

## Longitudinal Analyses of a Hierarchical Model of Peer Social Competence for Preschool Children

### Structural Fidelity and External Correlates

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

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

António J. Santos and Inês Peceguina *ISPA–University Institute,  
Lisbon, Portugal*

Gabrielle Coppola *G. D’Annunzio University, Chieti, Italy*

Achieving consensus on the definition and measurement of social competence (SC) for preschool children has proven difficult in the developmental sciences. We tested a hierarchical model in which SC is assumed to be a second-order latent variable by using longitudinal data ( $N = 345$ ). We also tested the degree to which peer SC at Time 1 predicted changes in positive adjustment from Time 1 to Time 2, based on teacher and peer ratings. Using a multiple-method data-collection strategy, information for three subdomains of SC (social engagement/motivation, profiles of social interaction and personality assets assessed with Q-sorts, peer acceptance) were collected across consecutive years in preschool programs. Longitudinal confirmatory factor analyses (CFAs) demonstrated invariance of both the measurement and the structural models across age levels and yielded a cross-time path weight of .74 for the second-order factor. Analyses

---

Nana Shin, Brian E. Vaughn, Mina Kim, and Lisa Krzysik, Human Development and Family Studies; Kelly K. Bost and Brent McBride, Human and Community Development; Antonio J. Santos and Inês Peceguina, Unit for the Investigation of Cognitive Psychology of Development and Education; Gabrielle Coppola, Department of Neuroscience and Imaging.

This research has been supported in part by grants from the National Science Foundation (BCS 9905926, BCS01-26163, and BCS06-23019) and by Department of Health and Human Services grant ACF 90CD0956.

Address correspondence to Brian E. Vaughn, Human Development and Family Studies, 203 Spidle Hall, Auburn University, Auburn, AL 36849. Phone: (334) 844-3235. Fax: (760) 465-7803. E-mail: vaughbe@auburn.edu.

*Merrill-Palmer Quarterly*, January 2011, Vol. 57, No. 1, pp. 73–103. Copyright © 2011 by Wayne State University Press, Detroit, MI 48201.

of latent means suggested significant increases in SC scores from the first year to second year of participation, and longitudinal cases in their second year of participation had higher scores than did age peers who entered the program as older children. Finally, Time 1 SC predicted increases from Time 1 to Time 2 for SC-relevant indicators rated by teachers and peers (standardized path coefficient of .29,  $p < .001$ ).

Even though developmental scientists endorse the notion that being or becoming “socially competent” is a critical, age-appropriate goal for preschool children, consensus regarding the definition of a “socially competent child” has not been achieved, and investigators have adopted different approaches to measuring the social competence (SC) construct (see Denham, 2006). Initial attempts to define and measure SC for young children emphasized the behavioral tactics (or social skills) that facilitated a child’s entry and integration into social groups and/or promoted the construction of supportive social relationships with peers and adults in those groups. Somewhat later, a second approach focused on interpersonal and intrapersonal outcomes or consequences of the child’s social activity in groups such as having friends, being accepted by peers, and/or enhancing self-esteem or social self-efficacy (for an empirical treatment of SC by using both types of definition/measurement approach, see Howes, 1987a; for comprehensive reviews of research from the two approaches, see also Ladd, 2005, chaps. 5 and 8).

Although both the social skills and social outcomes approaches supported generative research programs, each had weaknesses. For example, investigators defining SC as social skill(s) tended to adopt ad hoc definitions of the construct. That is, whichever behavioral tactics were relevant to the solution of a specific social challenge or puzzle constituted SC in that instance. This led to an ever-increasing list of “skills” that changed across problems, changed for the same challenge across differing social contexts, or changed as motor, cognitive, or emotional capacities matured. Likewise, investigators defining SC as outcome or consequence states must qualify the contexts in which these states occur. For example, being accepted by members of a deviant group and making friends in that group may not be signs of SC, and high self-regard based on the exercise of skills at prevarication, cheating in contests without being discovered, and defiance of adult authority may indicate psychopathology rather than competence. These problems and paradoxes prompted conceptual reconsiderations of the SC construct (e.g., Waters & Sroufe, 1983), and a third approach to defining and measuring SC was proposed.

