

## Hierarchical Models of Social Competence in Preschool Children: A Multisite, Multinational Study

Brian E. Vaughn, Nana Shin, Mina Kim,  
Gabrielle Coppola, and Lisa Krzysik  
*Auburn University*

António J. Santos, Inês Peceguina,  
João R. Daniel, and Manuela Veríssimo  
*Unidade de Investigação em Psicologia Cognitiva do  
Desenvolvimento e da Educação*

Anthon DeVries, Eric Elphick,  
and Xiomara Ballentina  
*Leiden University*

Kelly K. Bost  
*University of Illinois at Urbana-Champaign*

Wanda Y. Newell  
*Alcorn State University*

Ellaine B. Miller, J. Blake Snider,  
and Byran Korth  
*Auburn University*

The generality of a multilevel factorial model of social competence (SC) for preschool children was tested in a 5-group, multinational sample ( $N = 1,540$ ) using confirmatory factor analysis. The model fits the observed data well, and tests constraining paths for measured variables to their respective first-order factors across samples also fit well. Equivalence of measurement models was found at sample and sex within-sample levels but not for age within sample. In 2 groups, teachers' ratings were examined as correlates of SC indicators. Composites of SC indicators were significantly associated with both positive and negative child attributes from the teachers' ratings. The findings contribute to understanding of both methodological and substantive issues concerning SC in young children.

There is a general consensus among developmental scientists to the effect that social competence (SC) is a crucially important attribute (or attainment) of preschool children because it is presumed to underlie subsequent interpersonal adjustment and school readiness (for a recent treatment of these assumptions, see Ladd, 2005). However, despite this agree-

ment in the abstract concerning the desirability of preschoolers being or becoming socially competent, there is less consensus within the developmental science community concerning how, precisely, SC should be defined at the preschool age or how whatever it may mean to be socially competent might change over developmental time (for conceptual treatments of these questions, see Denham, 2006; Waters & Sroufe, 1983).

The research reported in this article has been supported by Grants SBR95-14563, BCS99-83391, BCS01-26163, and BCS06-23019 from the National Science Foundation; by Grant ACF 90CD0956, and by Grants POCTI/PSI/46739/2002 and PTDC/PSI/66172/2006 from the Portuguese Foundation for Science and Technology. Versions of findings reported here have been presented at the meetings of the Society for Research in Child Development, at the meetings of the International Society for the Study of Behavioral Development, and at the meetings of Society for Research on Human Development from 1995 through 2007. The authors are grateful for the participation and support from the many child-care centers, their directors, teachers, parents, and children who have welcomed observers into their classrooms.

Correspondence concerning this article should be addressed to Brian E. Vaughn, Human Development and Family Studies, 203 Spidle Hall, Auburn University, Auburn, AL 36849. Electronic mail may be sent to [vaughbe@auburn.edu](mailto:vaughbe@auburn.edu).

### *Definitions of SC*

In the bulk of empirical research on young children's SC, emphasis has been focused on specific "skills" and/or traits relevant to peer interactions and relationships that reflect the background and interests of the scientist conducting the study (e.g., Ladd, 2005, chap. 5). That is, those scientists

interested primarily in adjustment outcomes tend to emphasize traits and interpersonal skills presumed to foster harmonious peer and adult-child interactions and/or establishing and maintaining healthy relationships, whereas those interested in school readiness tend to emphasize behavioral attributes and/or character traits related to attention and behavioral regulation that contribute to the child's being manageable and teachable in the classroom. These qualities may, but need not, overlap. Most of the published studies with these emphases identify (at least some) behavioral and/or trait measures that do correlate with ("predict" if the adjustment and/or readiness measures are assessed subsequently to the trait or behavioral assessment) adjustment or school readiness and the list(s) of significant correlates become de facto definition(s) of SC.

The primary pitfall of this strategy for studying SC is that lists of SC-relevant traits and behaviors for most studies only incompletely overlap and the definition of the construct expands incrementally with each newly published study. In addition, the list of skills and desirable traits approach lacks a developmental perspective in the sense that there is no necessary relation between the list of attributes defining SC at one developmental period and the list defining it in subsequent periods (e.g., Is the list the same across age periods? Does the list expand?). Waters and Sroufe (1983) also noted that "skills" at one developmental period may be "liabilities" in subsequent periods (e.g., crying as a signal for contact or intervention in infants and toddlers can be seen as a skill but preschool-age children are expected to use words to communicate these needs and crying is often seen as a sign of immaturity or incompetence). Another, and less appreciated, pitfall of this approach is that any behavioral quality or trait (e.g., "aggression") that is negatively associated with a given desired outcome (e.g., popularity) is presumed to reflect deficits or deviance with respect to the skill set defining SC and may become a target for intervention (for an argument concerning the flaws of this logic, see Hawley, 2007).

A second approach to defining SC emphasizes the antecedents and consequences of individual differences in children's skill sets in the domains of understanding and feeling about self and others (e.g., Ladd, 2005, chap. 13). In terms of consequences, SC is defined as being accepted and popular with their peers, as being able to initiate and maintain positive relationships (friendships) with peers and not being exploited in such relationships,

as being socially self-efficacious (i.e., believing that one can act effectively in social environments), as being aware of personal social goals and the social goals of peers, and the social acceptability of personal social goals. In terms of antecedents, socially competent children are thought to process social information in a manner appropriate to the context in which that information is detected and to select behavioral tactics from their repertoires of social behavior that are contingent on the information processed (e.g., Crick & Dodge, 1994), and they are believed to have histories of interactions in their families that support their understanding of the social goals of self and others (e.g., McDowell, Parke, & Wang, 2003). Rose-Krasnor (1997) noted that definitions of SC at this level tend to reflect the balance and also the tension between intrapersonal and interpersonal social goals (e.g., autonomy vs. integration; agency vs. communion) that emerge from social transactions but are not defined as social skills, *per se*. She referred to measures at this level as "indexes" of the SC construct.

Index-level definitions have been used in many studies of psychological adjustment and school readiness (for a review, see Ladd, 2005). However, these studies often fail to distinguish between effects of the absence of SC and the presence of deviant behaviors as the sources of children being rejected by peers, having no close friends, low social self-efficacy, or low self-esteem, although research reviewed by Rubin, Bukowski, and Parker (2006) suggests that both deficits and deviance may independently undermine SC. Additionally, most studies using index measures as criteria for SC are not able to distinguish "good enough" from "optimal" levels (e.g., how many reciprocated friends is "enough?" Are social information processes more efficient or accurate for highly self-efficacious vs. average children?). Furthermore, index measures occasionally identify paradoxes between status and behavior (e.g., aggression and popularity may be positively correlated; for a review, see Cillessen & Mayeux, 2007). From a sociocultural perspective, should relations between specific index measures and SC be equivalent across sociocultural contexts (e.g., Chen & French, 2008) and how might cross-group similarities and/or differences be described and interpreted? Finally, the role(s) of status variables (e.g., age, sex) and sociocultural contexts have not been definitively explored for measures at the index level.

A third, less widely appreciated, approach to defining SC emphasizes children's social effectiveness, that is, the accomplishment of social goals

within a particular social context (for preschool children, their group mates and supervisory adults). Waters and Sroufe (1983) suggested that this sort of conceptual definition for the construct is useful for developmental studies because it need not be adjusted as a function of the developmental level of children, even though the tactics (i.e., skills) used to accomplish social goals changes through developmental time. They suggested that SC should be construed as an "organizational construct" for early childhood insofar as it implies an integration of developmental domains (e.g., cognitive, social, emotional) in an evolutionarily salient context (for a similar treatment of attachment as an organizational construct during infancy, see Sroufe & Waters, 1977). Conceptual and empirical variations on this definition have been offered by Attili (1990), Bost, Vaughn, Washington, Cielinski, and Bradbard (1998), Rubin and Rose-Krasnor (1992), and Wright (1982), among others, although most investigators add constraints on the means through which goal accomplishment is achieved to the definition of SC. In the present study, we adopt the definition offered by Bost et al. (1998), who defined SC as the flexible application of behavioral, cognitive, and affective resources in the service of achieving social goals without unduly constraining opportunities for peers to attain their own goals and without entering onto a developmental trajectory limiting opportunities for the child to attain future goals (that may not yet be known to the child). We believe that the conceptual and empirical advantages of defining SC at this level outweigh advantages of definitions at the skills or index levels but we also recognize that there are serious questions concerning assessment methods and measurement that must be addressed at this level (see Waters & Sroufe, 1983).

#### *SC: Measurement Issues*

The review so far suggests that there is diversity in the definition(s) of the SC construct across studies and that this diversity is compounded when it comes to measurement and instrumentation. By far, the most frequent form of assessment consists of the opinions of knowledgeable adults (i.e., parents and/or supervisory adults) concerning the behaviors, traits, or qualities of the children being studied, usually gathered as rated items on questionnaires. Questionnaire data tend to be acquired rapidly and relatively cheaply, so ratings appeal to investigators with large samples and/or limited budgets. Nevertheless, quick and cheap carries a cost in terms of reliability and validity. Rating scale studies with

multiple informants (e.g., parents and teachers) have found that the factor (dimensional) structure of the instrument and item means can differ for different informants and also differ from standardization samples (e.g., Azria, 1999; Gray, Clancy, & King, 1981) and that cross-rater agreement is modest to moderate (Gray et al., 1981; Renk & Phares, 2004). Renk and Phares (2004) reported an average weighted effect size of  $r = .38$  (< 15% common variance) between parents' and teachers' ratings for SC across 16 studies and the weighted effect size for mothers and fathers was somewhat lower ( $r = .36$ ). Although each informant's rating may have significant and substantial correlations with other ratings made by the same informant (e.g., mother ratings of SC vs. mother ratings of self-regulation or problem behaviors, etc.), associations with ratings made by other informants tend to be modest to moderate in size. Renk and Phares (2004) recommend that multiple informants referencing multiple contexts be used in SC studies relying on ratings.

