Direct Assessment of Children's Social-Emotional Comprehension

Authors: Clark McKown (ORCID: 0000-0001-9694-1179), Adelaide M. Allen (ORCID: 0000-0002-8410-1604), Nicole M. Russo-Ponsaran (ORCID: 0000-0002-9430-152X), Jason K. Johnson

Publication Year: 2013

Journal: Psychological Assessment

Publisher: https://www.apa.org/pubs/journals/pas/

Psychological Assessment
2013, Vol. 25, No. 4, 1154–1166

1040-3590/13/$12.00 DOI: 10.1037/a0033435

Direct Assessment of Children's Social-Emotional Comprehension

Clark McKown, Adelaide M. Allen, Nicole M. Russo-Ponsaran, and Jason K. Johnson
Rush University Medical Center

Social-emotional comprehension includes the ability to encode, interpret, and reason about social-emotional information. The better developed children's social-emotional comprehension, the more positive their social interactions and the better their peer relationships. Many clinical tools exist to assess children's social behavior. In contrast, fewer clinically interpretable tools are available to assess children's social-emotional comprehension. This study evaluated the psychometric properties of a group of direct assessments of social-emotional comprehension. Scores on these assessments reflected children's performance on challenging tasks that required them to demonstrate their social-emotional comprehension. In 2 independent samples, including a general education school sample ($n = 174$) and a clinic sample ($n = 119$), this study provided evidence that (a) individual assessments yield variably reliable scores, (b) composite scores are highly reliable, (c) direct assessments demonstrate a theoretically coherent factor structure and convergent and discriminant validity, and (d) composite scores yield expected age- and diagnostic-group differences. Implications for clinical practice, theory, and assessment development are discussed.

Keywords: social-emotional learning, child assessment, emotion recognition, theory of mind, social problem solving

Social-emotional comprehension includes the ability to encode, interpret, and reason about social-emotional information (Lipton & Nowicki, 2009). Prior research has demonstrated that the better developed children's social-emotional comprehension, the more positive their social interactions and the better their peer relationships (Baron-Cohen, 1989; Bauminger, Edelsztein, & Morash, 2005; Collins & Nowicki, 2001; Crick & Dodge, 1994; Denham, 2006; Duke, Nowicki, & Walker, 1996; McKown, 2007; McKown, Gumbiner, Russo, & Lipton, 2009; Nowicki, Duke, & van Buren, 2008). Better peer relationships are in turn associated with better mental health and other important life outcomes (Parker & Asher, 1987).

Practitioners working with socially marginalized children would thus be well-served to include assessments of children's social-emotional comprehension as part of a social assessment battery and use assessment findings to guide treatment plans. Yet few clinical tools are suited to assessing social-emotional comprehension. With noteworthy exceptions (e.g., the Meyer-Salovey-Caruso Emotional Intelligence Test, Youth Version [MSCEIT-YV]; Mayer, Caruso, & Salovey, 2005), existing assessment strategies are better suited to assessing social behavior than social-emotional comprehension. Examples include behavior rating scales (e.g., Social Skills Intervention System; Gresham & Elliott, 2008), behavioral observations (e.g., Peer Social Behavior Code; Walker & Severson, 1992), peer nominations (Chan & Mpofu, 2001; Moreno, 1933), and self-report (e.g., Trait Emotional Intelligence Questionnaire-Child Form; Mavroveli, Petrides, Shove, & Whitehead, 2008). In contrast, social-emotional comprehension involves the activation and deployment of internal mental processes. Accordingly, it would be better measured with direct assessments requiring children to solve social–emotional tasks.

Most of the few existing direct assessments narrowly assess a specific dimension of social-emotional comprehension, which limits their clinical utility. For example, the Diagnostic Analysis of Nonverbal Accuracy (DANVA; Nowicki & Duke, 1994) measures emotion recognition from faces, posture, and affective prosody. Two of the DANVA's strengths are that it includes child faces for the facial emotion recognition subtest and that it can be used from preschool through adulthood. The Mind in the Eyes Test Child Version (Baron-Cohen, Wheelwright, Scahill, Lawson, & Spong, 2001) is a mental-state inference task in which children view photos of the eye region of faces and indicate what the person is thinking or feeling. It is currently used for research purposes only, and there is no age range provided. The NEPSY-II (Korkman, Kirk, & Kemp, 2007) includes Affect Recognition and Theory of Mind (ToM) subtests for children ages 5 to 16 years old.

Two assessments measure social-emotional comprehension more broadly, but focus on targeted populations. The Social Information Processing Application (SIP-AP; Kupersmidt, Stelter, & Dodge, 2011) measures several dimensions of children's ability to solve hypothetical social problems and was designed to assess

This article was published Online First July 1, 2013.

Clark McKown, Adelaide M. Allen, Nicole M. Russo-Ponsaran, and Jason K. Johnson, Department of Behavioral Sciences, Rush University Medical Center.

This research reported here was supported by grants from the Dean and Rosemarie Buntrock Family Foundation and the Institute of Education Sciences through Grant R305A110143 to Rush University Medical Center. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Correspondence concerning this article should be addressed to Clark McKown, Department of Behavioral Sciences, Rush University Medical Center, RNBC 4711 Golf Road, Suite 1100, Skokie, IL 60076. E-mail: Clark_A_McKown@rush.edu

social-cognitive correlates of aggression in boys ages 8 to 12 years old (Kupersmidt et al., 2011). The MSCEIT-YV (Mayer et al., 2005) assesses the ability to perceive emotion from faces, match emotions to physical sensations, identify the cause of emotions, and identify effective ways to manage emotions and was designed for youth ages 10 to 17 years old (Rivers et al., 2012). Both of these assessments sample social-emotional comprehension broadly. Each is appropriate for a specific sex, age range, or clinical population.

Thus, existing direct assessments vary in the domain coverage, age range, and population for which they are appropriate. As a result, depending on a child's age, sex, and presenting issue, appropriate social-emotional comprehension assessments may not be available. With a limited range of tools suitable for the direct clinical assessment of social-emotional comprehension, clinicians may not be able to come to a fully informed understanding of contributors to social impairment in many clinical populations. Building on prior work, this study evaluated the psychometric properties of a group of social-emotional comprehension assessments in general education students and children with neurobehavioral disorders, including attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), and reading disorder (RD), across a wide age range (5 to 14 years). The assessments are designed to measure children's ability to encode, interpret, and reason about social-emotional information. Evaluating the psychometric properties of assessments in two distinct samples with a broad age range permits us to test the potential clinical usefulness of the assessments across diverse groups.

A Model of Social-Emotional Comprehension

Lipton and Nowicki (2009) proposed the Social Emotional Learning Framework (SELF) to conceptualize social-emotional comprehension. In this context "social-emotional" refers to a mixture of social and emotional phenomena. The SELF incorporates features of prominent theories of social-emotional comprehension, including social neuroscience (Adolphs, 2003), social information processing (Crick & Dodge, 1994), affective social competence (ASC; Halberstadt, Denham, & Dunsmore, 2001), and emotional intelligence (EI; Salovey & Mayer, 1989–1990). The SELF emphasizes three broad constructs that each resemble features of other models.

First, encoding of social-emotional information, or *Social Awareness*, is defined as the ability to label others' emotions from nonverbal cues. Social Awareness draws upon research on nonverbal communication (Nowicki & Duke, 1994) and is similar to "emotion perception" from EI (Salovey & Mayer, 1989–1990) and "receiving affective messages" from ASC (Halberstadt et al., 2001). Second, interpreting others' perspectives from their behavior and language, or *Social Meaning,* draws from research that finds that children's theory of mind and pragmatic judgment both involve interpreting intentions underlying others' words and actions (Capps, Kehres, & Sigman, 1998; Landa, 2000; McKown, 2007; Peterson, Wellman, & Liu, 2005; Wellman & Liu, 2004). Third, *Social Reasoning* is defined as the ability to reason about social problems and draws from research on social information-processing (Bauminger et al., 2005; Crick & Dodge, 1994; Denham, 2006; Dirks, Treat, & Weersing, 2007).

Empirical evidence supports the conclusion that these three dimensions of social-emotional comprehension are important. When examined separately, Social Awareness, Social Meaning, and Social Reasoning are each associated with social outcomes including peer status, interpersonal negotiating skill, aggression, and social withdrawal (Banerjee & Watling, 2005; Chung & Asher, 1996; Dodge & Price, 1994; Erdley & Asher, 1999; Hughes & Ensor, 2007; Nowicki & Duke, 1994; Slaughter, Dennis, & Pritchard, 2002; Yeates, Schultz, & Selman, 1991; Yeates, Schultz, & Selman, 1990). Work examining Social Awareness, Social Meaning, and Social Reasoning *together* suggests that the better children perform on assessments of several dimensions of social-emotional comprehension, the more others report they engage in socially competent behavior (McKown et al., 2009).