Waters and Sroufe (1983) and others (e.g., Bost, Vaughn, Washington, Cielinski, & Bradbard, 1998; Fabes et al., 1999; Rubin & Rose-Krasnor, 1992) suggested that SC should be defined as the achievement of social

goals at a particular time and in a specific context. That is, socially competent children are those who flexibly apply the behavioral, affective, and cognitive resources available to them in the service of attaining personal social goals in salient social contexts without impeding, too much, opportunities for group co-members to achieve their own goals, and without entering onto a developmental trajectory that puts the attainment of future goals (not yet known to the child) in jeopardy (see Bost et al., 1998; Rose-Krasnor, 1997; Waters & Sroufe, 1983). Defining SC at this level of abstraction is appealing because the definition applies across a range of ages, even though the behavioral, affective, and cognitive resources available and the tactics used to apply these resources to goal achievement changes with age. The confluence of behavior, affect, and cognition recruited to the attainment of social goals constitutes the child's skill set at a given period of development, and successful application of the skill set in salient social groups is presumed to enhance self-esteem, social self-efficacy, and peer acceptance, and also makes the formation of healthy peer relationships more likely (for an empirical and conceptual illustration of these processes across the toddler and preschool years, see Howes, 1987a, 1987b). Waters and Sroufe (1983) suggested that SC should be accorded the status of an organizational construct for early childhood that was analogous to attachment security in infancy, at least in terms of its relation to the adaptive challenges of the early-childhood peer group.

A critical task for investigators adopting this approach to defining and measuring SC concerns selection of measures to assess normative growth and individual differences among children. Waters and Sroufe (1983) argued that the organizational nature of the SC construct demanded broadband measurement. That is to say, measures of SC cannot simply assess the use of specific skills in specific contexts but must capture multiple behavioral tactics relevant to a wide range of contexts. Neither can SC be uniquely specified by one or another outcome or consequence variable (e.g., peer acceptance, self-esteem, friendship quality). Ideally, measures of SC would summarize a range of intrapersonal and interpersonal behavioral tactics relevant to goal achievement in social contexts at a given developmental period and would also capture variance from the domain of outcome-consequence variables. To the extent that such broadband measures cohere, they provide evidence for inferring a latent SC dimension reflecting the quality of the child's adaptation in the peer group. Finding coherence and stability for the latent SC dimension over time and across different groups provides evidence supporting the notion that SC has become internalized and generalized for the child, again, similar to the way that internal working models of attachment (Bowlby, 1973) allow parent/child relationships

to become properties of the child that are portable across contexts and remain available even when the attachment figure is not physically present.

Two independent research programs have considered aspects of the questions and challenges posed by Waters and Sroufe (1983). For characterizing SC, Howes (1987a, 1987b) used multiple measurement strategies, including teacher ratings of child personality and behavioral attributes, quality of peer play, affect expressed in play, behaviors used to initiate interactions, friendship quality, and peer acceptance (from sociometric tasks). She reported moderate to strong correlations among measurement methods both within and across time for children ranging in age from 13 to 60 months. Although Howes reported and interpreted her findings at the level of skills and consequences-outcomes, her data are consistent with the notion that SC can be understood as an organizing construct for toddlerhood and early childhood. Moreover, in her study, the quality of SC organization at earlier ages predicted individual differences along (a developmentally reorganized) SC dimension at later ages, just as argued by Waters and Sroufe (1983).

The second research program (e.g., Bost et al., 1998; Vaughn, 2001; Vaughn et al., 2009) has focused explicitly on the structural properties of peer SC for preschool children and was intended to test models implied by Waters and Sroufe (1983). In this work, seven measurement protocols (Q-sorts, sociometric interviews, direct observations of interaction, and visual attention directed to peers) were used to assess three specific constructs (i.e., behavioral and personality profiles characteristic of socially competent preschool children; peer acceptance; and social engagement/motivation), and these three constructs were presumed to cohere under a *Social Competence* construct. Confirmatory factor analyses (CFAs) suggest that this construct is organized hierarchically for preschool children with the second-order (*Social Competence*) factor influencing the three first-order constructs and the three first-order constructs influencing scores for their measured indicators. Vaughn et al. (2009) showed that the hierarchical structure was invariant across five samples, representing a range of racial, ethnic, socioeconomic status, and sociocultural populations. The hierarchical model fit better than two alternative models (i.e., a single-factor model with all variables loading on a single, first-order factor; and a two-factor first-order model). In addition, they found that both measurement and structural aspects of the model were invariant when sex was added to the CFA in a multigroup analysis. The measurement models were also invariant over age level (i.e., younger [ $<48$  months at the beginning of the academic year] vs. older [ $\geq 48$  months at the beginning of the academic year]), however, the structural properties of the model were not invariant over age level in their study. That is, the pathways between

the SC latent variable and the intermediate-level latent variables differed across age level.