Peer judgments and ratings and self-ratings for a range of attributes relevant to SC have also been reported in many studies. Peer judgments include sociometric assessments, nominations for roles in a "class play," and standard questionnaire items for children old enough to read and understand written or oral instructions and individual interviews for preschool-age children. These assessments tend to be more time consuming and expensive to administer than adult ratings and have the same reliability and validity issues. Renk and Phares (2004) reported that self-ratings tend to show low levels of convergence with parent, teacher, and peer ratings (weighted  $r$ s between .21 and .30 for over 100 effect size estimates). However, teacher and peer ratings tended to be greater than for any other informant pair reported by Renk and Phares (weighted  $r$ s of .48 for 70 effect sizes). This may be due to the fact that peer ratings aggregate judgments by several (or many) different peers, which should improve the reliability of the overall peer rating. These results underscore the importance of having multiple informants, although it may be that self-reports are less informative than the reports of others.

Direct observations are considerably more expensive than ratings or individual interviews and are less frequently reported in studies of school-age children and adolescents (but see Pellegrini, 2004; Pellegrini & Bartini, 2000; Pellegrini & Long, 2002, for a program of counterexamples). Such observational studies are, however, fairly common for samples of preschool children (e.g., Bost et al., 1998;

Charlesworth & LaFreniere, 1983; McGrew, 1972; Pellegrini et al., 2007; Roper & Hinde, 1978; Sluckin & Smith, 1977; Strayer, 1980; Vaughn & Waters, 1981), perhaps because group sizes tend to be smaller than for older children and adolescents and because many of these studies were informed by the theory and methods of human ethology (e.g., Blurton-Jones, 1972; Eibl-Eibesfeldt, 1989). Direct observations of moment-to-moment transactions provide more accurate estimates of the frequencies, durations, and relative proportions of behavioral categories as a function of the social and physical contexts than do ratings, and when observations are summarized using behavior-relevant items, as when children are described using standard Q-sorts (e.g., Block & Block, 1980; Waters, Garber, Gornal, & Vaughn, 1983; Waters, Noyes, Vaughn, & Ricks, 1985), a wide variety of behavioral and personality attributes can be scored. The major pitfall of observational studies is that sample sizes are often very small, sometimes limited to a single group (e.g., Vaughn & Waters, 1981), which opens questions concerning the generality of findings.

Each of the assessment approaches described has advantages and limitations. Ideally, investigators would follow the advice of Renk and Phares (2004) and assess SC using multiple, independent informants with multiple methods. In the study we report here, five samples were studied using a common protocol of observations and interviews with multiple, independent informants and, for some samples, teacher ratings were also available.

#### *Using Broadband Measures to Assess SC*

In addition to considerations of the implications of using different assessment methods, it is also necessary to consider the level of abstraction at which the SC construct has been defined. Waters and Sroufe (1983) argued that when the construct is defined as "social effectiveness," it is necessary to use measures that are very "broadband" to capture the meaning of the construct. Broadband measurement implies that individual assessments provide many "bits" of information and should be connected to different aspects of effectiveness across a range of social contexts. Furthermore, they suggest that the use of broadband measures should facilitate the use of common assessment protocols across different studies, to test the generality of findings. Broadband measures should also afford opportunities to test age-related (or developmental) hypotheses about normative changes with respect to SC in children, as well as affording possibilities of cross-

group (e.g., sex, culture) comparisons. For example, we might anticipate that older children would be better able than younger children to achieve their social goals in the peer group (i.e., normative "growth" for SC), and this change may be reflected by a tighter degree of integration among the broadband measures used to index the construct or by changes in the external correlates of SC. There may also be differences in the means girls and boys use to attain social goals (again, supporting the use of multiple broadband measures). Several studies have documented sex differences in peer rejection, with boys being more likely to be rejected than girls (e.g., Ladd & Price, 1987; O'Neil, Welsh, Parke, Wang, & Strand, 1997; Volling, MacKinnon-Lewis, Rabiner, & Baradaran, 1993). Thus, we might anticipate finding sex differences in broadband measures of peer acceptance (e.g., sociometric scores). Finally, age and/or sex differences might interact with sociocultural factors to shift the magnitude or the organization of SC indicators across groups from different nations.

The study reported here includes data collected over a 15-year period in five samples that differ in terms of social class, ethnic composition, age, and country of origin. The investigators collaborated to mount a large-scale, ethologically informed study using a core of common measures to assess SC. Assessments included direct observations of interactions (e.g., rates of social interactions) that also included a judgment of the hedonic tone of the interactions (based on expressed affect by one or both interaction partners). Additional observations were made concerning the distribution of visual attention to peers and receiving visual regard from peers. Different observers provided Q-sort descriptions of the children, based on 20+ hr of observation in each classroom. A third team conducted sociometric interviews of individual children. Taken together, these assessments provide broadband assessments of social engagement and motivation (i.e., initiated interaction and visual attention received), social centrality (i.e., sociometric peer acceptance), and profiles of social behaviors and personality attributes (i.e., Q-sorts summarized in terms of their similarity to a hypothetical "very socially competent" preschooler). These were the same measures used by Bost et al. (1998) and by Vaughn (2001), and we included the Bost et al. Head Start sample in confirmatory factor analyses (CFA) to examine the degree of model similarity across the five samples. For two samples, we also have extensive teacher ratings relevant to child temperamental and personality attributes,

social-cognitive attributes, problem behaviors, and social integration. These data were used as correlates of the several SC indicators.

In the Bost et al. (1998) and Vaughn (2001) reports, CFA results suggested that SC could be modeled as having a hierarchical structure with the second-order construct (SC) influencing the child's scores for three "families" of measurement (initiated interaction and visual attention received, Q-sort descriptions summarized in relation to SC, and sociometric acceptance) that were, in turn, causally related to scores on seven measured variables. On the assumption that achieving social goals in any preschool group would entail social engagement and motivation and the flexible application of available resources, and that peer acceptance would facilitate goal attainment, the authors suggested that the general hierarchical structure of SC would be reproduced in different samples of preschool-age children, even if the demographic or cultural characteristics of the samples differed from the sample of children attending Head Start that participated in their study. In addition to the original sample, this report also includes a community sample involving 11 different centers and 27 unique classrooms, a second large sample recruited from two university-managed and accredited programs with 29 different classrooms, a sample from Lisbon, Portugal that includes seven classrooms from three different centers, and a Dutch sample including four kindergarten classrooms in three different schools. The two non-U.S. samples should be viewed as opportunity samples insofar as principal investigators for those studies found the model described by Bost et al. appealing and had contacted the U.S. investigators to learn about the measures and analytic approach. These two studies were initiated to determine the utility of the general approach and the hierarchical model for studies of SC in their own countries and not because differences between the Dutch and Portuguese samples were hypothesized. The total sample was over 1,500 different children. Teacher rating data are available for the community sample and for the sample from the university-managed centers (over 800 children).

Methodological and conceptual goals of this study are organized around understanding what it means to be a "socially competent" preschooler and how what SC means can be measured. Achieving our methodological goals entails: (a) testing the generality of the hierarchical structural model from the Bost et al. (1998) study across the four new samples and comparing the fit of a model with a

hierarchical structure to alternative first-order (non-hierarchical) orthogonal dimension solutions using CFA (e.g., Brown, 2006); (b) testing nested sets of models with increasingly strong equivalence constraints (i.e., equivalence of path weights across samples at the level of measurement factors, then tests for equivalence of the path weights in the structural model across samples), again with follow-up tests for age and sex subgroups; and (c) for the samples with teachers' ratings, exploratory factor analyses of items examine the underlying dimensional structure present in the rated items.

Further examination of the measurement and structural models tested helps us to describe both age-related changes and sex differences with respect to the SC construct, as we have defined it. For example, we suggested that a developmental feature of SC might be an increasing integration of facets across age levels. Evidence supporting this interpretation might be seen if there were structural differences in patterns of latent variable loadings for younger and older preschool children within samples and might also be observed in the patterns of correlations among indicators at the two ages. Both possibilities are tested in these data. Sex and country-context differences (or similarities) are evaluated in the same manner to determine whether one or another aspect of the SC construct is more central to the core meaning for girls and boys (or for children from different sociocultural groups). Examining relations between our set of SC indicator variables and dimensions from the teachers' ratings further advances our conceptual goal of understanding the implications of SC in terms of behavior and personality attributes (as perceived by supervising adults), again examined in terms of age changes and sex differences. In constructing and testing these models, we include all children having data for at least one of the seven measured SC indicators. Full information maximum likelihood (FIML) modeling was used in tests of all structural equation modeling and CFA analyses to take advantage of as many cases as possible, including those with some missing data.

## Method

### *Participants*

Demographic particulars of each sample are provided in Table 1. In all samples ( $N = 1,540$ ), written consent of a parent or legal guardian was obtained for every participating child. Sample 1 consists of 471 children from the original Head Start sample

Table 1  
Sample Demographic Characteristics

Sample	Head Start	Community	University	Netherlands	Portuguese
Years of data collection	1992–1995	1996–2000	2000–2005	2000–2001	2004–2007
Program type	Public intervention	Mixed for and not-for-profit	University managed	Public kindergarten	Private preschool
No. centers	6	11	2	3	2
No. classrooms	30	29	25	4	5
Participation rate	> 90%	> 80%	> 90%	> 90%	> 95%
Total girls	232	227	160	55	65
Total boys	239	249	198	56	59
Total younger	243	126	130	0	24
Mean age younger	42 <sup>a</sup>	42.4	42.1	n/a	40.6
Total older	228	350	228	111	100
Mean age older	54 <sup>a</sup>	58.1	54.3	65.6	61.5
Ethnic mix	> 95% AA	30% AA 70% EA	32% AA 68% EA	50% European 50% African descent	European
SES mix	Low income	75% middle, 25% working class	Middle class	Middle class	Middle class

Note. AA = African American; EA = European American; SES = socioeconomic status.