Hypotheses

A main goal of this study was to demonstrate the psychometric properties of a group of direct assessments of social-emotional comprehension. To this end, we tested several hypotheses. First, we hypothesized that scores on a group of social-emotional comprehension assessments would exhibit high internal consistency, test–retest, and interrater reliabilities. Second, we hypothesized that assessments selected to reflect Social Awareness, Social Meaning, and Social Reasoning would reflect three separate but correlated latent variables. We tested this hypothesis using confirmatory factor analysis (CFA). Finally, we hypothesized that Social Awareness, Social Meaning, and Social Reasoning latent variables would each be more strongly associated with corresponding latent variables created using alternate measures (convergent validity) than with latent variables reflecting different dimensions of social-emotional comprehension (divergent validity). We tested this hypothesis using structural equation modeling (SEM; Campbell & Fiske, 1959; Messick, 1995).

Another goal was to determine whether the assessments yielded expected age and diagnostic differences, which would support their validity and clinical relevance. We hypothesized that (a) greater age would be associated with better performance on assessments of Social Awareness, Social Meaning, and Social Reasoning skills (Izard & Harris, 1995; Nowicki & Duke, 1994; Pillow, 2011; Wellman, Cross, & Watson, 2001; Zuckerman, Blanck, DePaulo, & Rosenthal, 1980); (b) compared to typical peers, children with ASD, ADHD, and RD would demonstrate significantly poorer Social Awareness, Social Meaning, and Social Reasoning skills; and (c) children with ASD would demonstrate greater social-emotional comprehension deficits than children with ADHD or RD (Channon, Charman, Heap, Crawford, & Rios, 2001; Embregts & van Nieuwenhuijzen, 2009; Hall & Richmond, 1985; Mikami, Lee, Hinshaw, & Mullin, 2008; H. L. Swanson & Malone, 1992; Tur-Kaspa & Bryan, 1994).

Method

Recruitment and Sample

The Rush University Medical Center Institutional Review Board (IRB) approved all study procedures and consent processes. All but two of the assessments were administered by research staff with bachelor's degrees or higher who were experienced in work-

ing with children and who were supervised by senior members of the research team. The measures used for quantifying autism characteristics were administered by either a clinical psychologist or research psychologist who was reliable in the administration and scoring of those measures.

Eligibility. All general education students in kindergarten through eighth grade at an urban parochial school and a suburban public school were invited to participate.

In addition, children with a diagnosis of ASD, ADHD, or RD who had average or above average cognitive abilities (full scale IQ $\geq$ 85) as measured by the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) were eligible to participate. When no prior diagnosis had been documented, ASD was operationalized as a score at or above 12 on the Social Communication Questionnaire (SCQ; Rutter, Bailey, Lord, & Berument, 2003) and scores above the diagnostic cutoff for autism spectrum disorder on the Autism Diagnostic Interview-Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994) and the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 1999). Children who had a prior diagnosis of ASD were required to score above the diagnostic cutoff for ASD on either the ADI-R or ADOS.

ADHD was operationalized as scoring in the clinical range on one teacher and one parent report of inattention or hyperactivity on either the Behavior Assessment System for Children (2nd ed.; BASC-2; Reynolds & Kamphaus, 2004) or the Swanson, Nolan and Pelham Scale (4th ed.; SNAP-IV; J. M. Swanson, 1995) *and* a history of difficulties before age 7 reported by a parent on the Kiddie-Schedule for Affective Disorders and Schizophrenia (K-SADS; Kaufman et al., 1997).

RD was operationalized as at least one core reading skill score on the Wechsler Individual Achievement Test (2nd ed.; WIAT-II; Wechsler, 2005) at least one standard deviation below full scale IQ on the WASI *and* a score at least one standard deviation below average on the phonological awareness subtest from the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999).

Recruitment and informed consent. For the school sample, informational parent meetings were held as part of back-to-school nights at the beginning of the school year. A consent form along with a letter from the principal was mailed to all parents. The clinic sample was recruited from a suburban outpatient mental health center associated with the research study, and from schools in the surrounding area and through therapists, physicians, and other professionals in the area working with the study population. Flyers were posted in the waiting rooms of the recruitment sites, e-mailed to parents in a clinical research data base, e-mailed to professionals in the community, and advertised in a newsletter. Parent informed consent and child assent were obtained for all participants.

Sample characteristics. Parents of 186 general education students ages four to 14 years old (M = 8.8 years, SD = 2.6 years) consented to have their children participate. One child was under 5 years old (4.9 years). The school sample included 84 boys (45.2%). Eighty children were White (43%), 53 were Black (28.5%), 23 were Latino (12.4%), 12 were Asian (6.5%), and 18 identified as multiracial (9.7%). The 119 participants in the clinic sample ranged from ages 5 to 14 (M = 10.3 years old, SD = 2.2 years). In total, 44 met criteria for ASD (36.7%), 46 met criteria for ADHD (38.3%), and 24 met criteria for RD (24.2%). The clinic sample included 96 boys (80.7%). Ninety children in this sample were White (75.6%), seven were Black (5.9%), two were Latino (1.7%), three were Asian (2.5%), and three were multiracial (2.5%). Fourteen children identified their ethnicity as "other" or did not indicate their ethnicity (11.8%).

The school sample included more girls than the clinic sample, $\chi^2(1) = 37.6$, $p < .05$, was more ethnically diverse, $\chi^2(5) = 65.7$, $p < .05$, and was older, $F(1, 291) = 243$, $p < .05$. On the Social Skills Rating System (SSRS; Gresham & Elliott, 1990), teachers reported that children in the school sample displayed average levels of socially skilled and problem behavior (Social Skills Standard Score = 100.0; Problem Behavior Standard Score = 100.2). In contrast, teachers reported that children in the clinic sample were below average in social skills and above average in problem behaviors (Social Skills Standard Score = 89.3; Problem Behavior Standard Score = 110.5). Sample differences in social skills and problem behaviors were significant ($F(1, 256) = 28.7$, $p < .05$, $d = 0.69$ for Social Skills, and $F(1, 256) = 35.4$, $p < .05$, $d = -0.76$ for Problem Behaviors). In addition, on a three-item questionnaire, teachers reported that children in the school sample were significantly more well accepted by peers than children in the clinic sample, $F(1, 260) = 17.3$, $p < .05$, $d = 0.53$.

Procedures

To minimize fatigue, more and less demanding assessments were alternated and were administered in the same order for all participants. School sample participants completed social-emotional comprehension assessments individually at their school. Each child participated in an average of two and one half hours of testing, broken into two or three sessions on different days. Breaks were offered on an as-needed basis to prevent testing fatigue. Participants from the clinic sample were tested individually at an outpatient clinic. Testing was broken into two sessions. In the first session, which lasted approximately 2 hr for children suspected of ADHD or RD diagnoses and up to 3 hr for children suspected of an ASD diagnosis, assessments were administered to confirm diagnostic eligibility. In the second session, which lasted approximately 2.5 hr, social-emotional comprehension assessments were administered.

Core Social-Emotional Comprehension Measures

Core measures were selected for which prior research had provided evidence of good reliability and validity.

Social Awareness. Four measures were used to assess Social Awareness. First, for the 24-item child faces subtest of the Diagnostic Analysis of Nonverbal Accuracy (DANVA; Nowicki & Duke, 1994), children viewed photographs of child faces and indicated whether each child was happy, sad, angry, or scared. Second, children completed the 22-item Match Emotional Prosody to Emotional Face (MEPEF) subtest of the Comprehensive Affect Testing System (CATS; Weiner, Gregory, Froming, Levy, & Ekman, 2006). Per CATS standard administration procedures, for each item, children listened to an audio recording of an adult making a statement. Children then selected one of five faces that displayed the same emotion conveyed by the speaker's tone of voice. Third, children completed a 22-item posture recognition task (Heberlein, Gläescher, & Adolphs, 2007). For this task, children viewed photographs

of adults with blurred faces and indicated whether the person was happy, sad, angry, or scared. Finally, children completed a 20-item point-light display task (Heberlein, Adolphs, Tranel, & Damasio, 2004). For this task, children viewed brief video clips of abstracted human forms walking. For each clip, children indicated whether the person was happy, sad, angry, or scared. Children received one point for each correct response.

Social Meaning. One theory of mind and one pragmatic judgment measure were used to assess Social Meaning. Theory of mind was assessed in children in kindergarten through fifth grade using 12 vignettes from Strange Stories (Happé, 1994; White, Hill, Happé, & Frith, 2009). In each vignette, a character states one thing but means something else. Respondents are asked whether what the character said is true and why the character said what he or she said. Children received 1 point for correctly inferring the speaker's intention or mental state. To assess pragmatic judgment, the 60-item pragmatic judgment subtest of the Comprehensive Assessment of Spoken Language (CASL; Carrow-Woolfolk, 1999) was administered to all participants. For this test, children were asked what they would do or say in a range of situations with social language demands, including greetings, requesting information, expressing sympathy, joining a conversation, and polite interruption. Scoring on the CASL followed standard procedures.

