i>	%
<b>Education</b>				
Some high school	9	10.6	11	13.0
High school diploma	32	37.6	31	36.5
Some college	28	33.0	20	23.6
College degree	10	11.8	10	11.8
<b>Occupational status</b>				
Homemaker	33	42.3	–	–
Clerical	18	23.1	2	2.9
Management	6	7.7	18	26.1
Trade	5	6.4	35	50.7
Professional	4	5.1	3	4.3
Education	3	3.8	–	–
Sales	3	3.8	1	1.4
Self-employed	3	3.8	–	–
Student	2	2.6	–	–
Military	1	1.3	4	5.8
Retired	–	–	1	1.4
Missing	7	8.2	16	18.8

Permission was granted from 85 parents, resulting in 43 male and 42 female participants ranging in age from 50 to 66 months ( $M = 57.74$ ,  $SD = 3.86$ ). Eighty children were Caucasian, four were African American, and one was Hispanic. Although 85 children participated in the study, the sample size varied for the different analyses because of the numbers of matched parent-teacher data. Eighty-three matched cases were used in the PIPPS analyses while 69 were used in the SEEC comparisons.

The average number of years of education for mothers and fathers was approximately 13 (mothers:  $M = 12.85$ , fathers:  $M = 12.78$ ). The most typical mother in the sample (42.3%) described herself as a homemaker and 50% of fathers reported working in trade occupations. Table 1 shows detailed information on characteristics of parent respondents. All eight teacher respondents were Caucasian and had at least a college degree.

Each teacher rated between 6 and 17 children (M = 4.93; SD = 2.26).

### **Instrumentation**

Penn Interactive Peer Play Scale. Parents' and teachers' ratings of behavior during peer play were assessed with the respective versions of the Penn Interactive Peer Play Scale (PIPPS; Fantuzzo, Coolahan et al., 1998; Fantuzzo, Mendez et al., 1998). The PIPPS is a 32-item rating scale designed to differentiate children who display positive versus negative behaviors during play with peers. The teacher version was established initially with a sample of 312 African American children, from families of low income and attending an urban Head Start program (Fantuzzo et al., 1995) with a revised version later validated with a different sample of 523 children who were African American, from families of low income, and attending an urban Head Start program (Fantuzzo, Coolahan et al., 1998). A parent version was developed with a sample of 297 African American children, from families of low income and attending an urban Head Start program, providing a 32-item assessment tool to use across settings and raters (Fantuzzo, Coolahan et al., 1998).

Examples of behaviors assessed by the PIPPS include helpfulness, fighting or arguing, showing positive or negative emotions, settling conflicts, and disrupting play (Fantuzzo, Coolahan et al., 1998; Fantuzzo, Mendez et al., 1998). Construct validity of scores was evidenced with the standardization samples by exploratory factor analyses, which revealed a three-factor structure for both versions: Play Interaction, Play Disruption, and Play Disconnection. Play Interaction is an indicator of children's prosocial behaviors and play strengths that includes items assessing helpfulness, creativity, and getting along with others. Play Disruption represents aggression and nonsocial play, as measured by items assessing demandingness, whining, and tattling. Play Disconnection is an indicator of withdrawn behaviors and

lack of participation in play activities and includes items assessing being ignored, confused, or remaining outside of play groups.

Based on normative data, raw scores are converted to t-scores, which were used for the current analyses. For the parent version, the Play Interaction and Play Disruption scales each consist of 10 items, while Play Disconnection consists of 9 items. For the teacher PIPPS, Play Interaction consists of 8 items, Play Disconnection consists of 9 items, and Play Disruption is comprised of 13 items.

Concurrent validity of PIPPS scores was established by comparing parent and teacher versions to the Social Skills Rating System (SSRS; Gresham & Elliot, 1990) in a sample of 312 children aged 38 to 63 months of African American descent from families of low income enrolled in a Head Start Program. Positive correlations were found between Play Interaction scores and scores on the social skills factors of the SSRS ( $r = .43$  to  $.63$ ) and between Play Disruption and Play Disconnection and the SSRS behavior problems factors ( $r = .10$  to  $.40$ ). Negative correlations were found between Play Interaction and the SSRS problem behaviors factors ( $r = 2.16$  to  $2.18$ ) and between Play Disruption and Play Disconnection and the social skills factors ( $r = 2.15$  to  $2.55$ ; Fantuzzo, Coolahan et al., 1998; Fantuzzo, Mendez et al., 1998).