<sup>a</sup>In the Head Start sample, actual birthdays were withheld for confidentiality reasons. However, children were generally assigned to classrooms on the basis of when their birthday fell in the academic year. Children < 48 months at the beginning of the academic year were assigned to 3-year-old classes, those ≥ 48 months were assigned to 4-year-old classes.

reported by Bost et al. (1998). Of these 471 children, 115 were seen in 2 consecutive years but they are included in the analyses reported here only for the 1st year of their participation in the project. Sample 2 consists of 476 children from a community sample recruited from two communities from Alabama. One center was located in a large urban area and the other in a smaller community in the east central region of the state. Two centers were managed by a major university and were accredited by the National Association for the Education of Young Children (NAEYC), one was a not-for-profit center serving lower income working families, and the remaining eight centers were for-profit centers serving primarily middle- and working-class families. Sample 3 consisted of 358 children recruited from two NAEYC-accredited centers managed by a major Southeastern university. Families served by these programs were predominantly middle class by the standards of their local communities. A total of 157 children from this sample were observed in consecutive years but for this report, only the assessments from the 2nd year of participation are used in analyses. Sample 4 consists of 111 children recruited from four kindergarten classrooms in the Netherlands educational system (two classrooms [ $N = 55$ ] located in den Haag and two classrooms [ $N = 56$ ] located on the island of St. Maarten, Netherlands Antilles). Children from the den Haag kindergartens were European and those from St. Maarten were of African descent. In the Dutch

educational system, children are entitled to enter kindergarten when they reach the age of 48 months, and over 90% of eligible children are enrolled when they reach their fourth birthday during the academic year. Children remain in the same kindergarten group for 2 years, so the oldest children reach 6+ years of age before moving to the next level (in this sample, the oldest child was 80 months of age when observed). Families served by these schools were considered middle class by the standards of the Netherlands and Netherlands Antilles. Sample 5 consisted of 124 children recruited from two different centers serving middle-class families in the region near Lisbon, Portugal. Preschool classrooms in Portugal are usually formed when a child is 36+ months of age and the group remains together until the children leave for primary school. As is common in the Lisbon area, the preschool classrooms were associated with a specific primary school and the children often remained together with familiar peers until the fourth-grade level. Although many children in this sample were assessed in consecutive years, the data reported here include only one assessment period for each child.

#### SC Assessments

The same set of seven broadband indicators was used to assess SC in each sample. These included two Q-sort descriptions (California Child Q-sort [CCQ], Block & Block, 1980; Preschool Q-set [PQ],

Bronson's adaptation of a Q-sort originally used by Baumrind, 1967), direct observations of initiated interaction and visual attention to peers (Bost et al., 1998), and two sociometric interviews (three-like, three-dislike nominations; paired comparison sociometric). Teachers' ratings relevant to temperament and personality attributes, social-cognitive skills, and qualities of peer-directed behaviors were available for two of these samples. With only a few exceptions, descriptions of the measures that follow apply to all samples, with qualifications specific to one or another sample elaborated. For all samples, excepting the two classrooms from the Netherlands sample assessed in St. Maarten, different teams of research staff collected data for the three measurement families (i.e., Q-sort attribute profiles, social engagement and motivation, peer acceptance), to protect the independence of the SC data. In the St. Maarten subsample, two observers collected data for all SC indicators. Numbers of participants for each of the seven SC indicators are presented in Table 2. In all samples, raw scores for the seven SC indicators were standardized within classrooms prior to analysis.

*Q-sort profiles.* Q-sort observers worked in teams of two for each classroom, with each observer spending a minimum of 20 hr observing the children in a given classroom. They took notes on the behaviors and personality attributes of individual children over this period, taking care to observe each child on several different days and across a variety of activity settings (e.g., meal times, small groups, free-play indoors, outdoor play, transition activities such as standing in lines or getting ready for nap time, and teacher-supervised picking up of toys). In three samples (community, university managed, Portugal) each observer described the children with both the CCQ and PQ item sets after completing his or her observations, according

to predetermined distributions of items to nine categories. In the Head Start and Dutch samples, observers split the sorting task such that one observer described half the children with the CCQ and the other half using the PQ item sets. The other observer described the first half of the class using the PQ and the other half using the CCQ. If a child was absent from the classroom for over half a given observer's observation hours, she or he did not provide a Q-sort description of that child. A total of 1,256 had Q-sort descriptions for both Q-sets.

Prior to data collection, observers were trained in the meanings of the items and were instructed about items they were not likely to be able to observe (such items were to be placed in the center categories [4, 5, 6] of the Q-sort). Both Q-sets were sorted according to rectangular distributions with equal numbers of items in each category (i.e., 9 piles of 11, with the odd item sorted into pile 5 for the CCQ and 9 piles of 8 for the PQ). The Q-sort descriptions of each child were used to derive SC "scores" for each child using the criteria published by Waters et al. (1985). Thus, the Q-sort description for a child provided by an observer was correlated with the profile of a hypothetical child at the extreme for SC that had been generated by aggregating the descriptions provided by developmental scientists with expertise in children's social development. The correlation between a Q-sort for a given child and the "criterion" sort for the construct becomes her or his "score" for that construct. This technique is commonly used to summarize Q-data and yields valid and reliable scores over a range of personality and behavior-relevant constructs for preschool-age children (e.g., Bost et al., 1998; Block, 1961/1978; Block & Block, 1980; Vaughn, Vollenweider, Bost, Azria-Evans, & Snider, 2003; Waters et al., 1985). Following the suggestions made by Waters et al. (1983) about influences of

Table 2  
Numbers of Participants for Each Social Competence Indicator

Data set	California child Q N	Preschool Q N	Visual attention received N	Initiated positive interaction N	Initiated neutral interaction N	Nominations sociometric N	Paired comparisons sociometric N
Full sample	1,318	1,321	1,480	1,462	1,462	1,390	1,389
Head Start	387	390	470	471	471	466	466
Community	371	376	451	428	428	381	368
University managed	337	332	332	334	334	326	330
Netherlands	105	105	106	106	106	108	108
Portugal	118	118	121	121	121	109	117

social desirability bias in Q-sort data, the scores were adjusted for social desirability response sets on the part of observers by controlling for the effects social desirability (i.e., partial correlations) while calculating the correlations between individual children and the criterion sorts.

For the 1,256 participants with both CCQ and PQ descriptions, CCQ SC scores averaged .07 (range =  $-.44$  to  $.58$ ), and for the PQ SC scores, the mean was .04 (range =  $-.61$  to  $.57$ ) across the five samples. The average cross-rater agreement for the CCQ SC criterion score was .59, and the average cross-rater agreement for the PQ was .62. Cross-rater, cross-Q-set correlations averaged .62 across the full sample (range =  $.57$  to  $.76$ ). Final scores were averaged across raters for each of the two Q-sort criterion scores for SC. As noted, final scores were standardized within classroom group prior to further analysis (i.e., inferential analyses all use z-score variables).

*Initiated interactions and visual attention.* Interaction and visual attention data were collected by teams of observers (between two and six for any given classroom) who worked independently from the Q-sort observation teams. Using the class roster, an observer watched a given child for a 15-s interval and recorded identifiers for all children with whom the target engaged in interaction. Codes for the initiator and affective valence (positive, neutral, negative) of the interaction episode were recorded. Interactions were coded as positive if one or both children showed positive affect in the context of the social exchange (i.e., smile, laugh, gesture, or vocalization indicating a positive feeling), unless such affect expression was accompanied by expressions of negative affect (e.g., crying, distress, pain, strong irritability) by the interactive partner. Interactions were coded as negative if one or both children expressed negative affects (e.g., anger, distress, fear, sadness) in a facial, vocal, or gestural mode, unless these expressions were made in the context of fantasy play (e.g., a Power Ranger attacks an enemy robot). All exchanges not identified as positive or negative were coded as neutral (e.g., exchanges of greetings or conversations during a meal or in the context of a school-related task that did not include the expression of affect, nonverbal exchanges that included physical contact and a reaction to contact). It is helpful to understand that these codes contain two levels of information. At the level of the individual, initiating and receiving distinguish the interactive partners. However, the coding system did not uniquely specify which child was responsi-

ble for affect valence. Thus, the expressed affect defining the interaction as positive or negative could have been exhibited by either interactive partner, and this aspect of the code is tagged for both initiator and recipient of initiation.

All children present in the group during a given observation round were watched for one 15-s interval before any child was watched twice. Observers did not work in pairs and rarely observed a given child simultaneously. Each observer started observing at a different position in the class roster for subsequent observation rounds. Scores were the total frequencies of positive, neutral, and negative interactions initiated by the target child. To adjust for absences from the classroom during observations and for differences in the number of observational rounds across classrooms (range = 100–228 rounds of observation in a given classroom), the total scores were converted to rate scores (i.e., dividing the total score by the number of observation rounds for which the target child was actually present in the classroom) and standardized within classroom. Children absent from the classroom for 50% or more of the observational rounds in any classroom were not given rate scores (i.e., treated as missing for these observations).

Observers received training in the observation system prior to beginning direct observations in the classroom. For most samples, rater agreement was estimated as the alpha coefficient for individual rate scores across raters. That is, the vector of rate scores from the observations of a given observer was treated as a single “item” and the standard internal consistency estimate (Cronbach’s alpha or Spearman–Brown prophecy correlation for classes in which only two observers provided data) was calculated. Reliability estimates ranged from .43 to .90 across all interaction categories (*Mdn* = .73 and only a single classroom less than .50 for any interaction category) across classrooms. For 30 classrooms, raters conducted separate joint observations and kappa coefficients were calculated. These ranged from .78 to 1.00 (*Mdn* = .87) across the three categories of interaction. These data suggest that observers were in agreement about the criteria for coding affective valence and were also in agreement about the overall rates of interaction for the observed children. For the purposes of this report, only the standardized rate scores for positive and neutral interactions initiated were retained for analysis (see Vaughn, 2001).