Social Reasoning. Five social problem-solving vignettes involving peer entry, peer pressure, peer provocation, and differences of opinion were administered to assess Social Reasoning (McKown et al., 2009). Vignettes were read aloud and children's answers were written down verbatim. Independent raters coded children's responses. For problem identification, children were asked to define the problem. Responses were awarded two points when the child accurately described the problem and referred to a consequence (e.g., "They are not talking to me and I feel sad."), one point when the child accurately described the problem but did not mention a consequence (e.g., "They are not talking to me."), and zero points were given when the child gave an inaccurate or implausible response. For goal generation, children were asked "How would you like things to turn out?" Children were awarded one point for a prosocial goal (e.g., "I want us to become friends.") and zero points for all other goals (e.g., "I want him to get in trouble"). Children were then asked to generate possible solutions. Each assertive, competent solution was awarded a point (e.g., "I would walk up and say hi."), while all other responses (aggressive, passive-avoidant or 3rd party intervention) were scored a "0." A final raw score for each child for each code was the average score across raters and vignettes.

Alternate Measures of Social-Emotional Comprehension

To aid in establishing construct validity, alternate child measures that reflected Social Awareness, Social Meaning, and Social Reasoning were selected and administered.

Social Awareness. The Davis Set of Emotion Expressions (Tracy & Robins, 2004) served as an alternate measure of Social Awareness. Children viewed photographs and indicated whether the emotion expressed by the person in each photograph matched a target emotion word (happy, sad, angry, or afraid). Eighteen photographs, reflecting varied emotion displays, were presented in random order for each target emotion. Children received one point each time they correctly indicated whether a photograph reflected the target emotion.

Social Meaning. The first 15 items of the NEPSY-II ToM (Korkman et al., 2007) served as an alternate measure of Social Meaning and was administered to children in kindergarten through fifth grade. Items measured false belief understanding, nonliteral language comprehension, and appearance-reality distinction.

Social Reasoning. An independent assessment of Social Reasoning was not administered. However, as part of the Social Reasoning task, children were asked to provide as many possible solutions to address the problem presented in each of the five vignettes. The number of solutions generated was not part of our model of social-emotional comprehension but is acknowledged by other theorists as an important indicator of social problem-solving skill (Spivack & Shure, 1974). Accordingly, the total number of solutions generated for each vignette was averaged across the five vignettes and used as an alternate measure of Social Reasoning.

Data for Temporal Stability Analyses

Children who participated in the first year of the school study and who remained at their respective school during the course of the study were invited to participate in each subsequent year. As a result, two waves of data were available for a subset of the school sample ($n = 72$). Children repeating the assessment completed a shortened group of assessments. The average interval between testing was 12.0 months ($SD = 2.0$ months). The DANVA Child Faces and Davis Set of Emotion Expressions were not administered during the subsequent years to reduce the burden on participants. Participants in the clinic sample were not retested.

Results

Descriptive Statistics

Table 1 includes descriptive statistics. Table 2 shows correlations between variables.

Missing Data

Cross-sectional data. In 12 cases from the school sample, because of absences or participant dropout, five or more assessment scores were missing, including all assessment scores from at least one, and often two, dimensions of social-emotional comprehension. Those cases were omitted from subsequent analyses. The remaining samples included 174 students in the school sample and 119 in the clinic sample.

Next, the percentage of participants completing each measure in each sample was calculated (Table 1). For 27 of 30 measures, data were collected from more than 95% of participants. However, in the school sample, 74% of children completed the Strange Stories and 70% completed the NEPSY-II ToM. These assessments were not administered to students above fifth grade because of developmentally expected ceiling effects.

Table 1
Description of Assessments

Social-emotional dimension and assessment	Items	School sample					Clinic sample				% Available	
		n	α	Stability	*M*	*SD*	*n*	α	*M*	*SD*	School	Clinic
Social Awareness												
Awareness composite			.95					.96				
DANVA Accuracy	24	174	.71		82.2	12.4	118	.70	78.4	13.4	100	99
CATS MEPEF	22	174	.67	.70	49.6	15.8	118	.72	51.8	17.5	100	99
Postures Accuracy	24	173	.80	.42	80.7	16.5	118	.82	79.2	17.9	99	99
PLW	20	173	.60	.33	65.4	15.4	118	.64	64.8	16.4	99	99
Social Meaning												
Meaning composite			.97					.97				
Strange Stories	12	128	.74	.64	5.2	2.8	119	.75	6.2	2.9	74	100
CASL Pragmatics	60	174	.96	.76	41.3	16.1	109	.94	45.5	14.3	100	92
Social Reasoning												
Reasoning composite			.95					.90				
Problem Identification	5	169	.61	.24	1.1	0.3	115	.50	1.1	0.3	98	100
Goal Quality	5	174	.69	.17	0.8	0.2	114	.49	0.8	0.2	97	100
No. Competent Solutions	5	169	.72	.31	1.4	0.6	108	.59	1.2	0.5	98	100
Criterion Measures												
Davis composite			.97					.96				
Davis Happiness	18	174	.64		84.9	10.6	119	.64	82.4	11.5	100	100
Davis Sadness	18	174	.76		85.3	15.3	119	.72	83.0	14.9	100	100
Davis Anger	18	174	.83		90.3	14.8	119	.75	88.4	13.8	100	100
Davis Fear	18	174	.82		86.4	16.8	119	.83	84.4	18.3	100	100
NEPSY-II ToM	15	122	.74	.67	15.4	4.0	113	.74	17.3	3.8	70	95
No. Solutions	5	169	.84	.24	2.3	0.8	108	.75	2.1	0.8	97	100

Note. DANVA = Diagnostic Analysis of Nonverbal Accuracy (Nowicki & Duke, 1994); CATS MEPEF = Match Emotional Prosody to Emotional Face subtest of the Comprehensive Affect Testing System (Weiner, Gregory, Froming, Levy, & Ekman, 2006); PLW = point-light walker; CASL = Comprehensive Assessment of Spoken Language (Carrow-Woolfolk, 1999); NEPSY-II ToM = NEPSY-II theory of mind (Korkman et al., 2007).

Controlling for age, children who were missing Strange Stories were slightly better at defining social problems, 1.2 versus 1.0, $\eta^2 = .03$, $F(1, 164) = 4.16$, $p < .05$, and produced slightly more competent potential solutions to hypothetical social problems, 1.5 versus 1.2, $\eta^2 = .03$, $F(1, 164) = 4.19$, $p < .05$. As with Strange Stories, controlling for age, children who were missing NEPSY-II ToM data were slightly better at defining social problems, 1.2 versus 1.0, $\eta^2 = .02$, $F(1, 164) = 3.96$, $p < .05$, and produced

Table 2
Zero-Order Correlations Between Variables

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Age	—	.37*	.43*	.50*	.42*	.53*	.63*	.39*	.46*	.33*	.29*	.34*	.29*	.29*	.55*	.05
2. DANVA	.46*	—	.32*	.59*	.39*	.33*	.46*	.31*	.20*	.23*	.34*	.36*	.36*	.37*	.37*	−.06
3. CATS MEPEF	.53*	.49*	—	.47*	.26*	.40*	.47*	.28*	.24*	.24*	.31*	.36*	.31*	.36*	.29*	.14
4. Posture	.42*	.60*	.40*	—	.46*	.49*	.55*	.31*	.31*	.12	.39*	.43*	.50*	.54*	.53*	−.06
5. PLW	.46*	.45*	.38*	.48*	—	.35*	.43*	.33*	.31*	.17†	.21*	.32*	.41*	.33*	.31*	.08
6. Strange Stories	.48*	.35*	.37*	.42*	.44*	—	.70*	.50*	.42*	.37*	.34*	.34*	.37*	.38*	.63*	.26*
7. CASL	.77*	.51*	.54*	.51*	.53*	.79*	—	.59*	.56*	.51*	.35*	.35*	.37*	.43*	.78*	.34*
8. Prob ID	.27*	.28*	.18*	.28*	.27*	.39*	.39*	—	.44*	.47*	.26*	.17†	.22*	.26*	.39*	.28*
9. Goal Quality	.26*	.21*	.21*	.13†	.19*	.27*	.30*	.10	—	.34*	.24*	.27*	.24*	.32*	.39*	.03
10. Competent Sol	.52*	.33*	.41*	.32*	.34*	.49*	.60*	.49*	.21*	—	.13	.20*	.27*	.26*	.37*	.63*
11. Davis Happy	.45*	.37*	.39*	.41*	.29*	.39*	.50*	.21*	.08	.24*	—	.35*	.36*	.35*	.33*	.06
12. Davis Sad	.32*	.39*	.33*	.42*	.34*	.34*	.40*	.13†	.20*	.22*	.25*	—	.58*	.51*	.30*	.06
13. Davis Angry	.32*	.36*	.32*	.35*	.32*	.34*	.42*	.15†	.22*	.26*	.35*	.54*	—	.61*	.33*	.07
14. Davis Afraid	.40*	.36*	.36*	.44*	.52*	.29*	.46*	.29*	.19*	.29*	.41*	.37*	.42*	—	.39*	.04
15. NEPSY-II ToM	.56*	.30*	.30*	.32*	.27*	.68*	.72*	.43*	.29*	.58*	.39*	.20*	.34*	.36*	—	.20*
16. # Solutions	.26*	.14†	.19*	.02	.17*	.23*	.34*	.38*	.04	.72*	.16*	−.01	.10†	.14†	.37*	—