Vineland Social-Emotional Early Childhood Scales. The Vineland Social-Emotional Early Childhood Scales (SEEC; Sparrow et al., 1998), an early childhood measure of socialemotional development using a semi-structured interview format with a familiar adult, was administered to parents and teachers of children in the study. The SEEC was derived from the Socialization Domain of the Vineland Adaptive Behavior Scales (ABS; Sparrow, Balla, & Cicchetti, 1984) and is appropriate for use with children from birth through 5 years, 11 months. Standard scores with a mean of 100 and standard deviation of 15 are obtained for each of the three scales and the composite. The Interpersonal Relationships scale, which

provides an indication of children's interactions with others, includes 44 items assessing behaviors such as emotional expression, desire to please others, and smiling. The 44 items from the Play and Leisure Time scale assess how children use play and leisure time, such as engaging in make-believe play and sharing. The ways in which children display responsibility and sensitivity to others is captured by the Coping Skills scale, which contains 34 items assessing behaviors such as following rules and not interrupting.

Evidence for the psychometric properties of SEEC scores is provided in the test manual. It should be noted that the norms for the instrument were derived from the normative data used in the development of the Vineland Adaptive Behavior Scales (ABS; Sparrow et al., 1984), which are almost 20 years old. Internal consistency of scores was satisfactory, ranging from .80 to .93 and from .85 to .96 for the 4-year-old portion of the sample (n = 200). Although test-retest reliability coefficients ranged from .74 to .79, they were lower for different interviewers, ranging from .47 to .50. Regarding the concurrent validity of SEEC scores, comparisons with the Battelle Developmental Inventory Personal-Social Domain (BDI; Newborg, Stock, Wnek, Guidubaldi, & Svinicki, 1984) and the Scales of Independent Behavior Early Developmental Scale (SIB; Bruininks, Woodcock, Weatherman, & Hill, 1984) revealed coefficients of .65 (N = 67) and .63 (N = 66), respectively (Johnson, Cook, & Kullman, 1992).

## **Procedures**

Of the school district's nine elementary schools, all of which had 4-year-old programs, eight participated in the study. Once informed consent was obtained, all parents and teachers of 4-year-old children in the participating schools were contacted to schedule SEEC interviews and to complete the PIPPS. The teacher measures were administered at least 6 to 8 weeks after the

beginning of the school year to allow time for the teachers to become familiar with the students. Data collection began in November and continued through February of the same school year. Three graduate students in school psychology with advanced training in interviewing and assessment conducted all interviews.

**Table 2**  
*Means, Standard Deviations, and Mean Score Differences on PIPPS and SEEC*

Measure	Parent <i>M</i> ( <i>SD</i> )	Teacher <i>M</i> ( <i>SD</i> )	<i>t</i>	$\eta^2$
PIPPS <i>t</i> scores	<i>n</i> = 83	<i>n</i> = 85		
Play interaction <sup>a</sup>	47.41 (8.02)	61.00 (7.92)	4.35*	.19
Play disruption <sup>b</sup>	52.69 (7.78)	27.09 (20.46)	3.80*	.19
Play disconnection <sup>b</sup>	47.67 (13.96)	47.60 (10.95)	-3.83*	.15
SEEC standard scores	<i>n</i> = 80	<i>n</i> = 73		
Interpersonal relationships <sup>c</sup>	99.69 (13.60)	95.90 (17.29)	1.59	.03
Play and leisure time <sup>c</sup>	94.88 (12.29)	94.26 (10.05)	.59	.00
Coping skills <sup>c</sup>	101.50 (10.72)	106.44 (11.72)	-3.90*	.16

*Note.* PIPPS = Penn Interactive Peer Play Scale; SEEC = Vineland Social-Emotional Early Childhood Scales; <sup>a</sup> = Higher scores denote more positive peer play behaviors on PIPPS. <sup>b</sup> = Higher scores denote more negative peer play behaviors on PIPPS. <sup>c</sup> = Higher scores denote more positive social-emotional skills on SEEC.  
\**p* < .001.

The primary investigator provided specific training on both measures. Training focused primarily on semi-structured interviewing and proceeded until the students' demonstrated proficiency in administration and scoring of the SEEC. The primary researcher observed each student during the first three administrations of the SEEC and conducted periodic checks on administration and scoring. The students met with most parents at school and administered the PIPPS and SEEC. The graduate student evaluators were assigned to specific classrooms so that each interviewed parentteacher pairs. Because of transportation difficulties and conflicts with work schedules, approximately 5% to 10% of parents completed the PIPPS at home and participated in phone interviews. Teachers were interviewed with the SEEC at school and completed the PIPPS independently.