Interaction observers were also responsible for collecting visual attention data. Observers were instructed to intersperse rounds of interaction and



visual regard observations (e.g., five interaction, five visual attention). An observer watched a given target child for a period of 6 s and recorded the identity codes for all children receiving a unit of visual regard from the observation target (no child was credited with receiving more than one unit of visual regard per 6 s interval, although several different children could each receive one unit from a given target in a single interval). No child present in the classroom was observed twice before all other peers were observed once. Two to six observers collected approximately 200 observation rounds in each classroom (range = 139–225 observation rounds across classrooms). Total scores were the sum of visual regard units received by a given child from all peers. To adjust for absences from the classroom during observations and for differences in the number of observational rounds across classroom, the total scores were converted to rates (i.e., dividing the total score by the number of observation rounds for which the target child was actually present in the classroom) and standardized within classroom. As with the interaction data, children who were not present in the classroom for 50% or more of the observation rounds were considered as having missing data for the visual regard observations.

Also as with the interaction data, interrater reliability was estimated from the vectors of scores for visual attention received from peers derived from the observations of each individual observer in each classroom. Alpha coefficients ranged from .53 to .90 (*Mdn* = .85) across all classrooms and kappa coefficients (based on joint observations in 30 classrooms) ranged from .74 to .91 across all rater pairs with joint observation data (*Mdn* = .81). These results suggest that visual regard was reliably assessed. As noted by Bakeman and Gottman (1997), this approach to assessing reliability is similar to comparing alternate forms of a test, for which the full score (rather than individual item scores) is of interest. That is, when multiple forms of a test are used to assess some ability, the investigator is not interested in whether an individual gets a particular item correct on both forms but rather in whether the total number of items correct is similar for each form. As the visual regard score of interest is the total number of units received by a child, examination of total scores for each observer is justified.

*Sociometric acceptance.* Positive and negative sociometric scores were derived from a nominations sociometric task (McCandless & Marshall, 1957) administered individually by a trained research staff member. Children were presented

with an array of photographs of their classmates and asked to identify a child they “especially liked.” After making three such choices, the children were asked to identify a classmate they “did not especially like.” They made three negative choices and then returned to the array to identify more “liked” children. The assistant requested choices until all class members’ photographs were turned over. Positive and negative choice scores were derived on the basis of the first three “liked” choices and the three “not liked” choices. Average values were calculated by dividing the total number of positive and negative choices received by the number of children making choices.

Sociometric acceptance was also scored from a paired comparisons task. For this task, cards for all pairs of children in the class were prepared, with each child’s photograph appearing on the left- or right-hand side of the stimulus cards an equal number of times. The order of presentation was such that no child was seen twice before all other children were seen once. Cards were presented one at a time and the child was asked, “Which of these two children do you especially like?” The number of pairs presented in this manner was substantial,  $n(n - 1)/2$ , for 190 pairs in a class of 20 children, and some children grew tired of the task. If a child’s attention wandered, the assistant stopped the task and continued it later. All children included in this report completed this task in one or two 15-min sessions. Positive acceptance scores were the total number of times a child was chosen by peers. These were averaged by dividing the total by the number of children making choices and then standardized within classroom.

#### *Teacher Ratings*

In both the community and university-managed samples described previously, teachers rated children’s social behavior, their social engagement tactics, and temperament/personality using items from widely used instruments: (Child Characteristics Questionnaire (CCQ), an age-appropriate extension of the Infant Characteristics Questionnaire (Bates, Freeland, & Lounsbury, 1979; 32 items); Social Competence and Behavior Evaluation Scale-Short Form (LaFreniere & Dumas, 1996; 30 items); Interpersonal Competence Scale (ICS; Cairns, Leung, Gest, & Cairns, 1995; 18 items); Teacher Rating of Social Skills (TRSS; Dodge & Somberg, 1987; 17 items); Social Behavior Scale (SBS; Cairns et al., 1995; seven-item aggressive engagement list). Typically, both the lead and associate teachers in a

classroom completed ratings and these were averaged for each item. Previous analyses of these item sets in large samples of preschool children (over 400 children with each item set; Akers, 2006; Snider, 1999) have suggested that the dimensional structures of these instruments for our samples do not correspond well with published accounts. Furthermore, Akers (2006) found that the underlying dimensional structure for these item-sets tended to differ for younger and older preschoolers. Finally, the content domains of items across scales show considerable redundancy, even though authors' dimension titles do not.

For these reasons, we redimensioned the set of items from the several instruments. After standardizing items within each instrument to reduce the effects of different scale ranges across instruments, we averaged the values of redundant items from different instruments (e.g., an item referring to getting into fights with peers appears in the TRSS, ICS, and SBS item sets), which reduced the total item set by 19 items. The resulting 86 items were dimensionalized using principal axis factoring with an oblique rotation, resulting in a 15-factor solution. One of these had a significant loading (i.e.,  $> .4$ ) for only one item ("How does this child react when you help dressing him or her?") and one factor was composed of two items, neither of which had a loading greater than .40. We dropped these two factors from subsequent analysis (see the Appendix). Scores were derived for each of the remaining factors by unit weighting all items with factor pattern loadings  $> .39$  and the resulting scores were submitted to a second principal axis factor analysis, again with an oblique rotation. Examination of the factor pattern matrix (see Hair, Anderson, Tatham, & Black, 1995; Tabachnick & Fidell, 2001, for arguments favoring the use of factor pattern matrices rather than factor loading matrices when oblique rotations are used in principal axis analysis) yielded a two-factor solution (see Table 3) with a cross-factor correlation of  $-.34$ . We labeled these factors "Peer Acceptance-Positive Mood" (nine dimensions had their highest loading on this factor) and "Negative Engagement" (four dimensions had their highest loading on this factor).

## Results

Results are presented in four sections. As preliminary analyses, comparisons of means across sample, sex, and age are evaluated. In these analyses, the contrasts of interest are for sex and sample or

Table 3  
Factor Pattern Matrix for Principal Axis Factor Analysis of First-Order Teacher Factors With Oblique Rotation

	Peer Acceptance- Positive Mood	Negative Engagement
Defiant	-.080	.713
Academic skills	.354	-.090
Desire for adult contact	.231	-.123
Adaptable	.740	-.005
Peer acceptance	.715	.003
Socially engaged	.643	.378
Aggressive-coercive style	.062	.924
Regularity	.335	-.144
Positive mood	.762	.002
Resistance to control	-.119	.674
Social awareness	.472	-.183
Social cognition	.558	-.333
Negative emotionality	-.274	.709

Age  $\times$  Sex interactions. No main effects for sample or age should be obtained because all SC indicators had been standardized within classroom and, because most sites grouped children of similar age grades (e.g., 3, 4, and 5 years) together, no main effects of age are anticipated. However, main effect sex differences are possible and interactions with the other independent variables could be found. Second, we present the primary analyses of the hierarchical model. In these analyses, the first model is unconstrained to determine whether a hierarchical structure for SC can be fit in every sample. The fit of this unconstrained model is compared to two alternative models. The first alternative is a single-factor model for which all variables have their primary loading on the first extracted component (i.e., treating SC as a single unitary trait). The second alternative assumes two orthogonal factors and is based on an actual principal components analysis of these data that used all cases with complete data. The comparison models use all cases and FIML methods for estimating the fit of the model. In the next section, the model is fit with increasingly strong constraints on equivalence of path weights in the model. In this section, we also report models that include sex and age (i.e., younger than 48 months of age at the beginning of the academic year vs. 48 months or older at the beginning of the academic year) as grouping variables. When age is included in these analyses, the Dutch sample is excluded because there were no children younger than 48 months of age in the Dutch kindergarten classrooms. In the last section, the variables from the model are examined in relation to

the scales derived from teachers' ratings to determine the overlap between teachers' evaluations of child behavior and personality relevant to SC and the constructs used to summarize our direct observations and sociometric interviews.

Sample Mean Differences

Values for the standardized SC indicators are broken down by sample and sex in Table 4. Because there are considerable missing data across samples, seven univariate ANOVAs were calculated, rather than MANOVAs. The analyses were calculated as two sets. In the first set of ANOVAs, sample and sex were independent variables, and in the second set sample, sex and age (i.e., younger vs. older children for the four samples that included younger children) were independent. For the first set of ANOVAs, five of seven main effects for sex were significant. Girls had higher scores for both of the sociometric measures and for the SC criterion score from the CCQ,  $F_s = 4.66, 14.02,$  and  $7.98, p_s < .05, .001,$  and  $.01,$  for the nominations and paired comparisons sociometric tasks, and CCQ criterion score, respectively. Boys had higher scores for receiving visual attention and for initiating interactions with a neutral valence,  $F_s = 5.33$  and  $6.23, p_s < .05.$

These results are qualified by significant interactions with sample in three of the seven analyses. Sex  $\times$  Sample interactions were significant for both sociometric scores,  $F_s = 3.17$  and  $5.27, p_s < .05$  and  $.001,$  for nominations and paired comparisons scores, respectively. Although girls had higher scores than boys in four of the five samples, the differences between boys and girls were significant in post hoc tests only for the Head Start and community samples. A significant Sex  $\times$  Sample interaction was also observed for one of the social engagement-motivation group variables (visual attention received),  $F = 4.90, p < .001.$  Post hoc analyses indicated that a sex difference favoring boys was significant in the Portuguese and community samples. To test whether these results might be due to the differences in age distributions across samples, we recalculated the ANOVAs after excluding younger children from the analysis. Four of five main effects for sex remained significant (with the fifth dropping to  $p < .10$ ) and two of three Sex  $\times$  Sample interactions remained significant in these tests.