Failure to demonstrate structural invariance for SC across age level could arise for several reasons. There might be a meaningful developmental reorganization among the first-order constructs between the 36- to 48-month and 48- to 60-month age periods. Howes (1987a, 1987b) advanced this argument in her studies of SC for preschool children to accommodate the age-related changes in play quality she believed to be grounded in the cognitive developmental advances of early childhood. The cross-sectional design of Vaughn et al.'s (2009) study might have also contributed to the failure to detect structural invariance. For example, in the kindergarten sample from the Netherlands, no younger children were included, because, in the Dutch school system, children do not enter kindergarten until they reach 48 months of age, and this sample was excluded from analyses testing for invariance across age level. However, even if younger preschool children are more heterogeneous with respect to peer SC (compared to older preschoolers), we may find within-child coherence and model invariance if the same children are assessed over time. Accordingly, the central rationale for this report is to examine the structure and longitudinal stability of the hierarchical model of SC with peers in a longitudinal data set.

In two of the samples from the Vaughn et al. (2009) report, data were collected over consecutive years in the same child-care centers, and so we had access to observation and interview data for 272 children seen over both years (longitudinal cases were included at only one time point in the cross-sectional analyses reported by Vaughn et al.). Longitudinal data for these children (plus additional cases not included in the Vaughn et al. report, final  $N = 345$  for longitudinal sample) are analyzed for this study. In each sample, a common protocol (i.e., direct observations of interaction and visual attention, Q-sort descriptions, sociometric interviews) for data collection was followed, and all SC indicator variables were collected, with one exception. For the additional cases, a single Q-set was used in data collection. All measured variables had been standardized within classroom for each study, but it is possible that main effects of sex and interactions of sample, age, and sex could be observed. To test these possibilities, preliminary analyses examined each of the seven indicators with sample, sex, and age as grouping variables. When main effects of an independent variable were detected in these analyses, residuals (controlling for the factor having the effect in preliminary analyses) were calculated for the indicator prior to testing the longitudinal models.

Primary analyses test the invariance of the measurement and structural aspects of the SC construct at both times in the longitudinal data set and also test the cross-time autocorrelations among the manifest variables.

Tests on the means for Time 1 and Time 2 (using standardized scores for the measured variables) data allow an examination of possible change over time with respect to peer SC. Given that longitudinal cases are older and have more peer experience at Time 2, it follows that age changes should be directional (i.e., older children would have higher scores). We note that our strategy of standardizing values for the SC indicators within each classroom might eliminate potential age differences; however, the standardization groups included cases for whom we did not have longitudinal data (i.e., children at Time 1 who were no longer attending the child-care programs at Time 2, and children at Time 2 who had not attended the programs at Time 1). We had no way of knowing which children would become longitudinal cases at Time 2 and which would not; consequently, we anticipated that no mean differences would be detected between the two types of children for Time 1 data. However, at Time 2, children new to the programs would lack experiences specific to the program itself that were available to longitudinal cases. Thus, Time 1 vs. Time 2 differences favoring the longitudinal cases might be detectable. We designed analyses to test these possibilities.

Our final set of analyses test the degree to which peer SC at Time 1 predicts changes in child adaptation from Time 1 to Time 2, based on teacher and peer ratings. Teachers rated children's behaviors by using items adapted from several widely used questionnaires at both Time 1 and Time 2 (for details, see Vaughn et al., 2009), and children had rated how much they liked each of their peers on a 3-point scale at both time points. Two scales were formed from items rated by the teachers (peer acceptance/positive mood and classroom adjustment), and the average peer rating scale score was the third indicator for this outcome variable. The content/meanings of these variables overlap with the content/meanings of the SC indicator set, and we would expect to find that early SC should positively predict these scores at Time 2. Time 2 residual scores (i.e., residual values after the correlation between Time 1 and Time 2 measured variables is calculated) were computed to serve as the "change over time" indicators, and these scores were the measured variables for a latent variable predicted from our Time 1 latent SC variable in a structural equation model (SEM) analysis.

## Method

### *Participants*

As already noted, data for children from two different studies were used in this report. The full sample of children ( $N = 961$ ) included 493 younger (250 girls and 243 boys) and 468 older (206 girls and 262 boys). Classes

were constituted at the beginning of the academic year based on the ages of participating children. At the beginning of the academic year, children between 36 and 48 months of age ( $M = 40.2$  months,  $SD = 3.4$ ) were grouped together, and children between 48 and 60 months of age ( $M = 53.4$  months,  $SD = 3.7$ ) were grouped together. However, because we observed children throughout the academic year, many children had a birthday before being observed. To avoid confusion, we refer to the groups as *younger* and *older*. In both studies, written consent of a parent or legal guardian was obtained for every participating child.