Note. DANVA = Diagnostic Analysis of Nonverbal Accuracy (Nowicki & Duke, 1994); CATS MEPEF = Match Emotional Prosody to Emotional Face subtest of the Comprehensive Affect Testing System (Weiner, Gregory, Froming, Levy, & Ekman, 2006); PLW = point-light walker; CASL = Comprehensive Assessment of Spoken Language (Carrow-Woolfolk, 1999); Prob ID = problem identification; Sol = solutions; NEPSY-II ToM = NEPSY-II theory of mind (Korkman et al., 2007). Correlations below the diagonal are from the school sample; correlations above the diagonal are from the clinical sample.
† $p < .10$. * $p < .05$.

slightly lower quality goals in response to hypothetical social problems, 0.9 versus 1.0, $\eta^2 = .05$, $F(1, 164) = 8.53$, $p < .05$. Children who were missing Strange Stories and NEPSY-II ToM did not differ from children who were not missing those measures on any other available measures. Having missing data from other variables was not associated with performance on other measures.

Because data from most variables were missing at random (MAR), confirmatory factor analyses (CFAs) and structural equation models (SEMs) were estimated with full information maximum likelihood (FIML). When data are MAR and variables associated with missingness are included in the model, FIML results are equivalent to stochastic multiple imputation (McCartney, Bub, & Burchinal, 2006).

Temporal Stability. We evaluated the extent to which children in the school sample with two waves of data differed from those with one. In the context of independent samples *t* tests, on 12 of 16 observed variables, there were no significant differences between children who did and did not have two waves of data. Compared to children who participated in the study for only 1 year, children with two waves of data were younger, 8.7 versus 9.7 years old, $t(291) = 2.75$, $p < .05$, and provided more competent potential solutions to hypothetical social problems, 1.5 versus 1.2, $t(275) = 3.10$, $p < .05$. There were no other significant differences between children who did and who did not have two waves of data available.

Reliability

Interrater reliability. Interviewers transcribed children's verbatim responses to Strange Stories during the assessment. A team of four raters independently coded children's verbatim responses to each item. Across the 12 Strange Stories items, average pairwise kappa was .72, and intraclass correlation was .90. A team of five raters coded all social problem-solving responses for two consecutive years of data collection. Average pairwise kappa for problem identification, goal quality, competence of solutions generated, and competence of best responses was .67, .54, .66, .70, and .97, respectively. The intraclass correlation for these codes was .92, .84, .95, .92, and .99, respectively.

Reliability over 1 year. One-year measurement stability is presented in Table 1 for measures and participants for which successive waves of data were available. The average 1-year stability was $r = .45$, with a range of .17 to .76. All stability coefficients were statistically significant ($p < .05$) except for Goal Quality and the number of solutions generated.

Internal consistency. Cronbach's alpha was computed for each assessment and sample (see Table 1). In the school sample, internal consistency reliability ranged from .60 to .96. Ten out of 15 assessments achieved an alpha of .70 or greater and average internal consistency was .75. In the clinical sample, internal consistency reliability ranged from .49 to .94. Ten out of 15 scales achieved an alpha of .70 or greater and average internal consistency was .70. For four of the 14 measures, differences in alphas between samples (Feldt, 1969) were statistically significant ($W = .67$, $p < .05$ for CASL; $W = .61$, $p < .05$ for Goal Quality; $W = .68$, $p < .05$ for the number of competent solutions generated; $W = .68$, $p < .05$ for Davis Anger recognition).

Composite reliability. When the elements of a composite score are correlated, the reliability of the composite score is higher than the average reliability of those elements (Nunnally & Bernstein, 1994). Factor scores for Nonverbal Awareness, Social Meaning, and Social Reasoning were created and standardized and used in all validity analyses. Nunnally and Bernstein (1994, p. 271) defined the reliability of a composite that is a weighted sum as follows:

$$r_{yy} = 1 - \frac{\sum b_i^2\sigma_i^2 - \sum b_i^2\sigma_i^2 r_{ii}}{\sigma_Y^2},$$

where r_{yy} is the reliability of the latent variable, b_i is factor weights of each measure i associated with the latent variable, σ_i^2 is the variance of each measure i, r_{ii} is the reliability of measure i, and σ_Y^2 is the summed variance of the obtained scores.

The reliability of the Social Awareness composite was .95 for the school sample, and .96 for the clinic sample. The reliability of the Social Meaning composite was .97 for the school and clinic samples. The reliability of the Social Reasoning composite was .95 for the school sample, and .90 for the clinic sample. Sample differences in the reliabilities of Social Awareness and Social Meaning scores were not statistically significant ($W = .80$ for Social Awareness, *ns*, $W = 1.00$ for Social Meaning). The sample difference in the reliability of the Social Reasoning composite was statistically significant ($W = .50$, $p < .05$).

Validity

Factor structure. We used Amos (17.0.2; Arbuckle, 2008) to construct and test confirmatory factor analysis (CFA) and structural equation models (SEM). The fit of all models was evaluated with overall χ^2 goodness-of-fit, χ^2/df adjusted goodness-of-fit, the comparative fit index (CFI), the incremental fit index (IFI), and root-mean-square error of approximation (RMSEA). Models were interpreted as a good fit with the data if the χ^2/df adjusted goodness-of-fit statistic was <2, CFI and IFI were both $\geq .90$, RMSEA $\leq .08$ (Browne & Cudeck, 1993), and hypothesized coefficients were significant and in the predicted direction.

We compared three models in each sample. Summary findings can be found in Table 3. First, we tested a one-factor model in which all indicator variables loaded onto a single factor. The fit of this model to the data was marginal. Next, we tested a two-factor model in which the DANVA, CATS MEPEF, posture recognition and point-light display scores loaded on one factor, and Strange Stories, pragmatic judgment, problem identification, goal quality, and the number of competent solutions loaded on a second factor. The fit of the data to this model in both samples was acceptable and was significantly better than the one-factor model. Next, we tested a three-factor model corresponding to our theory of social-emotional comprehension, which is depicted in Figure 1. The data fit the three-factor model was significantly better than the two-factor model in both samples, and, as seen in Table 2, the fit of the data to the model was excellent. Across samples, the three-factor model coefficients were the same valence and similar magnitude. Because the three-factor model fit our conceptualization of social-emotional comprehension and was superior to simpler models in both samples, the remaining analyses used this model. Factor scores reflecting Social Awareness, Social Meaning, and Social Reasoning were saved, standardized, and used in validity analyses focused on age- and diagnostic differences.

Table 3
Goodness-of-Fit Indices of Alternative Models, by Sample

Model	*df*	χ^2	χ^2/df	$\Delta\chi^2(\Delta df)$	IFI	CFI	RMSEA (90% CI)
School sample							
One factor	27	83.5	3.09		0.90	.90	.110 (.084–.137)
Two factor	26	46.1	1.77	29.3 (1)*	0.97	.96	.067 (.033–.098)
Three factor	24	36.4	1.52	9.7 (2)*	0.98	.98	.055 (.000–.089)
Clinic sample							
One factor	27	64.7	2.40		0.90	.89	.068 (.047–.089)
Two factor	26	30.1	1.16	34.6 (1)*	0.99	.99	.023 (.000–.053)
Three factor	24	22.3	0.93	7.8 (2)*	1.01	1.00	.000 (.000–.043)

Note. IFI = incremental fit index; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; CI = confidence interval.
* $p < .05$.

Post hoc analyses also revealed that, in both samples, the data fit a model in which Social Awareness, Social Meaning, and Social Reasoning loaded on a second-order Social-Emotional Comprehension latent variable, IFI/CFI = .98, RMSEA = .055 (.000–.089), $\chi^2(24) = 36.4$, *ns*, $\chi^2/df = 1.52$ for the school sample; and IFI = 1.00, CFI = 1.00, RMSEA = .058 (.029–.082), $\chi^2(25) = 24.8$, *ns*, $\chi^2/df = 0.99$ for the clinic sample. This second-order factor may be useful to achieve some research goals (see McKown et al., 2009). We were interested in the convergent and discriminant validity of Social Awareness, Social Meaning, and Social Reasoning. Accordingly, we opted to examine the three-factor model without consideration of the higher order factor.