Data were collected from parents and teachers of the same students within a time period of 2 to 4 weeks. Although attempts were made to obtain information on both measures from each subject, in some cases this was not possible. PIPPS scores were obtained from 83 parents and 85 teachers, while SEEC scores were collected from 80 parents and 73 teachers. With regard to the lower number of teacher SEEC scores, the time-consuming nature of the SEEC interviews precluded teachers from completing this aspect of the data for all participating children in their classrooms. Of the 85 total subjects, 16 did not have both parent and teacher SEEC scores. Thus, the number of subjects for the SEEC analyses was 69.

## RESULTS

To evaluate parent-teacher agreement, Table 2 shows an overview of the mean scores, mean score differences, and effect sizes for the PIPPS and SEEC. Statistically and practically significant mean score differences were obtained between raters on all three PIPPS scales but on only one of the SEEC scales.

The relationships between parent and teacher ratings (i.e., consistency of scores) on the PIPPS and SEEC were analyzed using the Pearson product-moment correlation coefficient. Preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity, and homoscedasticity. Only pairs of raters were included in the analyses, thus explaining differences in the sample sizes from the original 85 participants reported previously.

As shown in Table 3, no statistically significant or noteworthy correlations emerged between parent and teacher PIPPS scores, which is in contrast to the findings from the Fantuzzo, Mendez et al. (1998) study, also shown in Table 3. In contrast, comparisons of the parent and teacher SEEC responses revealed statistically significant but low to moderate, positive correlations between

the Interpersonal Relationships ( $r(69) = .31, p < .01$ ) and Coping Skills ( $r(69) = .46, p < .01$ ) scales, but not for the Play an Leisure Time scale ( $r(69) = .19, p < .12$ ), as shown in Table 4.

**Table 3**  
*Correlations between Parent and Teacher PIPPS Subscale Scores for Present Sample and Fantuzzo, Mendez, & Tighe, (1998) Sample (n = 83)*

Parent factors	Teacher Factors					
	Present sample			Fantuzzo et al. sample		
	Play interaction	Play disruption	Play disconnection	Play interaction	Play disruption	Play disconnection
Play interaction	.18	-.08	-.19	.36**	-.21*	-.24**
Play disruption	.02	.05	-.17	-.08	.24**	.08
Play disconnection	.08	.03	-.16	-.12	.18*	.25**

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

To examine whether the parent and teacher ratings on the PIPPS and SEEC assessed common underlying constructs, canonical correlation analyses were performed. Canonical correlation is a multivariate technique used to examine complex interactions between two sets of variables when each set contains two or more variables (Thompson, 1984, 2004; Weiss, 1972). Canonical correlation analysis was considered appropriate to use with the present data set because of the multifaceted nature of the variables involved. In canonical correlation analysis, the observed variables in each set are combined statistically to represent unobserved, or latent, variables (Sherry & Henson, 2005). The correlation between these latent variables is examined and can be conceptualized as a bivariate correlation. Three important statistics typically are reported from canonical correlation analyses. The canonical correlation coefficient represents the relationship between the latent variables. The standardized canonical function coefficient also is reported. The canonical function is a set of standardized canonical function coefficients for the variable

sets, each of which is independent of the other. The number of functions is equal to the number of variables in the smaller set. Finally, structure coefficients, which are correlations between each observed variable in both sets and their respective latent variables, are reported. Standard scores for all three scales on the teacher and parent PIPPS scores and teacher and parent SEEC scores were included in the analyses, with parent scores in one variable set and teacher scores in the other for both measures. Table 5 shows the standardized canonical function coefficients and structure coefficients for all resulting functions for both measures.

**Table 4**  
*Correlations between Parent and Teacher SEEC Subscale Scores (n = 69)*

Parent factors	Teacher factors		
	Interpersonal relationships	Play and leisure	Coping skills
Interpersonal relationships	.31**	.22	.25*
Play and leisure	.26*	.19	.28*
Coping skills	.13	.35**	.46**

*Note.* SEEC = Vineland Social-Emotional Early Childhood Scales.  
\*  $p < .05$ , \*\* $p < .01$ .