The next set of analyses tested for sample, age, and sex differences after excluding the Dutch sample. Results were similar, but not identical, to the pattern to the previous analyses. Three of seven

Table 4  
Mean and Standard Deviations for Standardized Social Competence Indicators by Sample and Sex

	Dutch	HS	Community	University	Portuguese	$F_{sex}$	$F_{sex \times Sample}$
Nominations	Boys vs. girls -.11 (.91 vs. .93)	-.19 vs. .13 (.85 vs. 1.08)	-.10 vs. .11 (.95 vs. 1.03)	-.16 vs. .17 (1.00 vs. .96)	.23 vs. -.22 (1.11 vs. .97)	4.66*	3.17*
acceptance							
Pairs	Boys vs. girls -.23 (.95 vs. .88)	-.29 vs. .21 (.95 vs. .95)	-.18 vs. .19 (.95 vs. .95)	-.03 vs. .00 (1.00 vs. .94)	.11 vs. -.10 (1.02 vs. .95)	14.02***	5.27***
acceptance							
CCQ social	Boys vs. girls -.17 vs. .16 (.95 vs. .88)	-.13 vs. .06 (.93 vs. 1.03)	.02 vs. -.01 (.93 vs. .89)	-.13 vs. .17 (.94 vs. .99)	-.03 vs. .03 (.96 vs. 1.01)	7.98**	1.71
competence							
PQ social	Boys vs. girls -.00 vs. .00 (1.04 vs. .85)	-.11 vs. .01 (.93 vs. 1.01)	.08 vs. -.09 (.93 vs. .88)	.09 vs. -.07 (.96 vs. .97)	.15 vs. -.13 (.92 vs. 1.03)	2.49	1.73
competence							
Visual	Boys vs. girls -.17 vs. .17 (.85 vs. .96)	-.00 vs. -.08 (.98 vs. .90)	.09 vs. -.10 (.97 vs. .97)	.07 vs. .01 (1.02 vs. 1.04)	.39 vs. -.36 (1.00 vs. .83)	5.33*	4.90**
attention							
Positive	Boys vs. girls -.12 vs. .12 (.92 vs. .91)	-.04 vs. -.06 (.91 vs. 1.00)	.00 vs. -.05 (1.03 vs. .83)	-.07 vs. -.07 (.90 vs. .99)	.18 vs. -.16 (.97 vs. .98)	0.25	1.40
interaction							
Neutral	Boys vs. girls .00 vs. -.00 (.99 vs. .87)	.03 vs. -.10 (1.03 vs. .92)	.03 vs. -.04 (.96 vs. .96)	.10 vs. -.04 (1.01 vs. 1.08)	.24 vs. -.22 (.91 vs. 1.01)	6.23*	1.01
interaction							

Note. Standard deviations are in parentheses. CCQ = California Child Q-sort; PQ = Preschool Q-set.  
\* $p < .05.$  \*\* $p < .01.$  \*\*\* $p < .001.$

main effects for sex were significant (i.e., paired comparison sociometric, visual attention received, and initiated neutral interactions). Two of seven Sex  $\times$  Sample interactions were significant (i.e., paired comparison sociometric acceptance and visual attention received from peers). Neither the main effect of age nor any interaction involving age reached significance. Taken together, these analyses do not suggest that age differences across samples account for the sex main effects or the Sex  $\times$  Sample interactions (readers interested in the statistical details of these analyses may receive them from the first author).

#### *A Hierarchical Model of SC: Factorial Invariance*

*Sample invariance.* To determine whether a hierarchical model of SC is applicable across five samples, a CFA was carried out using the FIML estimation procedure of the Amos 6 program (Arbuckle, 2005). This model showed a good fit to the data, suggesting that a hierarchical structure of SC represented the data well. Model chi-square and other fit indices are shown in Table 5 and the measurement and structural path weights are presented in Figure 1. For ease of comprehension, we do not include disturbance or error terms in Figure 1. We then tested two alternative models (also presented in Table 5). For these tests, we combined the data from all samples and tested the hierarchical model against: (a) a single factor model in which all seven SC indicators were indicators of a single latent variable (SC) and (b) a model with two orthogonal factors (observation/Q-sort and sociometric interview). These two models reflect the actual structure of the measured

variables (for those cases with complete data) in an exploratory factor analysis. The first alternative model did not fit the data well, indicating that SC is not best described as a single first-order dimension governing the behavior of the SC indicators. The second model produced a better fit than did the single-factor model; nevertheless, it described relations among the measured variables significantly less adequately than did the hierarchical model. Based on these analyses, a hierarchical model of SC was chosen to examine the factorial invariance across samples, age, and gender.

A series of hierarchically nested models was computed to test whether the second-order factor structure of SC is replicable across the five samples (see Table 6). Starting with an unconstrained model, each successive model added increasingly stringent constraints to the previous model. Three models were tested: (a) the unconstrained model (Model 1), (b) invariance of measurement factor loadings (Model 2), and (c) invariance of measurement *and* structural factor loadings (Model 3). Evidence for factorial invariance was examined by means of chi-square difference tests and the overall fit of the models. These models and their corresponding fit statistics are displayed in Table 6. In an absolute sense, all models that we tested produced an acceptable fit. The chi-square values divided by its degree of freedom were small, and fit indices were above the recommended cut-points (see Bentler & Bonett, 1980; Browne & Cudeck, 1993; Marsh, Balla, & McDonald, 1988). The difference in chi squares for the measurement model (i.e., Model 2, constraints on variables to first-order factors) and Model 1 was not significant, suggesting that first-order factor loadings are not significantly

Table 5  
*Comparing a Hierarchical Model of Social Competence With Two Alternative Models*

Fit indices	Unconstrained multigroup hierarchical model	Single-sample hierarchical model	Single-sample one-factor model	Single-sample two orthogonal factors
$\chi^2$ (df)	120.18*** (55)	40.60*** (11)	663.13*** (14)	253.64*** (13)
$\chi^2/df$	2.19	3.39	47.37	19.51
NFI	.96	.99	.77	.91
RFI	.91	.97	.55	.81
IFI	.98	.99	.78	.92
NNFI	.95	.97	.55	.82
CFI	.98	.99	.78	.92
RMSEA	.03	.04	.17	.11

*Note.* NFI = normed fit index; RFI = relative fit index; IFI = incremental fit index; NNFI = non-normed fit index; CFI = comparative fit index; RMSEA = root mean square error of approximation.

\*\*\* $p < .001$ .

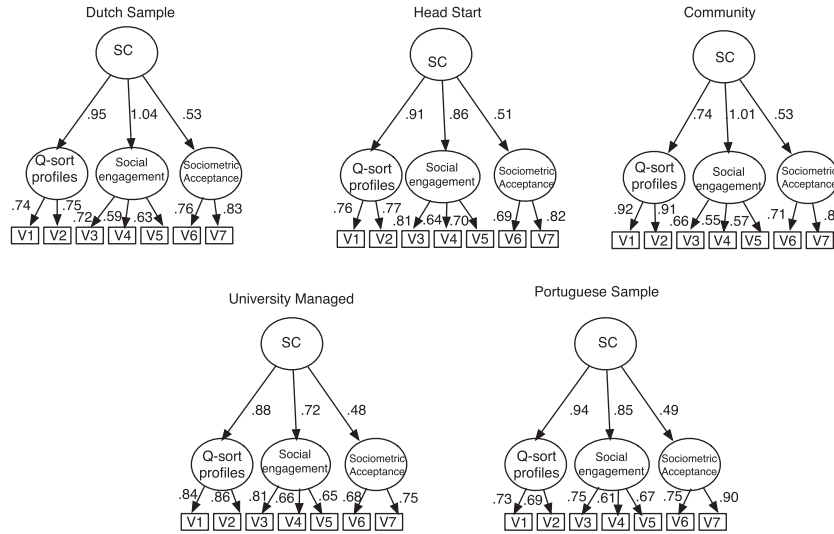


Figure 1. Measurement and structural path weights for Model 2 in five samples.

Table 6  
Factorial Invariance Across Sample (A), Sex (B), and Age (C)

	$\chi^2$	df	$\chi^2/df$	NFI	NNFI	CFI	RMSEA	$\Delta\chi^2$
Model 1: Unconstrained model								
A	120.18	55	2.19	.96	.95	.98	.03	
B	181.42	110	1.65	.95	.94	.98	.02	
C	149.58	88	1.70	.95	.95	.98	.02	
Model 2: Measurement loadings								
A	138.53	71	1.95	.96	.96	.98	.03	18.35
B	219.69	146	1.151	.94	.96	.98	.02	38.27
C	243.39	116	2.18	.92	.92	.96	.03	93.80***
Model 3: Structural loadings								
Model 3a								
A	153.14	75	2.04	.95	.95	.98	.03	14.62**
B	234.30	155	1.51	.93	.95	.98	.02	14.61
Model 3b								
A	162.65	79	2.06	.95	.95	.97	.03	24.21**
B	251.38	164	1.53	.93	.95	.97	.02	31.70*
Model 3c								
A	176.13	83	2.12	.95	.95	.97	.03	37.61***
B	268.11	173	1.55	.92	.95	.97	.02	48.42**

Note. Model 3a = constraining only the first second-order factor loadings; Model 3b = constraining the first two second-order factor loadings; Model 3c = constraining all second-order factor loadings. Chi squares for Model 3 are compared with chi squares for Model 2. A = 5 groups; B = 10 groups (gender within each sample); C = 8 groups (age within each sample, excluding Dutch sample). NFI = normed fit index; NNFI = non-normed fit index; CFI = comparative fit index; RMSEA = root mean square error of approximation.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

different across these five samples. In Model 3, all three structural factor loadings were constrained to equality across the samples. The chi-square difference test between Model 3cA and Model 2A was

significant, providing support for weak invariance of structural factor loadings. To determine which structural factor loading was responsible for weak invariance, we ran further tests on the structural

factor loadings by looking at the structural paths one at a time (Models 3aA and 3bA). Results of these analyses indicated that no single path of combination of paths could be constrained to equivalence across the five samples. Thus, while the hierarchical structure for SC is apparent in each sample, the structural path weights differ from group to group.

*Invariance for sex and age.* The measurement and structural models were also tested for invariance across sex (10 groups) and age (8 groups, Dutch sample excluded). Results of these analyses also are presented in Table 6. The unconstrained model had good fit to the data in both the sex and age analyses (Models 1B and 1C). Again using the chi-square difference test as the criterion of difference or similarity, we found that the measurement model was invariant (i.e., did not differ significantly from the unconstrained model) for the Sex  $\times$  Sample breakdown (Model 2B). The next set of analyses (Model 3B) tested for invariance of the structural aspects of the model at the Sex  $\times$  Sample level. The fully constrained model (Model 3cB in Table 6) was not invariant by the chi-square test; however, tests on single paths indicated that constraining the first path (Q-sort observations) to equivalence did not produce a significant increase in the chi square from Model 2B to Model 3aB. That is, the path from the SC (second-order hierarchical latent variable) to the Q-sort profile factor is equivalent for boys and girls within each sample.