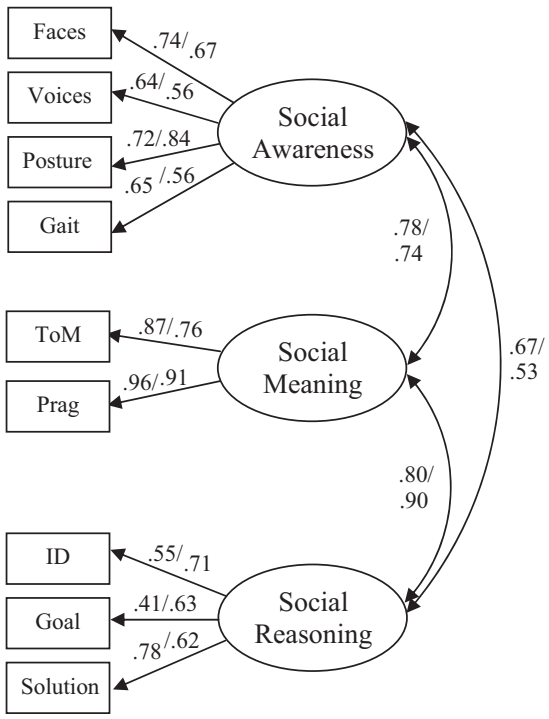

Figure 1. Three-factor confirmatory model of social-emotional comprehension. Note that coefficients are standardized. All values to the left of the "/" are from the school sample; values to the right of the "/" are from the clinic sample. Bold signifies paths and path coefficients reflecting convergent validity. All coefficients are significant at the $p < .05$ level. $\chi^2(24) = 36.4/22.3$, *ns*; $\chi^2/df = 1.52/0.93$; incremental fit index = 0.98/1.01; comparative fit index = .98/1.00; root-mean-square error of approximation (90% confidence interval) = .055 (.000–.089)/.000 (.000–.043). ToM = theory of mind; Prag = pragmatic judgment subtest of the Comprehensive Assessment of Spoken Language; ID = problem identification.

Convergent and discriminant validity. We next evaluated the convergent and discriminant validity of the social-emotional comprehension assessments. For each sample, SEM was used to evaluate the extent to which latent variables reflecting Social Awareness, Social Meaning, and Social Reasoning created with core assessments were more related to parallel latent variables created with alternate assessments than they were related to other latent variables created with alternate assessments. Convergent validity was demonstrated by strong associations between Social Awareness and the alternate Social Awareness latent variable, between Social Meaning and the alternate Social Meaning latent variable, and between Social Reasoning and the alternate Social Reasoning latent variable. Discriminant validity was demonstrated by weaker associations between Social Awareness and the alternate Social Meaning and Social Reasoning latent variables, between Social Meaning and the alternate Social Awareness and Social Reasoning latent variables, and between Social Reasoning and the alternate measures of Social Awareness and Social Meaning latent variables.

SEMs corresponding to this model were fit separately to the school sample data and the clinic sample data. In both cases, the overall fit of the model to the data was very good, IFI/CFI = .96, RMSEA = .055 (.035–.073), $\chi^2(81) = 126.0$, $p < .05$, $\chi^2/df = 1.56$ for the school sample; and IFI/CFI = .95, RMSEA = .058 (.029–.082), $\chi^2(81) = 113.4$, $p < .05$, $\chi^2/df = 1.40$ for the clinic sample. These models are depicted in Figure 2. Inspection of model paths revealed that in both samples, the structural coefficients linking parallel latent variables with alternate assessments were large and statistically significant. Furthermore, most of the structural coefficients linking latent variables in different domains were nonsignificant. Those that were statistically significant were nevertheless much smaller in magnitude than coefficients linking parallel latent variables.

Age differences in social-emotional comprehension. As shown in Table 2, performance on all social-emotional comprehension assessments were significantly and positively associated with age in both samples. Not surprisingly, therefore, as illustrated in Figure 3, factor scores reflecting Social Awareness, Social Meaning, and Social Reasoning were positively associated with age in both samples. What is clear from this

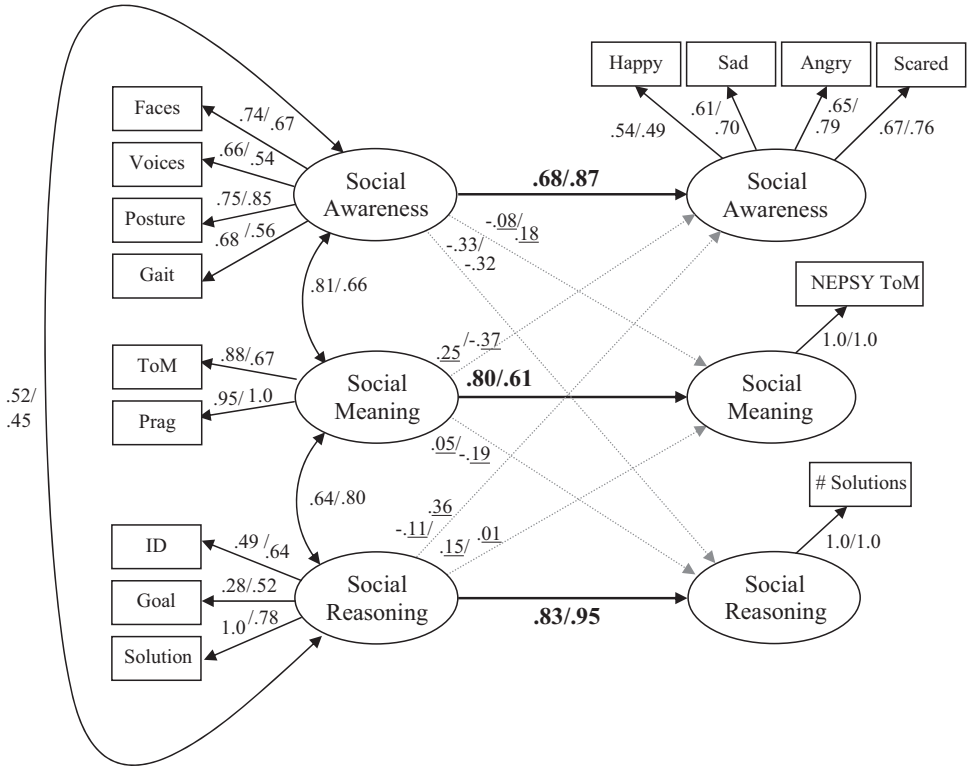

Figure 2. Convergent and discriminant validity social-emotional comprehension assessments. Note that coefficients are standardized. All values to the left of the "/" are from the school sample; values to the right of the "/" are from the clinic sample. Bold signifies paths and path coefficients reflecting convergent validity. All coefficients are significant at the $p < .05$ level *except* underlined values, which are nonsignificant. $\chi^2(81) = 126.0/113.4$, $p < .05$; $\chi^2/df = 1.56/1.40$; incremental fit index = .96/.96; comparative fit index = .96/.95; root-mean-square error of approximation (90% confidence interval) = .055 (.035–.073)/.058 (.029–.082). ToM = theory of mind; Prag = pragmatic judgment subtest of the Comprehensive Assessment of Spoken Language; ID = problem identification.

figure is that the slope of age-related changes in social-emotional comprehension are very similar in school and clinic samples, with higher mean scores in the school sample than in the clinic sample.

The plot of social-emotional comprehension across age suggests a steeper curve at younger ages, gradually flattening with age. Post hoc analyses evaluated the extent to which age-related changes in domain scores was curvilinear for each factor score reflecting Social Awareness, Social Meaning, and Social Reasoning. Specifically, each factor score was regressed on age and an age-squared quadratic term. Analyses were run separately by dimension of social-emotional comprehension and sample. If the quadratic term was significant, this was interpreted as evidence for the presence of a curvilinear relationship between age and the outcome measure in question. Additional analyses were run adding participant characteristics that differed across groups (sex, ethnicity, SSRS Social Skills and Problem Behavior) as covariates. In all cases, findings were unchanged. We therefore maintained the simpler models reported in Table 4.

Diagnostic differences in social-emotional comprehension. To further evaluate validity, we compared children from different diagnostic groups on standardized factor scores reflecting Social Awareness, Social Meaning, and Social Reasoning. Age-adjusted standardized means and standard errors are reflected in Figure 4. Controlling for age, children in the school sample scored significantly higher than children in the clinic sample on the composite measures of Social Awareness, Social Meaning, and Social Reasoning, $F(1, 289) = 16.9$, $p < .05$ for Social Awareness; $F(1, 289) = 25.0$, $p < .05$ for Social Meaning; and $F(1, 289) = 19.5$, $p < .05$ for Social Reasoning. Pairwise comparisons also revealed that children with ASD performed worse than typically developing children and worse than children with either ADHD or RD on composite measures of Social Awareness, Social Meaning, and Social Reasoning. Additional analyses were run adding participant characteristics that differed across groups (sex, ethnicity, SSRS Social Skills and Problem behavior scores) as covariates. In all cases, findings were unchanged. We therefore used the simpler models reported above and depicted in Figure 4.