For the PIPPS analyses, the full canonical model was not statistically significant or noteworthy,  $L = .84$ ,  $F(9, 188) = 1.58$ ,  $p < .125$ , accounting for only 16% of the variance between the two sets of variables across all canonical functions. Although three canonical functions resulted from the PIPPS analyses, none were statistically significant,  $L = .84$ ,  $F(9, 188) = 1.58$  ( $p < .125$ );  $L = .93$ ,  $F(4, 156) = 1.52$  ( $p < .198$ ); and  $L = .99$ ,  $F(1, 79) = .24$  ( $p < .623$ ), respectively. The canonical correlations for the three pairs were .31, .27, and .06 and explained 9.6%, 7.1%, and less than 1% of the variance between the variable sets, respectively. Sherry and Henson (2005) recommend that only those canonical functions that explain a meaningful amount of variance should be interpreted. Because each of



the resulting functions from the PIPPS analyses explained less than 10% of the variance, they were not interpreted further.

**Table 5**  
*Canonical Structure of Parent and Teacher SEEC (n = 69)*

Variable	Function 1		Function 2		Function 3	
	Coefficient	$r_s$	Coefficient	$r_s$	Coefficient	$r_s$
Parent PIPPS subscales						
Play interaction	-.94	-.73	-.43	-.66	-.17	-.19
Play disruption	-.40	-.36	.16	.47	.97	.81
Play disconnection	-.48	-.35	.72	.89	-.64	-.30
Teacher PIPPS subscales						
Play interaction	-.44	-.81	-.20	-.31	-1.16	-.50
Play disruption	-.12	.37	1.04	.88	-.45	-.29
Play disconnection	.74	.94	-.59	-.04	-.82	-.35
Parent SEEC subscales						
Interpersonal relationships	-.08	.20	.73	.94	-1.15	-.29
Play and leisure time	-.38	.31	.51	.81	1.41	.50
Coping skills	1.21	.94	-.33	.31	.21	.14
Teacher SEEC subscales						
Interpersonal relationships	-.71	.06	.09	.97	.04	-.23
Play and leisure time	.41	.65	-.38	.44	-1.47	-.62
Coping skills	.94	.84	.21	.54	1.13	.09

*Note.* SEEC = Vineland Social-Emotional Early Childhood Scales; PIPPS = Penn Interactive Peer Play Scale; *coefficient* = standardized canonical function coefficient;  $r_s$  = structure coefficient. Structure coefficients  $\geq .45$  are considered appreciable.

The SEEC comparisons yielded a full canonical model that was statistically significant and noteworthy,  $L = .65$ ,  $F(9, 153) = 3.36$ ,  $p < .001$ , accounting for 35% of the variance among the two sets of variables across all canonical functions. Though three canonical functions emerged, only the first was statistically significant and meaningful, with a canonical correlation coefficient of .52. This function accounted for 27% of the variance shared across raters in relation to the latent variable represented by this canonical function and was characterized by strong positive relations among parent and teacher reports on the Coping Skills scale and teacher reports on the Play and Leisure Time scale,  $L = .65$ ,  $F(9, 153) = 3.36$ ,  $p < .001$ . These results are supported by the large structure

coefficients for these variables (see Table 5). Given that the behaviors assessed by these scales reflect sharing, cooperation, interest and engagement in play, using manners, controlling impulses, and being sensitive to others, this function was labeled Sensitivity.

## **DISCUSSION**

Findings from this study revealed important estimates of reliability, including agreement/consensus, consistency, and measurement of a common underlying construct for a playbased rating scale and a semi-structured interview of social-emotional behavior for a sample of preschoolers at risk for academic difficulties. First, statistically significant mean differences were found between parent and teacher scores on all PIPPS subscales, but only on one SEEC subscale. Second, there was a lack of noteworthy correlations between parent and teacher scores on the PIPPS, though low to moderate correlations emerged for the SEEC. Third, canonical correlations revealed no underlying constructs measured by parent and teacher ratings on the PIPPS, though the SEEC parent and teacher ratings assessed underlying constructs related to sensitivity and responsiveness.

The lack of agreement, consistency, and assessment of underlying constructs for PIPPS scores was consistent with the growing body of literature finding a lack of consistency across parents' and teachers' ratings (Fagan & Fantuzzo, 1999; Hinshaw et al., 1992). Results from the initial validation study of the parent version of the PIPPS, which revealed statistically significant low to moderate correlations between parent and teacher ratings (Fantuzzo, Mendez et al., 1998), were a notable exception. Characteristics of study samples might explain differences across studies. The current study sample of 85 primarily Caucasian children residing in a relatively rural geographic area was smaller than the original standardization sample of 297 Head Start children from an urban setting in the Fantuzzo et al study. Further, the majority of studies examining

convergence have focused on score consistency rather than also examining agreement and measurement of underlying constructs, which was evaluated in the present study.