Analyses testing equivalence in the Age  $\times$  Sample breakdown suggest differences in the measurement model for younger and older children (i.e., the latent factor loadings are not equivalent for younger and older preschoolers, Model 2C). Because we did not find measurement equivalence across age level, we did not further test for invariance for age at the next (structural) level.

#### *Observation and Interview Assessments of SC and Teacher-Rated Scales*

Examination of Table 3 shows that three variables with their highest loadings on the Peer Acceptance-Positive Mood factor (i.e., regularity, desire for adult contact, academic skills) are less saturated with common variance (i.e., factor loadings  $< .40$ ) than the other six variables loading on this factor. We deleted these variables before calculating unit-weighted factor scores for the two dimensions. Initial analyses of the teacher rating factors tested for mean differences across age (younger vs. older) and sex. Main effects of both sex and age were significant. Older children and girls had higher scores for the Peer Acceptance-Positive Mood factor,  $F_s = 15.12$  and  $23.47$ ,  $p_s < .001$ , for age and sex, respectively. Younger children and boys had higher scores for the Negative Engagement factor,  $F_s = 6.31$  and  $30.31$   $p_s < .05$  and  $.001$ , respectively. The Sex  $\times$  Age interaction did not approach significance. Because sex and age differences were obtained in analyses of the teacher-rated scales, we

Table 7  
Correlations Between Social Competence Composites and Teacher-Rated Factors Broken Down by Age and Sex

	1	2	3	4	5
Older children <sup>a</sup>					
1. Q-sort profiles	—	.59***	.36***	.30***	.10
2. Social Engagement-Motivation	.61***	—	.33***	.29***	.20**
3. Peer Acceptance (sociometric)	.37***	.45***	—	.34***	-.14
4. Peer Acceptance-Positive Mood	.39***	.41***	.29***	—	-.44***
5. Negative Engagement	.22**	.15*	-.11	-.32***	—
Younger children <sup>b</sup>					
1. Q-sort profiles	—	.59***	.46***	.31**	.10
2. Social Engagement-Motivation	.29**	—	.51***	.26*	.08
3. Peer Acceptance (sociometric) <sup>c</sup>	.25*	.07	—	.31**	-.14
4. Peer Acceptance-Positive Mood	.25*	.20	.20	—	-.40***
5. Negative Engagement	.09	-.01	-.06	-.40***	—

<sup>a</sup>Scores for girls ( $n = 198$ ) are above the diagonal. Scores for boys ( $n = 255$ ) are below the diagonal. <sup>b</sup>Scores for girls ( $n = 108$ ) are above the diagonal. Scores for boys ( $n = 117$ ) are below the diagonal. <sup>c</sup>Correlations between Peer Acceptance and both Q-sort profiles and Social Engagement-Motivation composites for younger boys are significantly different from the analogous correlations for both younger girls and older boys.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

examined age and sex groups separately in our next analyses relating teacher ratings to our SC indicators.

Table 7 displays within- and between-set correlations for the SC composites and the factors derived from teachers' ratings. As expected (from the oblique rotation of dimensions in these data), scores for the two factors from the teachers' ratings were negatively correlated in each of the four Age  $\times$  Sex breakdowns. Furthermore, the SC composites were positively and significantly correlated in each breakdown except for younger boys. In this subsample, significant correlations were lower than anticipated, and the correlation between Peer Acceptance (from sociometric data) and the Social Engagement-Motivation composite ( $r = .07$ ) did not reach significance. Subsequent tests ( $r$  to  $Z$  transformations) on the differences between correlation magnitudes revealed that the relation between Peer Acceptance and Social Engagement-Motivation for younger boys was significantly lower than for younger girls and for both older boys and girls. In addition, the  $r = .29$  correlation between the Q-sort profiles and Social Engagement-Motivation composites for younger boys was also significantly lower than the analogous correlation in the other three subgroups. These patterns of association suggest the possibility that the failure to find age equivalences in the measurement and structural models described earlier may be due to reduced coherence of the SC model for younger preschool boys.

The Peer Acceptance-Positive Mood factor from the teachers' ratings had consistently positive and significant associations with the SC composites for the older children (six of six associations are significant). This same pattern was observed for the younger girls (three of three correlations were significant). Only the younger boys are outliers (one of three correlations reach significance); however, even for the younger boys, the nonsignificant relations are positively signed and they are not significantly lower than correlations for the other subsamples. Clearly, the content of this factor from the teachers' ratings overlaps with the notion that SC entails accomplishing one's social goals. The Negative Engagement factor from teachers' ratings did not have much overlap with the SC composites. For younger children, the absolute values of correlations between this factor and the SC composites ranged from .01 to .14 (zero of six significant). For the older children, three of six correlations reached significance but no correlation exceeded  $r = .22$  and each of the significant correlations was signed positively. For these preschool children, levels of nega-

tive engagement as rated by teachers did not appear to reflect children's measured SC. This seems counter to the generally accepted wisdom of the developmental sciences and we return to this finding in the Discussion.

## Discussion

We have suggested that preschool children's SC cannot be conceptualized satisfactorily at a single level or along a single dimension (e.g., targets sets of skills or peer acceptance) and we agreed with Waters and Sroufe (1983), and others, that accomplishing social goals using the behavioral, cognitive, and affect resources available to the child at a given point in developmental time within specific social contexts was the best working definition of the SC construct. From this perspective, measurement of SC for young children must entail multiple methods covering multiple levels of social activity. Following Waters and Sroufe's suggestion, we identified broadband measures of social engagement-motivation, peer acceptance, and skill-attribute profiles, each assessed with multiple indicators, to capture the meaning and implications of the SC construct. We anticipated finding sufficient overlap among measurement families to justify characterization of SC as a latent, second-order construct. For this report, we were able to take advantage of a unique data set involving five distinct and relatively large samples that shared a common assessment protocol. Such a data set is rare in social development research. Using data for the five samples, we posed questions concerning whether and how the hierarchical model of SC described by Bost et al. (1998) and Vaughn (2001) would hold up in new samples and whether the SC indicators showed systematic variation across age, sex, and perhaps culture.

### *Measurement and Structure of SC*

Consistent with our expectations, the multisample CFA showed that the general model fits the observed data well across the five samples. Furthermore, the measures held together at the level of first-order latent factors across the five samples and pathways were also invariant for the measurement model when sex within sample was examined (see Table 6, Models 2A and 2B). There were, however, age differences for the measurement model (Table 6, Model 2C). Our analyses of the SC composites and the factors from teachers' ratings

suggest that the measurement model may fit least well for the younger boys in the sample. For example, the two sociometric tasks used to index peer acceptance had a correlation of .37 for younger boys but ranged from .55 to .66 in the other Age  $\times$  Sex cells, and the differences among these correlations were significant. No other significant differences were found in comparisons of correlations among measurement variables across the Age  $\times$  Sex breakdowns. Perhaps because the Peer Acceptance latent variable was less well measured for younger boys, we also found lower correlations among the indicators for the first-order latent variable composites in this group than in any other Age  $\times$  Sex breakdowns, suggesting that SC may be a less coherent construct (or perhaps less reliably measured) for the younger boys (in comparison with same-age girls and all older children). If true, this may account for our finding that factors from the teachers' ratings tended to have lower levels of association with the three SC composites for younger boys than for other Age  $\times$  Gender breakdowns in the sample (although differences in correlation magnitudes across subsamples were not themselves significant).

We did not find equivalence across samples at the structural level of the model (i.e., pathways among latent variables could not be constrained to equivalence across samples, Model 3cA in Table 6). The source(s) of variation among the samples that account for this failure to converge at the structural level cannot be specified from our data. The samples differ in terms of ethnic composition, social class, age distributions, and culture (size also differs across samples) and any of these might have contributed to differences in the pathways between the SC construct and its measurement family indicators. For example, we observed some differences in the structural paths for the Dutch and Portuguese samples but no strong theoretical framework demands or explains why it is that the Social Engagement-Motivation family of measures is more saturated with SC variance for the Dutch sample (and the community sample in the United States) than the other two first-order latent factors. In fact, there were very few differences between the Dutch and Portuguese groups in terms of mean values for measured variables in the contrasts concerning sex, with the exception that Portuguese boys had significantly higher scores than Portuguese girls for receiving visual attention from peers. In the Dutch sample, this contrast did not reach significance and girls had higher mean scores. We conclude that cultural differences did not strongly

influence either the measurement or the structural aspects of the models tested.

There is also evidence for collinearity at the structural level in two of the samples (i.e., a structural coefficient exceeding 1.0). This could happen when a measured variable assigned to one first-order latent factor also has a strong relation to variables on a different first-order latent factor. Indirect support for this interpretation is found in the Age  $\times$  Sex correlations reported in Table 7. The cross-factor correlations for the Q-sort profile and Social Engagement-Motivation indicators are somewhat higher for older than for younger children; however, this is true only for boys. Thus, in this sample, SC factors become more intertwined with increasing age (at least for boys). As we continue this program of work, we plan to assess children longitudinally from preschool to primary school and this speculative hypothesis can be tested directly.

To summarize the meaning of our findings from a measurement and methods standpoint, the measurement model proposed by Bost et al. (1998) was supported by our results. In the original Head Start sample and in each of the four new samples included in our analyses, measurement factors were composed by the same set of measured variables and the measurement factors each had a significant path from the single second-order dimension (i.e., SC). In each sample, the three measurement families (Q-sort criterion scores, social engagement-motivation variables, sociometric acceptance) are at least moderately correlated. We conclude that these indicator variables are a valid set of measures that could be used in many group-care or early childhood education settings with similar results.