Discussion

Direct assessments of social-emotional comprehension are critical elements in the clinical toolkit, enabling assessors to evaluate, identify, and conceptualize contributors to social impairment. This study contributes to a growing body of literature finding that direct

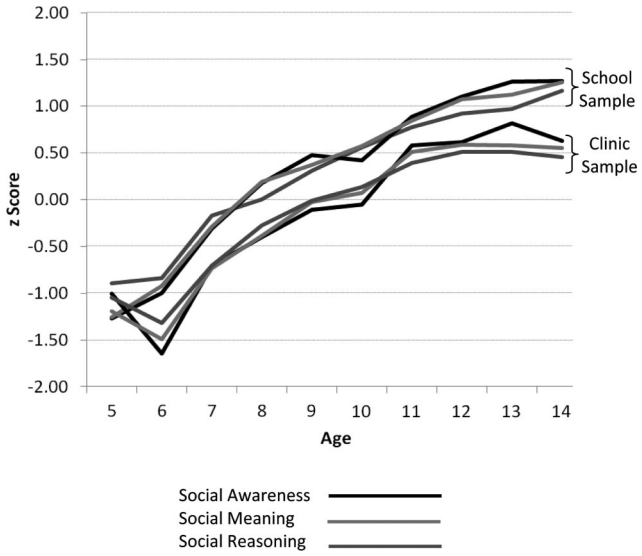

Figure 3. Relationship between age (in years) and social-emotional comprehension.

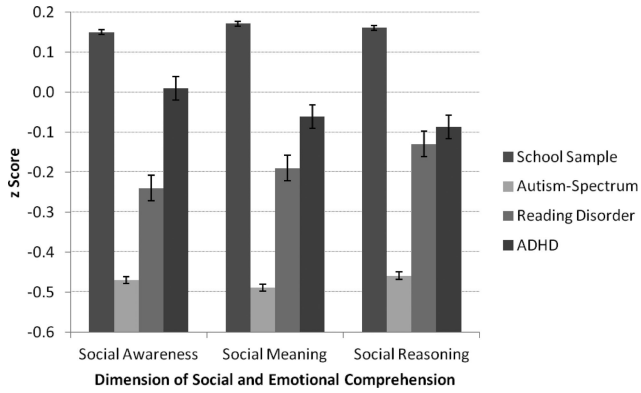

Figure 4. Social-emotional comprehension, by sample and diagnostic group. Error bars indicate one standard error. ADHD = attention-deficit/hyperactivity disorder.

assessment of social-emotional comprehension is feasible, that direct assessments yield reliable scores, and that the scores obtained from direct assessments measure what they are designed to measure. We found that (a) direct assessments of social-emotional comprehension reflected three latent variables that we called Social Awareness, Social Meaning, and Social Reasoning; (b) the reliability of the factor scores was excellent; (c) the latent variables demonstrated convergent and discriminant validity; and (d) the assessments differentiated children from different age and diagnostic groups. Further development is needed to increase the usability and feasibility of assessments that cover a wide age range, offer broad construct coverage, and yield subtest scores with greater reliability.

Relationship to Existing Theory

Findings from this study are consistent with and extend theoretical models of social-emotional comprehension. Consistent with Lipton and Nowicki (2009), this study suggests that three broad dimensions—Social Awareness, Social Meaning, and Social Reasoning—reflect salient aspects of social-emotional comprehension. Undoubtedly, those constructs may be further subdivided and assessed. This study suggests that for investigators and clinicians wishing to obtain a broad assessment of social-emotional comprehension, assessing across these three dimensions is feasible, theoretically and empirically justified, can be done efficiently, and includes important factors that affect children's social behavior. Thus, the present study, and the model of social-emotional comprehension that it supports, offers a straightforward and empirically supported heuristic with which investigators may organize their research hypotheses and clinicians may gather data to develop case formulations.

Psychometrics of Social-Emotional Comprehension Assessment

This study demonstrated that direct assessments of Social Awareness, Social Meaning, and Social Reasoning yield reliable composite scores. This is consistent with prior research reporting the reliability of direct social-emotional comprehension assessments that measure Social Awareness, Social Meaning, or Social Reasoning (Banerjee & Watling, 2005; Hughes & Ensor, 2007; Kupersmidt et al., 2011; McKown et al., 2009; Nowicki & Duke, 1994; Slaughter et al., 2002; Yeates et al., 1991). Study findings also support the conclusion that direct assessment of social-emotional comprehension is valid for its intended purpose. CFA and SEM analyses both found that Social Awareness, Social Meaning, and Social Reasoning are distinct, but correlated con-

Table 4
The Relationship Between Age and Social-Emotional Comprehension Factor Scores, by Sample

	Criterion					
	School sample			Clinic sample		
Parameter	Social Awareness	Social Meaning	Social Reasoning	Social Awareness	Social Meaning	Social Reasoning
Intercept	−4.05*	−5.73*	−5.84*	−6.13*	−6.80*	−6.42*
Age *B*	0.65*	0.97*	1.00*	0.94*	1.11*	1.07*
Age^2 *B*	−0.02*	−0.03*	−0.04*	−0.03*	−0.04*	−0.04*
R^2	.49*	.62*	.57*	.41*	.49*	.42*

Note. Parameter estimates are unstandardized regression coefficients.
* $p < .05$.

structs that demonstrate convergent and discriminant validity. Those constructs in turn reflect a higher order social-emotional comprehension skill. Moreover, the present study provides evidence that measuring broadly across distinct dimensions of social-emotional comprehension can yield findings that may more comprehensively reflect an individual child's strengths and needs than assessments that are more narrowly tailored.

The psychometric properties of the social-emotional comprehension assessments were consistent in two very different samples of children—one from a general education setting and one from a clinic, lending confidence to the conclusion that the social-emotional comprehension scores are reliable, and that the assessments demonstrate robust construct validity. In addition, direct assessments discriminated children from different diagnostic and age groups. As expected, the school sample performed better than the clinic sample. Furthermore, children with autism-spectrum disorders performed worse than children with ADHD or RD. This is consistent with previous research documenting severe social-emotional comprehension deficits in children with ASD (e.g., Bauminger, 2002; Capps, Yirmiya, & Sigman, 1992; Channon et al., 2001; Clark, Winkielman, & McIntosh, 2008; Dodd, Ocampo, & Kennedy, 2011; Embregts & van Nieuwenhuijzen, 2009; Flood, Julian Hare, & Wallis, 2011; Mazefsky & Oswald, 2007; Meyer, Mundy, Van Hecke, & Durocher, 2006; Volden, Mulcahy, & Holdgrafer, 1997). These findings also mirror work suggesting less severe social deficits among children with ADHD and RD (Hall & Richmond, 1985; Jackson, Enright, & Murdock, 1987; Nowicki & Duke, 1994; Sciberras, Ohan, & Anderson, 2012; H. L. Swanson & Malone, 1992; Tur-Kaspa & Bryan, 1994).

Limitations and Future Directions

Construct representation. Social behavior and social functioning are influenced by an array of individual and situational factors, many of which were beyond the scope of the present study and were thus not examined. This study presented strong evidence of the validity of social-emotional comprehension assessments covering a conceptually coherent set of dimensions. Nevertheless, alternative models are clearly possible. Further research should explore other factors that influence social behavior and functioning, such as self regulation.

Sample size. This studies included two relatively small samples (n = 174 and n = 119 in the school and clinic samples, respectively) that substantially differed in characteristics such as sex and racial-ethnic composition. Despite the small sample sizes, in both studies, the fit of the data to the models was excellent and the coefficients were robust, in the predicted direction, and of comparable magnitude. Similar findings in two very different samples suggest that these findings are robust and generalizable to children with a broad array of social-emotional challenges. Future research with larger samples will permit multiple-group analyses to evaluate whether the model coefficients differ between groups.

Measurement. Certain measurement limitations from this study leave open questions for future research. For example, longitudinal data were only available in the school sample and only for a subset of measures. This limits our understanding of the temporal stability of social-emotional comprehension in children with neurobehavioral disorders. In addition, the alternative measure for Social Reasoning was taken from the same measure as our core assessments. Thus, the strong relationship between the focal Social Reasoning latent variable and the alternate Social Reasoning latent variable may be inflated by common method variance. Therefore, the convergent validity findings regarding Social Reasoning should be interpreted with caution. Nevertheless, Social Reasoning did demonstrate good evidence of discriminant validity.

The internal consistency of individual assessments was variable. The average internal consistency was in the mid .70s, and some measures were much lower than this. All scores were retained because of their importance to our model of social-emotional comprehension. In contrast, the reliabilities of the composite factor scores was generally well above .90. On a practical level, this means that scores from most of the individual assessments are insufficiently reliable for interpreting the strengths and needs of individual children from those scores and are certainly too low for high-stakes decision making. At the composite level, reliabilities are adequate for interpreting individual student scores. An important limitation of these assessments, therefore, is that to obtain reliable estimates of social-emotional comprehension requires the administration and complex aggregation of many individual assessments, which reduces the usability and utility of such assessments in real-world practice. Further assessment development is needed to increase the consistency of measurement and reduce the number of measures required to achieve this goal. Until that development occurs, interpretation of individual student social-emotional comprehension test scores should proceed with caution.