Several plausible explanations exist when considering the lack of agreement found between parents and teachers in the present study. First, it is likely that parents observed qualitatively different behaviors than teachers given the different settings in which they observed the children. Many of the preschoolers had been cared for in their homes by their mothers, without being exposed to a large number of children in a structured situation before entering the preschool program. While completing the PIPPS, some parents commented that they had not had the opportunity to observe a few of the behaviors described in particular items on the scale, as some children did not have siblings at home and some lived in rural areas without other children nearby.

Second, parents' and teachers' perspectives were likely quite different. Over one-third of the mothers in the study did not work outside of the home and many did not have their children involved in activities with other children on a regular basis. In contrast, teachers had numerous other children to use as comparisons. Parent-child interactions also are qualitatively different than teacher-child interactions (Kemple, 1991), and these differences might contribute to different behaviors being exhibited in the presence of different adults. This also might be an explanation for why parent and teacher ratings did not assess similar underlying constructs.

Third, individual differences in adults' tolerance for various behaviors might have led to differences in ratings, as past research has shown (Kaufman et al., 1980; Youngstrom et al., 2000). Although the items on the PIPPS represent clearly defined, easily identifiable behaviors, raters might have differed in the extent to which they perceive the occurrence or severity of the behaviors. Parents tended to rate children's play behavior

as more disruptive and less interactive than did teachers. It is likely that this is due to interactions of many of the aforementioned explanations, such as observing the children in different settings and with different peers (e.g., siblings as opposed to classmates), divergent perspectives on behavior, and varying degrees of tolerance.

It is also noteworthy that each parent rated only one child while each teacher rated between 6 and 17 children. Thus, additional factors relating to the teacher ratings might have contributed to the lack of agreement, including fatigue, response bias, or contrast effects.

In contrast to the lack of convergence between parent and teacher ratings on the PIPPS, there was only one statistically significant difference between parent and teacher scores on the SEEC. In addition, modest correlations emerged between parent and teacher ratings on the SEEC.

The use of canonical correlations helped to clarify the multivariate relationships between different informants on this measure of social emotional functioning. The first canonical variate that emerged from this analysis, called sensitivity, reflected strong, positive relationships between parent and teacher reports on the Coping Skills scale and teacher reports on the Play and Leisure scale. This synthetic or latent variate, which accounted for almost one-third of the shared variance, indicated that children who scored high on parent ratings of Coping Skills also scored high on teacher ratings of Coping Skills and Play and Leisure Time. The Coping Skills scale reflects responsibility and sensitivity to others and assesses children's use of manners, responses to limits and rules, impulse control, and management of feelings (Sparrow et al., 1998). The activities engaged in during play are reflected in scores on the Play and Leisure Time scale, which assesses the ways in which children share, cooperate, and play with others. Parents and teachers were able to identify consistently and describe qualitatively children's coping skills, and children who

displayed positive coping skills also were observed by their teachers to engage in positive and appropriate play activities. This finding is consistent with results of a study conducted by Walker and Bracken (1996), indicating that raters were more consistent when considering positive rather than negative behaviors.

Considering the results of the present study, the question arises as to why there was convergence on the SEEC but not on the PIPPS. One possible explanation involves the nature of the measures, which are qualitatively different. The PIPPS is a rating scale that is completed individually by parents or teachers and does not require the presence of an examiner. In contrast, the SEEC is an individually administered, semi-structured interview, involving discussion and allowing for a more elaborate description of children's behaviors. Through the dialogue generated by the SEEC, it might have been possible for parents and teachers to report a more accurate description of children's behaviors, thus contributing to the resulting agreement, consistency, and assessment of underlying constructs. Not only are the measures different in terms of structure, they also involve different domains of behavior. Although there are similarities between some items, the PIPPS captures the quality of peer play interactions while the SEEC examines more general aspects of social-emotional functioning.

Another explanation might involve the difference in the length of each instrument. Because the total number of items on the SEEC is greater than on the PIPPS, it would be expected that higher inter-rater score reliability coefficients would be obtained with a longer test. Assuming the items are related sufficiently to the construct being measured, longer tests provide a more representative sample of items through which to assess the domain of interest.