#### *Age, Sex, and the Meaning of SC*

We are satisfied that the hierarchical model of SC proposed by Bost et al. (1998) and Vaughn (2001) provides a good characterization of the construct as it is reflected in the day-to-day transactions of children in preschool groups. We now turn to the more substantive questions concerning the ways in which SC is manifested at younger and older age groups and how children's SC may be differentiated for boys and girls. For these purposes, we consider both the measurement family composites and the summary factor scores from the teachers' ratings. Consistent with expectations from current research reviews (see Rubin et al., 2006), socially competent children were characterized by teachers as having more positive moods, as being



more accepted by peers, more adaptable to perturbations in the social and physical surround, more engaged socially, and using more advanced social cognitive skills (Factor 1 from teachers' ratings). The scales with higher loadings on this factor overlap traditional definitions of SC (see Rubin et al., 2006) and also seem to us to satisfy the criterion for "broadband" measurement of SC suggested by Waters and Sroufe (1983). We found that, with the exception of younger boys, scores for the factor had significant positive correlations with each of the three measurement family composites from our test battery. Even for younger boys, the association between the first teachers' ratings factor and the Q-sort profiles composite yielded a significant correlation, and the other two correlations (both  $r = .20$ ) were positive and approached significance. We noted earlier that SC may have been less well measured for younger boys in this sample, but an alternative interpretation could be that younger boys were less socially competent than their female peers and less competent than older preschool children. Analyses of the teachers' ratings data provide partial support for this interpretation insofar as older children and girls received higher ratings on the items contributing to the Peer Acceptance-Positive Mood factor than did younger children and boys (although the Sex  $\times$  Age level interaction was not significant). This finding is partly corroborated by the finding that girls (both in analyses of the full sample and in the community and university-managed samples) have significantly higher scores for the sociometric acceptance variables. We interpret this to mean that younger preschool age girls are somewhat advantaged in comparison with their male peers both with regard to personal resources supportive of peer acceptance and for the consolidation of individual differences across domains of SC.

A second, and seemingly paradoxical, result from our analyses of the teachers' ratings concerns the positive SC correlates of the Negative Engagement factor. Most current construals of SC suggest that the kinds of behavior reflected in this factor (e.g., defiance of adults, aggressive or coercive behavior directed to peers, negative emotionality) are the antithesis of SC (or, at the least, undermine opportunities for the child to achieve personal goals in social contexts); nevertheless, these scores positively predicted the Q-sort profiles (for older boys only) and the Social Engagement-Motivation composites (for older but not younger children) and did not negatively predict peer acceptance in any of the Age  $\times$  Sex breakdowns of the sample (see

Table 7). Examination of the raw scores for items included in the scales making up this factor indicated that the median item value was 2 or less (with a modal value of 1–1.1), with the 75th percentile being less than 3.00. That is, the vast majority of "high"-scoring children had scores below the midpoint of most item-level ranges. Furthermore, children with high social motivation scores (i.e., those who initiate more positive and neutral interactions and receive elevated levels of visual attention from peers) also tend to be described as "negatively engaged" (see Table 7). It seems likely that older children in the present sample with mid-scale ratings from teachers were more socially engaged and were also involved in activities seen by teachers as less socially desirable as a result of their higher levels of social engagement. In previous reports (e.g., Vaughn et al., 2003), such children have been characterized as socially dominant. Hawley (2007) has argued that SC and social dominance are closely intertwined across childhood and adolescence and that overt, physical aggression is an aspect of behavior that is characteristic of socially competent preschool-age children. Additional research will be needed to determine whether this finding in our data is a reflection of increased salience of dominance relations for older preschool children.

There are additional aspects of the teacher-rated data that do not bear directly on the SC construct or on the primary emphases of this article. For example, even though we obtained teacher ratings from many different teachers and they rated a large sample of children, we were unable to reproduce the dimensional structures published by instrument authors precisely for most of the instruments used (the exception being Cairns et al.'s, 1995, Social Behavior scale for which all seven items loaded on a single dimension). Nearly 20% of the items from the total item pool were redundant (i.e., addressing the same behavior or personality construct), and several of these items had been phrased in a way that a single item from one instrument contained references to two items from a different instrument. Furthermore, many items had to be reverse coded to give the same meaning to high versus low scores. This suggests that more work on scale construction would be helpful before teams other than the originators use an item set for research. Furthermore, after redimensionalizing the items from the several questionnaires into 15 obliquely organized (i.e., correlated) dimensions, our second factor analysis collapsed these to two (somewhat negatively) related dimensions. All scales with socially desirable

content loaded on the larger "positive" teacher rating factor and all dimensions with less desirable content loaded on the "negative" factor. This suggests that teachers seem to be rating children from the basic premise that "all good things go together" (and conversely, so too do "all bad things go together"). As heuristics, these premises may serve early childhood teachers well in their day-to-day transactions with the children in their care, but they may also obscure the diverse patterns of accomplishments, proficiencies, and motivations that characterize those children, and they also obscure the potential utility of "bad things" in the construction of SC. To the extent that developmental scientists rely on teachers' evaluations and ratings of children's SC, we suggest that such ratings and evaluations should always be supplemented with direct assessments (observational or test) of the children by other objective observers.

To summarize, we find support in our data for the notion that peer SC in preschool-age children is a multilevel construct involving at least three measurement families (behavioral, cognitive, and affective skills; social engagement and motivation; and peer acceptance). These domains are adequately assessed using the battery of indicators that we chose, and the conceptual model adequately characterized SC in five relatively diverse samples of preschool children. There is, of course, more to be done to specify the model and to determine the sources of influence leading to different structural patterns and external correlates across samples, sex, and age levels. Future research should explore these issues in more detail. Finally, additional research is needed to determine whether and how SC levels attained at earlier ages predict SC and general social adaptation at later ages.

## References

- Akers, C. V. (2006). *The expression of affect as correlate and predictor of social competence for preschool children*. Unpublished master's thesis, Department of Human Development and Family Studies, Auburn University, Auburn, AL.
- Arbuckle, J. L. (2005). *Amos 6.0 user's guide*, Chicago: SPSS.
- Attili, G. (1990). Successful and disconfirmed children in the peer group: Indices of social competence within an evolutionary perspective. *Human Development, 33*, 238–249.
- Azria, M. R. (1999). *Moderators of mother-teacher disagreement on reports of young children's behavior*. Unpublished doctoral dissertation, Auburn University, Auburn, AL.
- Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). New York: Cambridge University Press.
- Bates, J. E., Freeland, C. A., & Lounsbury, M. L. (1979). Measure of infant difficulty. *Child Development, 50*, 794–803.
- Baumrind, D. (1967). Child care practices anteceding three patterns of preschool behavior. *Genetic Psychology Monographs, 75*, 43–48.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588–606.
- Block, J. (1978). *The Q-sort method in personality assessment and psychiatry research*. Palo Alto, CA: Consulting Psychologists Press. (Reprint of original work published 1961 by Charles Thomas & Sons)
- Block, J. H., & Block, J. (1980). The role of ego-control and ego-resiliency in the organization of behavior. In W. A. Collins (Ed.), *Minnesota symposium on child psychology: Vol. 13. Development of cognition affect, and social relations* (pp. 39–101), Hillsdale, NJ: Erlbaum.
- Blurton-Jones, N. (1972). *Ethological studies of child behavior*, Cambridge, UK: Cambridge University Press.
- Bost, K. K., Vaughn, B. E., Washington, W. N., Cielinski, K. L., & Bradbard, M. R. (1998). Social competence, social support, and attachment: Demarcation of construct domains, measurement, and paths of influence for preschool children attending Head Start. *Child Development, 69*, 192–218.
- Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.
- 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. 111–135), Beverly Hills, CA: Sage.
- Cairns, R. B., Leung, M., Gest, S. D., & Cairns, B. D. (1995). A brief method for assessing social development: Structure, reliability, stability, and developmental validity of the Interpersonal Competence Scale. *Behavioral Research and Therapy, 33*, 725–736.
- Charlesworth, W. R., & LaFreniere, P. (1983). Dominance, friendship, and resource utilization in preschool children's groups. *Ethology and Sociobiology, 4*, 175–186.
- Chen, X., & French, D. C. (2008). Children's social competence in cultural context. *Annual Review of Psychology, 59*, 591–616.
- Cillessen, A. H. N., & Mayeux, L. (2007). Variations in the association between aggression and social status: Theoretical and empirical perspectives. In P. H. Hawley, T. D. Little, & P. C. Rodkin (Eds.), *Aggression and adaptation: The bright side to bad behavior* (pp. 135–156), Mahwah, NJ: Erlbaum.
- Crick, N., & Dodge, K. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*, 74–101.
- 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.