The internal consistency of some measures and composite scores differed between samples. Future work on the development of direct assessments should be particularly mindful of the potential for measurement reliability to be lower in clinical populations and should seek to identify and minimize the sources of unreliability. In the present study, age and diagnostic differences were tested on composite scores, not on scores on individual assessments. At the composite level, sample differences between estimated reliabilities were low, with a maximum difference of .07. The likely impact of this difference is that it may have made it more difficult to detect differences between samples. We conclude, therefore, that group differences were robust.

Temporal stability analyses suggest that, over the course of 12 months, general education children generally score consistently on direct assessments of social-emotional comprehension. However, reliability coefficients over this span varied greatly and were generally lower than desirable for clinical measures. Future work should examine the temporal stability over a shorter time span to more clearly evaluate the test–retest reliability of these assessment strategies.

A final limitation is that this study did not address the incremental validity of these assessments, over and above other variables that may be more readily assessed in clinical practice. If, for example, the relationship between social-emotional comprehension and relevant outcomes were mediated entirely by child IQ, then these assessment approaches would not demonstrate incremental criterion-related validity and, as a result, would be of limited value. It is reassuring that prior research has suggested that social-emotional comprehension assessments account for variance in outcomes, over and above IQ (McKown, 2007; Rivers, Brackett, & Salovey, 2008). Furthermore, prior research has found that a higher order factor reflecting Social Awareness, Social Meaning, and Social Reasoning is more strongly associated with teacher

report of social behavior than any one of those factors alone (McKown et al., 2009). Further research is necessary to fully evaluate the incremental validity of these assessment approaches.

Complexity. There are practical limitations of the assessments examined in this study. They require varying procedures and level of training. Even a shortened group of assessments takes more than 1 hr to administer. The existence and quality of norms varies, so interpreting scores is challenging. As a result, these assessment approaches may be useful in the context of clinical practice when the clinician has obtained sufficient training to administer, score, and interpret findings. As assessment development moves forward, clinicians are advised to use the best contemporary standardized assessments designed to measure important dimensions of social-emotional comprehension in children and youth. Assessments such as the NEPSY-II ToM and Affect Recognition subtests, the SIP-AP, MSCEIT-YV, and the Social Language Development Test are examples of currently available instruments.

Conclusion

Clinicians and researchers can use direct assessments of social-emotional comprehension to understand the strengths and limitations that affect each child's social relationships. The evidence presented in this article suggests that direct assessment, used in conjunction with other well-validated strategies, such as behavior rating scales, may provide important and clinically useful information about factors that affect social relationships. In addition to being able to characterize strengths and limitations, it is important that assessments be able to guide treatment planning. Accordingly, an important next step will be to investigate strategies for linking the careful assessment of social-emotional comprehension to intervention strategies that help children develop the skills they need to be successful.

References

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience, 4,* 165–178. doi:10.1038/nrn1056

Arbuckle, J. L. (2008). *Amos* (Version 17.0). Chicago, IL: SPSS.

Banerjee, R., & Watling, D. (2005). Children's understanding of faux pas: Associations with peer relations. *Hellenic Journal of Psychology, 2,* 27–45.

Baron-Cohen, S. (1989). The autistic child's theory of mind: A case of specific developmental delay. *Child Psychology & Psychiatry & Allied Disciplines, 30,* 285–297. doi:10.1111/j.1469-7610.1989.tb00241.x

Baron-Cohen, S., Wheelwright, S., Scahill, V., Lawson, J., & Spong, A. (2001). Are intuitive physics and intuitive psychology independent? *Journal of Developmental & Learning Disorders, 5,* 47–78.

Bauminger, N. (2002). The facilitation of social-emotional understanding and social interaction in high-functioning children with autism: Intervention outcomes. *Journal of Autism and Developmental Disorders, 32,* 283–298. doi:10.1023/A:1016378718278

Bauminger, N., Edelsztein, H. S., & Morash, J. (2005). Social information processing and emotional understanding in children with LD. *Journal of Learning Disabilities, 38,* 45–61. doi:10.1177/00222194050380010401

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Beverly Hills, CA: Sage.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56,* 81–105. doi:10.1037/h0046016

Capps, L., Kehres, J., & Sigman, M. (1998). Conversational abilities among children with autism and children with developmental delays. *Autism, 2,* 325–344. doi:10.1177/1362361398024002

Capps, L., Yirmiya, N., & Sigman, M. (1992). Understanding of simple and complex emotions in non-retarded children with autism. *Child Psychology & Psychiatry & Allied Disciplines, 33,* 1169–1182. doi:10.1111/j.1469-7610.1992.tb00936.x

Carrow-Woolfolk, E. (1999). *Comprehensive assessment of spoken language manual*. Circle Pines, MN: American Guidance Service.

Chan, S., & Mpofu, E. (2001). Children's peer status in school settings: Current and prospective assessment procedures. *School Psychology International, 22,* 43–52. doi:10.1177/01430343010221004

Channon, S., Charman, T., Heap, J., Crawford, S., & Rios, P. (2001). Real-life-type problem-solving in Asperger's syndrome. *Journal of Autism and Developmental Disorders, 31,* 461–469. doi:10.1023/A:1012212824307

Chung, T., & Asher, S. R. (1996). Children's goals and strategies in peer conflict situations. *Merrill-Palmer Quarterly, 42,* 125–147.

Clark, T. F., Winkielman, P., & McIntosh, D. N. (2008). Autism and the extraction of emotion from briefly presented facial expressions: Stumbling at the first step of empathy. *Emotion, 8,* 803–809. doi:10.1037/a0014124

Collins, M., & Nowicki, S. J. (2001). African American children's ability to identify emotion in facial expressions and tones of voice of European Americans. *The Journal of Genetic Psychology: Research and Theory on Human Development, 162,* 334–346. doi:10.1080/00221320109597487

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115,* 74–101. doi:10.1037/0033-2909.115.1.74

Denham, S. A. (2006). Social-emotional competence as support for school readiness: What is it and how do we assess it? *Early Education and Development, 17,* 57–89. doi:10.1207/s15566935eed1701_4

Dirks, M. A., Treat, T. A., & Weersing, V. R. (2007). The situation specificity of youth responses to peer provocation. *Journal of Clinical Child and Adolescent Psychology, 36,* 621–628. doi:10.1080/15374410701662758

Dodd, J. L., Ocampo, A., & Kennedy, K. S. (2011). Perspective taking through narratives: An intervention for students with ASD. *Communication Disorders Quarterly, 33,* 23–33. doi:10.1177/1525740110395014

Dodge, K. A., & Price, J. M. (1994). On the relation between social information processing and socially competent behavior in early school-aged children. *Child Development, 65,* 1385–1397. doi:10.2307/1131505

Duke, M. P., Nowicki, S., & Walker, E. (1996). *Teaching your child the language of social success*. Atlanta, GA: Peachtree.

Embregts, P., & van Nieuwenhuijzen, M. (2009). Social information processing in boys with autistic spectrum disorder and mild to borderline intellectual disabilities. *Journal of Intellectual Disability Research, 53,* 922–931. doi:10.1111/j.1365-2788.2009.01204.x

Erdley, C. A., & Asher, S. R. (1999). A social goals perspective on children's social competence. *Journal of Emotional and Behavioral Disorders, 7,* 156–167. doi:10.1177/106342669900700304

Feldt, L. S. (1969). A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests. *Psychometrika, 34,* 363–373. doi:10.1007/BF02289364

Flood, A. M., Julian Hare, D., & Wallis, P. (2011). An investigation into social information processing in young people with Asperger syndrome. *Autism, 15,* 601–624. doi:10.1177/1362361310387803

Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system manual*. Circle Pines, MN: American Guidance Service.

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

Halberstadt, A. G., Denham, S. A., & Dunsmore, J. C. (2001). Affective social competence. *Social Development, 10,* 79–119. doi:10.1111/1467-9507.00150

Hall, C. W., & Richmond, B. O. (1985). Non-verbal communication, self-esteem and interpersonal relations of LD and non-LD students. *Exceptional Child, 32,* 87–91. doi:10.1080/0156655850320204

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders, 24,* 129–154. doi:10.1007/BF02172093

Heberlein, A. S., Adolphs, R., Tranel, D., & Damasio, H. (2004). Cortical regions for judgments of emotions and personality traits from point-light walkers. *Journal of Cognitive Neuroscience, 16,* 1143–1158. doi:10.1162/0898929041920423

Heberlein, A. S., Gläescher, J., & Adolphs, R. (2007). *Emotion recognition from human postures with masked facial expressions*. Unpublished manuscript.

Hughes, C., & Ensor, R. (2007). Positive and protective: Effects of early theory of mind on problem behaviors in at-risk preschoolers. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 48,* 1025–1032. doi:10.1111/j.1469-7610.2007.01806.x

Izard, C. E., & Harris, P. (1995). Emotional development and developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology, Vol. 1: Theory and methods* (pp. 467–503). Oxford, England: Wiley.