While differences between raters can be expected in behavioral assessments, method variance might affect the level of agreement

and consistency attained (Walker & Bracken, 1996). Kazdin (1979) emphasized the importance of considering the effect that the type of assessment has on performance and the resulting scores. Because two measures were used in the present study, it was possible to evaluate whether rater differences were related to variance in instrumentation or if the differences reflected actual variations in parents' and teachers' observations of the same behaviors. The lack of consistency between parent and teacher ratings on the PIPPS as compared to the SEEC suggests method variance as a primary source of error, though, as mentioned previously, other sources of error, such as situation specificity, differences in the extent of exposure to normative samples of children, and individual differences in parents' and teachers' tolerance for various behaviors, also could contribute to the resulting differences.

The current findings suggest that when using rating scales to evaluate behavior, evaluators might anticipate low levels of agreement and consistency and thus, should consider ratings from each informant as individual contributions to the assessment. In contrast, the use of semi-structured interviews might facilitate more reliable descriptions

of children's behaviors in which parents' and teachers' ratings are consistent.

## **Limitations**

Though the current findings contribute meaningfully to the extant literature on parent and teacher ratings of preschool social and emotional behavior, limitations should be noted. The first involves the use of the SEEC. Though the psychometric properties of the SEEC are well established, it is important to note that the norms for the instrument were derived from the normative data used in the development of the Vineland Adaptive Behavior Scales (ABS; Sparrow et al., 1984). The ABS standardization sample was based on 1980 Census figures, and thus, the norms are almost 20 years old. When interpreting

scores from a norm-referenced instrument, consideration of the appropriateness of the norms is imperative. Consistent with potential environmental effects that contribute to the Flynn effect described within the context of intelligence testing (Flynn, 1984), it is possible that changes in the early childhood experiences of young children and families, such as improvements in physical health, nutrition, education, and technology, and increased access to information and services in health care and education, might have altered substantially the characteristics of young children's social-emotional development and influenced the current results.

Another limitation involves procedural variations. Although the majority of parents participated in face-to-face SEEC interviews and completed the PIPPS with the researchers present, in a small percentage of the cases this was not possible because of difficulties with transportation or work schedules, and thus, some parents completed the PIPPS at home and participated in phone interviews. Additionally, teachers were interviewed in person yet completed the PIPPS independently to minimize the burden on their time. It is also noteworthy that the teachers rated more than one child while the parents rated only one. These differences in data collection procedures might have affected the results and therefore, are important to consider.

Another limitation involves the fidelity of the data collection process. Though a procedure was followed for training the graduate student researchers, which included observations of the first three administrations of the SEEC by the primary investigator, the validity of the findings would be enhanced by continued evaluation of the fidelity of the data collection procedures. Additionally, the relatively small sample size might have affected power. Finally, the limited data on parent, family, and teacher demographics also pose restrictions in terms of the generalizability of the results.

## Implications and Future Directions

These findings have potential implications for individuals involved in the assessment and development of interventions for children of different ages, though identification of social-emotional difficulties is particularly important for young children, as early identification is critical for the prevention of negative outcomes. Furthermore, assessments involving multiple measures, multiple traits, and multiple informants are recommended practice in early intervention (Neisworth & Bagnato, 2005).

Results from this study suggest that the SEEC might be useful in gathering data from both parents and teachers that converge in terms of agreement, consistency, and measurement of the underlying construct of sensitivity. More caution might be warranted when using the PIPPS if the intent is to gather information that can be verified across settings and raters. This is not to say that the standardized assessment of play is not important; rather, evaluators should be very clear about the type of information they wish to gather about play. For example, is the purpose of the evaluation to assess perspectives about the child's play within particular settings (e.g., large groups of same-aged children) or to assess how this behavior might vary as a possible function of rater, setting, and playmates?

Although strong psychometric properties of measures are critical, the accumulating evidence suggests that behavior and perspectives about behavior vary considerably depending on whether a parent or teacher is rating the behavior. This suggests these differences should be explored systematically within the context of early childhood evaluation and assessment. For example, it would be of great interest to assess variables that might lead to teachers' rating children's behavior as more interactive and less disruptive. Perhaps the classroom structure or expectations for behavior are ones that assist children in this regard, which might be useful to share with parents.