- Dodge, K. A., & Somberg, D. R. (1987). Hostile attributional biases among aggressive boys are exacerbated under conditions of threats to the self. *Child Development, 58*, 213–224.
- Eibl-Eibesfeldt, I. (1989). *Human ethology*. New York: Aldine De Gruyter.
- Gray, C. A., Clancy, S., & King, L. (1981). Teacher versus parent reports of preschoolers' social competence. *Journal of Personality Assessment, 45*, 487–493.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis with readings*. Upper Saddle River, NJ: Prentice Hall.
- Hawley, P. H. (2007). Social dominance in childhood and adolescence: Why social competence and aggression may go hand in hand. In P. H. Hawley, T. D. Little, & P. C. Rodkin (Eds.), *Aggression and adaptation: The bright side to bad behavior* (pp. 1–30), Mahwah, NJ: Erlbaum.
- Ladd, G. W. (2005). *Children's peer relations and social competence: A century of progress*, New Haven, CT: Yale University Press.
- Ladd, G. W., & Price, J. M. (1987). Predicting children's social and school adjustment following the transition from preschool to kindergarten. *Child Development, 58*, 1168–1189.
- LaFreniere, P. J., & Dumas, J. E. (1996). Social competence and behavior evaluation in children ages 3 to 6 years: The Short Form (SCBE-30). *Psychological Assessment, 8*, 369–377.
- Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*, 391–410.
- McCandless, B. R., & Marshall, H. R. (1957). A picture sociometric technique for preschool children and its relation to teacher judgments of friendship. *Child Development, 28*, 139–148.
- McDowell, D. J., Parke, R. D., & Wang, S. (2003). Differences between mothers' and fathers' advice-giving style and content: Relations with social competence and psychological functioning behavior in middle childhood. *Merrill-Palmer Quarterly, 49*, 55–76.
- McGrew, W. C. (1972). *An ethological study of children's behaviour*. London: Methuen.
- O'Neil, R., Welsh, M., Parke, R. D., Wang, S., & Strand, C. (1997). A longitudinal assessment of the academic correlates of early peer acceptance and rejection. *Journal of Clinical Child Psychology, 26*, 290–303.
- Pellegrini, A. D. (2004). *Observing children in the natural worlds: A methodological primer* (2nd ed.). Mahwah, NJ: Erlbaum.
- Pellegrini, A. D., & Bartini, M. (2000). An empirical comparison of methods of sampling aggression and victimization in school settings. *Journal of Educational Psychology, 92*, 360–366.
- Pellegrini, A. D., & Long, J. D. (2002). A longitudinal study of bullying, dominance, and victimization during the transition from primary school through secondary school. *British Journal of Developmental Psychology, 20*, 259–280.
- Pellegrini, A. D., Roseth, C. J., Mliner, S., Bohn, C. M., Van Ryzin, M., Vance, N., et al. (2007). Social dominance in preschool classrooms. *Journal of Comparative Psychology, 121*, 54–64.
- Renk, K., & Phares, V. (2004). Cross-informant ratings of social competence in children and adolescents. *Clinical Psychology Review, 24*, 239–254.
- Roper, T., & Hinde, R. (1978). Social play in a play group: Consistency and complexity. *Child Development, 49*, 570–579.
- Rose-Krasnor, L. (1997). The nature of social competence: A theoretical review. *Social Development, 6*, 111–135.
- Rubin, K. H., Bukowski, W. M., & Parker, J. G. (2006). Peer interactions, relationships, and groups. In W. Damon, & R. M. Lerner (Series Eds.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 571–645). New York: Wiley.
- Rubin, K. H., & Rose-Krasnor, L. (1992). Interpersonal problem solving. In V. B. Van Hassett & M. Hersen (Eds.), *Handbook of social development* (pp. 283–323). New York: Plenum.
- Sluckin, A., & Smith, P. K. (1977). Two approaches to the concept of dominance in preschool children. *Child Development, 48*, 917–923.
- Snider, J. B. (1999). *Measurement of social competence in preschool-age children: Comparing direct assessments of child behavior and sociometric acceptance with teacher ratings*. Unpublished master's thesis, Auburn University, Auburn, AL.
- Sroufe, L. A., & Waters, E. (1977). Attachment as an organizational construct. *Child Development, 48*, 1184–1199.
- Strayer, F. F. (1980). Social ecology of the preschool peer group. In W. A. Collins (Ed.), *Minnesota symposium on child development* (Vol. 13, pp. 165–196). Hillsdale, NJ: Erlbaum.
- Tabachnick, B., & Fidell, L. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn & Bacon.
- Vaughn, B. E. (2001). A hierarchical model of social competence for preschool-age children: Cross-sectional and longitudinal analyses. *Revue Internationale de Psychologie Sociale, 14*, 13–40.
- Vaughn, B. E., Vollenweider, M., Bost, K. K., Azria-Evans, M. R., & Snider, J. B. (2003). Negative interactions and social competence for preschool children in two samples: Reconsidering the interpretation of aggressive behavior for young children. *Merrill-Palmer Quarterly, 49*, 245–278.
- Vaughn, B. E., & Waters, E. (1981). Attention structure, sociometric status, and dominance: Interrelations, behavioral correlates, and relationships to social competence. *Developmental Psychology, 17*, 275–288.
- Volling, B. V., MacKinnon-Lewis, C., Rabiner, D., & Baradaran, L. P. (1993). Children's social competence and sociometric status: Further exploration of aggression, social withdrawal, and peer rejection. *Development and Psychopathology, 5*, 549–583.

Waters, E., Garber, J., Gornal, M., & Vaughn, B. E. (1983). Q-sort correlates of social competence. *Developmental Psychology, 19*, 550–560.

Waters, E., Noyes, D., Vaughn, B. E., & Ricks, M. (1985). Q-sort definitions of social competence and self esteem: Discriminant validity of related constructs in theory and data. *Developmental Psychology, 17*, 275–288.

Waters, E., & Sroufe, L. A. (1983). Social competence as a developmental construct. *Developmental Review, 3*, 79–97.

Wright, M. J. (1982). Measuring the social competence of preschool children. *Canadian Journal of Behavioral Science, 12*, 17–3.

## Appendix

### Factor Loadings for Teacher Rated Items Used to Derive Scale Scores

Factor label	Item	Source	Factor loading
Adaptable ( $\alpha = .90$ )	How does this child typically respond to a new person? (R)	BATES	0.75
	How does this child adopt to new experiences eventually? (R)	BATES	0.73
	How does this child typically respond to being in a new place? (R)	BATES	0.72
	How does this child typically respond to new playthings? (R)	BATES	0.56
	How does this child typically respond to new foods? (R)	BATES	0.52
	How does this child respond to disruptions and changes in the everyday routine, such as when you go to on field trips, swimming, dance, or have unexpected visitors in class, etc.? (R)	BATES	0.50
Peer Acceptance ( $\alpha = .86$ )	A composite of "This child is accepted by the peer group" and "No friends (R)"	TRSS, ICS	0.81
	Other children like this child and seek him/her out for play	TRSS	0.65
	A composite of "This child gets along well with peers of same sex" and either "Popular among boys" or "Popular among girls" (scores were separately calculated for boys and girls)	TRSS, ICS	0.61
	A composite of "This child gets along well with peers of opposite sex" and either "Popular among boys" or "Popular among girls" (scores were separately calculated for boys and girls)	TRSS, ICS	0.61
	Not good at sports (R)	ICS	0.52
	Never wins (R)	ICS	0.45
Positive Mood ( $\alpha = .88$ )	Not good looking (R)	ICS	0.44
	What kind of mood is this child generally in? (R)	BATES	0.71
	Never smiles (R)	ICS	0.71
	How much does this child smile and make happy sounds? (R)	BATES	0.68
	How excited does this child become when people play with or talk to him or her? (R)	BATES	0.62
Social Cognition ( $\alpha = .94$ )	Always friendly	ICS	0.47
	Accurately interpreting what a peer is trying to do	TRSS	0.71
	Generating many solutions to interpersonal problems	TRSS	0.7
	Being aware of the effects of his behavior on others	TRSS	0.68
	Being socially aware of what is happening in a situation	TRSS	0.64
	A composite of "Generating good quality solutions to interpersonal problems" and "Negotiates solutions to conflicts with other children"	TRSS, SCBE	0.59
	Understanding others' feelings	TRSS	0.56
	Refraining from over-impulsive responding	TRSS	0.54
Engaged ( $\alpha = .66$ )	Takes other children and their points of view into account	TRSS	0.40
	A composite of "Worries (R)" and "Always worries (R)"	SCBE, ICS	0.65
	Timid, afraid (R)	SCBE	0.55
	Goes unnoticed in group (R)	SCBE	0.43

## Appendix

*Continued*

Factor label	Item	Source	Factor loading
Social Awareness ( $\alpha = .67$ )	Attentive toward younger children	SCBE	0.55
	Comforts or assists another child in difficulty	SCBE	0.43
Academic Skills ( $\alpha = .91$ )	Very good number knowledge	ICS	0.93
	Letter and word knowledge not good (R)	ICS	0.91
Desire for Adult Contact ( $\alpha = .70$ )	How much does this child want to be held?	BATES	0.70
	How much does this child cuddle and snuggle when held? (R)	BATES	0.60
Negative Emotionality ( $\alpha = .94$ )	A composite of "How much does this child cry/fuss in general?," "Screams or yells easily," and "Never cries (R)"	BATES, SCBE, ICS	0.72
	A composite of "Irritable, gets mad easily" and "How easily does this child get upset?"	SCBE, BATES	0.71
	How many times per day, on the average, does this child get fussy and irritable?	BATES	0.58
	When this child gets upset, how vigorously or loudly does she/he cry and fuss?	BATES	0.55
	How changeable is this child's mood?	BATES	0.5
	How easy of difficult is it for you to calm or soothe this child when he or she is upset?	BATES	0.49
	A composite of "Easily frustrated" and "When removed from something he or she is interested in but should not be getting into, does this child get upset?"	SCBE, BATES	0.43
	Gets angry when interrupted	SCBE	0.41
	Aggressive/Coercive Style ( $\alpha = .96$ )	A composite of "This child says mean things to peers in teasing or name-calling" and "Says uncomplimentary or unpleasant things to other children: engages in name calling, ridicule, verbal derogation"	TRSS, SBS
A composite of "Forces other children to do things they don't want to do" and "Uses coercive tactics to force the submission of peers; manipulates, threatens"	SCBE, SBS	0.73	
A composite of "This child gets into verbal arguments with peers," "Always argues," and "Argues and must have the last word in verbal exchanges"	TRSS, ICS, SBS	0.71	
Speaks to others in an impatient or cranky tone of voice	SBS	0.67	
A composite of "This child starts fights with peers," "Never gets in a fight (R)," and "Displays physical aggression toward objects or person"	TRSS, ICS, SBS	0.66	
A composite of "This child disrupts the peer group by inappropriate or attention-getting behavior" and "Disturbs other children; teases, provoke fights, interrupts others"	TRSS, SBS	0.62	
Openly strikes back with angry behavior in response to other children's teasing	SBS	0.59	
Gets into conflicts with other children	SCBE	0.58	
Hits, bites, or kicks other children	SCBE	0.40	
Defiant ( $\alpha = .83$ )	Defiant when reprimanded	SCBE	0.71
	Hits you or destroys things when angry with you	SCBE	0.7
	Opposes your suggestions	SCBE	0.56
Resistance to Control ( $\alpha = .56$ )	How does this child react to being confined?	BATES	0.54
	Does this child persist in playing with objects when she/he is told to leave them alone?	BATES	0.40

## Appendix

*Continued*

Factor label	Item	Source	Factor loading
Regularity ( $\alpha = .64$ )	How consistently does this child stick to her eating routine? (R)	BATES	0.8
	How consistently does this child stick to her sleeping routine? (R)	BATES	0.54

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.