Jackson, S. C., Enright, R. D., & Murdock, J. Y. (1987). Social perception problems in learning disabled youth: Developmental lag versus perceptual deficit. *Journal of Learning Disabilities, 20,* 361–364. doi:10.1177/002221948702000612

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., . . . Ryan, N. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children–Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry, 36,* 980–988. doi:10.1097/00004583-199707000-00021

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY–Second Ed. (NEPSY-II)*. San Antonio, TX: Pearson Assessment.

Kupersmidt, J. B., Stelter, R., & Dodge, K. A. (2011). Development and validation of the social information processing application: A web-based measure of social information processing patterns in elementary school-age boys. *Psychological Assessment, 23,* 834–847. doi:10.1037/a0023621

Landa, R. (2000). Social language use in asperger syndrome and high-functioning autism. In A. Klin, F. R. Volkmar, & S. S. Sparrow (Eds.), *Asperger syndrome* (pp. 125–155). New York, NY: Guilford Press.

Lipton, M., & Nowicki, S. (2009). The social emotional learning framework (SELF): A guide for understanding brain-based social emotional learning impairments. *Journal of Developmental Processes, 4,* 99–115.

Lord, C., Rutter, M., DiLavore, P., & Risi, S. (1999). *Autism Diagnostic Observation Schedule-WPS edition*. Los Angeles, CA: Western Psychological Services.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview—Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24,* 659–685. doi:10.1007/BF02172145

Mavroveli, S., Petrides, K. V., Shove, C., & Whitehead, A. (2008). Investigation of the construct of trait emotional intelligence in children. *European Child & Adolescent Psychiatry, 17,* 516–526. doi:10.1007/s00787-008-0696-6;10.1007/s00787-008-0696-6

Mayer, J. D., Caruso, D. R., & Salovey, P. (2005). *The Mayer-Salovey-Caruso Emotional Intelligence Test–Youth Version (MSCEIT-YV) Research Version*. Toronto, Ontario, Canada: Multi-Health Systems.

Mazefsky, C. A., & Oswald, D. P. (2007). Emotion perception in Asperger's syndrome and high-functioning autism: The importance of diagnostic criteria and cue intensity. *Journal of Autism and Developmental Disorders, 37,* 1086–1095. doi:10.1007/s10803-006-0251-6

McCartney, K., Bub, K. L., & Burchinal, M. (2006). Best practices in quantitative methods for developmentalists: VI. Selection, detection, and reflection. *Monographs of the Society for Research in Child Development, 71,* 105–126.

McKown, C. (2007). Concurrent validity and clinical usefulness of several individually administered tests of children's social-emotional cognition. *Journal of Clinical Child and Adolescent Psychology, 36,* 29–41.

McKown, C., Gumbiner, L. M., Russo, N. M., & Lipton, M. (2009). Social-emotional learning skill, self-regulation, and social competence in typically developing and clinic-referred children. *Journal of Clinical Child and Adolescent Psychology, 38,* 858–871. doi:10.1080/15374410903258934

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741–749. doi:10.1037/0003-066X.50.9.741

Meyer, J. A., Mundy, P. C., Van Hecke, A. V., & Durocher, J. S. (2006). Social attribution processes and comorbid psychiatric symptoms in children with Asperger syndrome. *Autism, 10,* 383–402. doi:10.1177/1362361306064435

Mikami, A. Y., Lee, S. S., Hinshaw, S. P., & Mullin, B. C. (2008). Relationships between social information processing and aggression among adolescent girls with and without ADHD. *Journal of Youth and Adolescence, 37,* 761–771. doi:10.1007/s10964-007-9237-8

Moreno, J. L. (1933). Psychological and social organization of groups in the community. *Proceedings & Addresses. American Association on Mental Deficiency, 38,* 224–242.

Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior, 18,* 9–35. doi:10.1007/BF02169077

Nowicki, S., Duke, M. P., & van Buren, A. (2008). *Starting kids off right*. Atlanta, GA: Peachtree Press.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Parker, J. G., & Asher, S. R. (1987). Peer relations and later personal adjustment: Are low-accepted children at risk? *Psychological Bulletin, 102,* 357–389. doi:10.1037/0033-2909.102.3.357

Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in theory-of-mind development for children with deafness or autism. *Child Development, 76,* 502–517. doi:10.1111/j.1467-8624.2005.00859.x

Pillow, B. H. (2011). *Children's discovery of the active mind: Phenomenological awareness, social experience, and knowledge about cognition*. New York, NY: Springer Science and Business Media.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior assessment system for children* (2nd ed.). San Antonio, TX: Pearson Assessment.

Rivers, S. E., Brackett, M. A., Reyes, M. R., Mayer, J. D., Caruso, D. R., & Salovey, P. (2012). Measuring emotional intelligence in early adolescence with the MSCEIT-YV: Psychometric properties and relationship with academic performance and psychosocial functioning. *Journal of Psychoeducational Assessment, 30,* 344–366. doi:10.1177/0734282912449443

Rivers, S. E., Brackett, M. A., & Salovey, P. (2008). Measuring emotional intelligence as a mental ability in adults and children. In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment* (Vol. 2, pp. 440–460). Los Angeles, CA: Sage. doi:10.4135/9781849200479.n21

Rutter, M., Bailey, A., Lord, C., & Berument, S. K. (2003). *Social communication questionnaire*. Los Angeles, CA: Western Psychological Services.

Salovey, P., & Mayer, J. D. (1989–1990). Emotional intelligence. *Imagination, Cognition and Personality, 9,* 185–211. doi:10.2190/DUGG-P24E-52WK-6CDG

Sciberras, E. F., Ohan, J. F., & Anderson, V. (2012). Bullying and peer victimisation in adolescent girls with attention-deficit/hyperactivity disorder. *Child Psychiatry and Human Development, 43,* 254–270. doi:10.1007/s10578-011-0264-z

Slaughter, V., Dennis, M. J., & Pritchard, M. (2002). Theory of mind and peer acceptance in preschool children. *British Journal of Developmental Psychology, 20,* 545–564. doi:10.1348/026151002760390945

Spivack, G., & Shure, M. B. (1974). *Social adjustment of young children: A cognitive approach to solving real-life problems*. Oxford, England: Jossey-Bass.

Swanson, H. L., & Malone, S. (1992). Social skills and learning disabilities: A meta-analysis of the literature. *School Psychology Review, 21,* 427–443.

Swanson, J. M. (1995). *The Swanson, Nolan, and Pelham Teacher and Parent Rating Scale (SNAP-IV)*. Irvine, CA: Child Development Center, University of California.

Tracy, J. L., & Robins, R. W. (2004). Show your pride: Evidence for a discrete emotion expression. *Psychological Science, 15,* 194–197. doi:10.1111/j.0956-7976.2004.01503008.x

Tur-Kaspa, H., & Bryan, T. (1994). Social information-processing skills of students with learning disabilities. *Learning Disabilities Research & Practice, 9,* 12–23.

Volden, J., Mulcahy, R. F., & Holdgrafer, G. (1997). Pragmatic language disorder and perspective taking in autistic speakers. *Applied Psycholinguistics, 18,* 181–198. doi:10.1017/S0142716400009966

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing*. Austin, TX: PRO-ED.

Walker, H. M., & Severson, H. (1992). *Systematic screening for behavior disorders*. Longmont, CO: Sopris West.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2005). *Wechsler Individual Achievement Test 2nd ed. (WIAT II)*. London, England: Psychological Corporation.

Weiner, S. G., Gregory, A., Froming, K. B., Levy, C. M., & Ekman, P. (2006). *Emotion processing: The comprehensive affect testing system user's manual*. Sanford, FL: Psychology Software.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72,* 655–684. doi:10.1111/1467-8624.00304

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75,* 523–541. doi:10.1111/j.1467-8624.2004.00691.x

White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the strange stories: Revealing mentalizing impairments in autism. *Child Development, 80,* 1097–1117. doi:10.1111/j.1467-8624.2009.01319.x

Yeates, K. O., Schultz, L. H., & Selman, R. L. (1990). Bridging the gaps in child-clinical assessment: Toward the application of social-cognitive developmental theory. *Clinical Psychology Review, 10,* 567–588. doi:10.1016/0272-7358(90)90097-T

Yeates, K. O., Schultz, L. H., & Selman, R. L. (1991). The development of interpersonal negotiation strategies in thought and action: A social-cognitive link to behavioral adjustment and social status. *Merrill-Palmer Quarterly, 37,* 369–405.

Zuckerman, M., Blanck, P. D., DePaulo, B. M., & Rosenthal, R. (1980). Developmental changes in decoding discrepant and nondiscrepant nonverbal cues. *Developmental Psychology, 16,* 220–228. doi:10.1037/0012-1649.16.3.220

Received August 28, 2012
Revision received May 13, 2013
Accepted May 15, 2013 ■