Several directions for future research are suggested by the results of the present study. To identify further the reasons for relatively low parent-teacher convergence on the assessment of play, parents and teachers might be asked to observe the child in the same setting with the same peers for a period of time and conduct ratings. This could help identify whether rater differences or actual differences in behavior contribute to score variability. If it is found that it is actual differences in children's behaviors, manipulating variables within the settings (e.g., free play vs. structured activities, large group vs. dyadic play) could yield additional information about how play behaviors differ. In addition, studies are needed to assess the extent to which demographic variables and larger contextual issues (e.g., urban vs. rural settings) influence actual and rated play behavior and social-emotional functioning.

The results of this study underscore the importance of an ecological approach to the evaluation of preschoolers, which includes multiple methods and informants. Parents and teachers serve as essential informants of children's social-emotional and behavioral functioning. Though their reports tend to differ according to the type of methodology employed, both sources provide valuable information that should be used to understand better young children's behavior and to guide intervention efforts.

## REFERENCES

- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of crossinformant correlations for situational specificity. *Psychological Bulletin*, 101, 213–232.
- Bagnato, S. J., & Neisworth, J. T. (1991). *Assessment for early intervention: Best practices for professionals*. New York: Guilford Press.
- Bruininks, R. H., Woodcock, R. W., Weatherman, R. F., & Hill, B. K. (1984). *Scales of independent behavior*. Chicago: Riverside.
- Culbertson, J. L., & Willis, D. J. (1993). Introduction to testing young children. In J. L. Culbertson & D. J. Willis (Eds.), *Testing young children: A reference guide for developmental, psychoeducational, and psychosocial assessments* (pp. 1–10). Austin, TX: PRO-ED.
- Diamond, K. E., & Squires, J. (1993). The role of parental report in the screening and assessment of young children. *Journal of Early Intervention*, 17, 107–115.
- Elliott, S. M., Busse, R. T., & Gresham, F. M. (1993). Behavior rating scales: Issues of use and development. *School Psychology Review*, 22, 313–321.
- Fagan, J., & Fantuzzo, J. W. (1999). Multirater congruence on the social skills rating system: Mother, father, and teacher assessments of urban Head Start children's social competencies. *Early Childhood Research Quarterly*, 14, 229–242.
- Fantuzzo, J., Coolahan, K., Mendez, J., McDermott, P., & Sutton-Smith, B. (1998). Contextually-relevant validation of peer play constructs with African American Head Start children: Penn interactive peer play scale. *Early Childhood Research Quarterly*, 13, 411–431.

Fantuzzo, J., Mendez, J., & Tighe, E. (1998). Parental assessment of peer play: Development and validation of the parent version of the Penn interactive peer play scale. *Early Childhood Research Quarterly*, 13, 659–676.

Fantuzzo, J., Sutton-Smith, B., Coolahan, K. C., Manz, P. H., Canning, S., & Debnam, D. (1995). Assessment of preschool play interaction behaviors in young low-income children: Penn interactive peer play scale. *Early Childhood Research Quarterly*, 10, 105–120.

Fewell, R. R. (1986). *Play assessment scale* (5th rev.). Seattle: University of Washington.  
Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932–1978. *Psychological Bulletin*, 95, 29–51.

Gresham, F. M., & Elliott, S. M. (1990). *The social skills rating system*. Circle Pines, MN: American Guidance Services.

Hinshaw, S. P., Han, S. S., Erhardt, D., & Huber, A. (1992). Internalizing and externalizing behavior problems in preschool children: Correspondence among parent and teacher ratings and behavior observations. *Journal of Clinical Child Psychology*, 21, 143–150.

Individuals with Disabilities Education Improvement Act, Amendments of 2004, Pub. L. No. 108-446, U.S.C. 20, 11400 et seq.  
Johnson, L. J., Cook, M. J., & Kullman, A. J. (1992). An examination of the concurrent validity of the Batelle developmental inventory as compared with the Vineland adaptive scales and Bayley scales of infant development. *Journal of Early Intervention*, 16, 353–359.

Kaufman, A. S., Swan, W. W., & Wood, M. M. (1980). Do parents, teachers, and psychoeducational evaluators agree in their perceptions of the problems of black and white emotionally disturbed children? *Psychology in the Schools*, 17, 185–191.

- Kazdin, A. E. (1979). Situational specificity: The two-edged sword of behavioral assessment. *Behavioral Assessment*, 1, 57–75.
- Keith, L. K., & Campbell, J. M. (2000). Assessment of social and emotional development in preschool children. In B. A. Bracken (Ed.), *The psychoeducational assessment of preschool children* (3rd ed., pp. 364–382). Boston: Allyn and Bacon.
- Kemple, K. M. (1991). Preschool children's peer acceptance and social interaction. *Young Children*, 46, 47–55.
- Kolko, D. J., & Kazdin, A. E. (1993). Emotional/behavioral problems in clinic and nonclinic children: Correspondence among child, parent, and teacher reports. *Journal of Child Psychology and Psychiatry*, 34, 991–1006.
- Linder, T. (1993). *Transdisciplinary play-based assessment*. Baltimore: Brookes.
- Lowenthal, B. (1997). Useful early childhood assessment: Play-based, interviews, and multiple intelligences. *Early Child Development and Care*, 129, 43–49.
- Mardell-Czudnowski, C., & Goldenberg, D. S. (1998). *Developmental indicators for the assessment of learning* (3rd ed.). Circle Pines, MN: American Guidance Service.
- Martin, R. P. (1986). Assessment of the social and emotional functioning of preschool children. *School Psychology Review*, 15, 216–232.
- Merrell, K. W. (1996). Social-emotional problems in early childhood: New directions in conceptualization, assessment, and treatment. *Education and Treatment of Children*, 19, 458–473.
- Nagle, R. J. (2000). Issues in preschool assessment. In B. A. Bracken (Ed.), *Psychoeducational assessment of preschool children* (3rd ed., pp. 19–32). Needham Heights, MA: Allyn & Bacon.

Neisworth, J. T., & Bagnato, S. J. (2005). DEC recommended practices: Assessment. In S. Sandall, M. L. Hemmeter, B. J. Smith & M. E. McLean (Eds.), *DEC recommended practice: A comprehensive guide for practical application in early intervention/early childhood special education* (pp. 45–70). Longmont, CO: Sopris West.

Newborg, J., Stock, J. R., Wnek, L., Guidubaldi, J., & Svinicki J (1984). *Battelle developmental inventory*. Chicago: Riverside.

Nickerson, A. B., & Nagle, R. J. (2001). Interrater reliability of the Devereux behavior rating scale –school form: The influence of teacher frame of reference. *Journal of Psychoeducational Assessment*, 19, 299–316.

Northam, E., Prior, M., Sansom, A., & Oberklaid, F. (1987). Toddler temperament as perceived by mothers versus daycare givers. *Merrill Palmer Quarterly*, 33, 213–229.

Paget, K. D. (1999). Ten years later: Trends in the assessment of infants, toddlers, preschoolers, and their families. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 476–496). New York: Wiley.

Pellegrini, A. D. (1998). Play and the assessment of young children. In O. N. Saracho & B. Spodek (Eds.), *Multiple perspectives on play in early childhood education* (pp. 220–239). Albany, NY: State University of New York Press.

Sherry, A., & Henson, R. (2005). Conducting and interpreting canonical correlation analysis in personality research: A user-friendly primer. *Journal of Personality Assessment*, 84, 37–48.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland adaptive behavior scales*. Circle Pines, MN: American Guidance Service.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1998). *Vineland social-emotional early childhood scales (SEEC)*. Circle Pines, MN: American Guidance Service.

Squires, J., Bricker, D., Heo, K., & Twombly, E. (2001). Identification of social-emotional problems in young children using a parent-completed screening instrument. *Early Childhood Research Quarterly*, 16, 405–419.

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9. Retrieved June 7, 2006 from <http://pareonline.net/getvn.asp?v=9&n=4>

Szatmari, P., Archer, L., Fisman, S., & Streiner, D. L. (1994). Parent and teacher agreement in the assessment of pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24, 703–717.

U.S. Department of Health and Human Services, Administration for Children and Families. (2005). *A guide to good start grow smart and other federal early learning initiatives*. Washington, DC: Author.

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

Thompson, B. (1984). *Canonical correlation analysis: Uses and interpretation*. Newbury Park, CA: Sage Publications.

Verhulst, F. C., Koot, H. M., & Van der Ende, J. (1994). Differential predictive value of parents' and teachers' reports of children's problem behaviors: A longitudinal study. *Journal of Abnormal Child Psychology*, 5, 531–546.

Walker, K. C., & Bracken, B. A. (1996). Interparent agreement on four preschool behavior rating scales: Effects of parent and child gender. *Psychology in the Schools*, 33, 273–283.

Westby, C. E. (1980). Assessment of cognitive and language abilities through play. *Language, Speech, and Hearing Services in Schools*, 11, 154–168.

Weiss, D. J. (1972). Canonical correlation analysis in counseling psychology research. *Journal of Counseling Psychology*, 19, 241–252.

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.