One sample (university affiliated) consisted of 490 children (224 girls, 250 under 48 months of age at the beginning of the academic year) who were recruited from two geographic regions in the United States (Southeast and Midwest). In the Southeastern site, participants were recruited from two National Association for the Education of Young Children (NAEYC)-accredited centers managed by a major university (25 classrooms). Data for these children were included in our earlier report (Vaughn et al., 2009), but longitudinal cases were included only at 1 time point in that study. At the Midwestern site, participants were recruited from a single university-affiliated day-care center (12 classrooms). Data for these children have not been reported previously. Participation rates ranged from 80% to 100% across classrooms for both sites. A total of 230 children from this study were observed in consecutive years. The sample was ethnically diverse (approximately 32% minority) but was predominantly middle class in terms of education and income levels, by the standards of the local communities. The second sample included 471 children (232 girls, 243 under 48 months of age at the beginning of the academic year) from the Head Start sample reported by Bost et al. (1998). The children had been recruited from five different Head Start centers (30 separate classrooms) in Alabama. Of these, 115 were seen in 2 consecutive years. Participation rates averaged 90% to 100% across classrooms. All participating children were eligible for Head Start services on the basis of family income, and over 95% were African American. Within-study analyses of variance (ANOVAs) comparing children seen longitudinally with peers who were only seen once indicated that SC indicators did not distinguish the longitudinal cases from nonlongitudinal age peers in their first contact year (as "younger" children). Thus, SC level did not predict which children were retained in the participating centers during consecutive years.

### *Procedures*

All assessments took place in the day-care centers. Children were observed across all available settings (e.g., free-play and group activities in

the classroom, meals, playground, transitions between activities). All classrooms were organized similarly with science, reading, and dramatic play areas. All classrooms had tables for art activities and meals. Finally, in all classrooms an area was available for large group activities. All centers had an accessible playground, and children were given an opportunity to play outdoors at least once each morning and/or afternoon, weather permitting. To avoid distractions, interviews were conducted in private spaces outside the classrooms.

### *Social Competence Indicators*

A common set of SC indicators was assessed in both studies. These included two Q-sort descriptions (California Child Q-sort [CCQ; Block & Block, 1980]; and Preschool Q-set [PQ], Bronson's adaptation [unpublished] of a Q-sort originally used by Baumrind [1967]), direct observations of initiated interaction and visual attention to peers (Bost et al., 1998; Vaughn & Waters, 1981), and two sociometric interviews (3 like, 3 dislike nominations; paired-comparisons sociometric). With a single exception, descriptions of the measures that follow apply to both samples. In the Midwestern subsample for the first study, a single Q-sort (CCQ) description of each child was collected. Models using the full sample were computed using full information maximum likelihood (FIML) estimation for analyses with latent variables.

*Q-sort descriptions.* Q-sort observers worked in teams of two for each classroom. Each observer spent a minimum of 20 hr observing the children in a given classroom. They took notes on the behaviors and 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). When observations were completed, each child was described with both the CCQ and PQ item sets (i.e., the Q-set items were sorted into categories reflecting their salience as descriptors of a given child), according to predetermined rectangular distributions of items to nine categories. In the Head Start sample, different observers completed the CCQ-sort and the PQ-sort for each child. In the university-affiliated sample, both observers described all children by using both Q-sorts. If a child was absent from the classroom for over half a given observer's observation hours (i.e., for more than 10 hr), the observer did not complete a Q-sort for that child. Due to absences, only 852 of the 961 children from the full sample were described by using the CCQ. Due to absences and by

design (in the Midwestern subsample), only 721 children were described with the PQ. A total of 661 children were described with both Q-sets. In the longitudinal sample, 216 (of 345) children were described by using both Q-sorts in 2 consecutive years of observation.

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, and 6] of the Q-sort). Both Q-sets were sorted according to rectangular distributions with equal numbers of items in each category. The Q-sort descriptions of each child (i.e., the profile of scores for all items in each Q item set) were used to derive SC scores for each child by using the SC profiles published by Waters, Noyes, Vaughn, and Ricks (1985). Thus, the Q-sort description for a child provided by a given observer was correlated with the profile of a hypothetical child at the extreme for SC that had been generated by averaging the descriptions provided by experts in children's social development (for average item scores and the list of expert sorters, see Waters et al., 1985). The correlation between a Q-sort for a given child and the "criterion" sort for the construct becomes her or his Q-sort score for that construct. It is this Q-sort score that serves as the SC indicator for this study. This technique yields valid and reliable scores over a range of personality- and behavior-relevant constructs for children (e.g., Block, 1978; Block & Block, 1980; Waters et al., 1985). Following the rationale suggested by Waters, Garber, Gornal, and Vaughn (1983), the scores were adjusted for the social desirability response of observers by controlling for social desirability in the Q-set while calculating the correlations between individual children and the criterion sorts (i.e., these are partial correlations).

For the CCQ the Q-sort scores (social desirability controlled) averaged .06 (range,  $-.44$  to  $.58$ ) across both studies, and for the PQ the average Q-sort score was .01 (range,  $-.60$  to  $.57$ ). Cross-rater agreement for the CCQ criterion score was .59, and for the PQ the cross-rater correlation was .62 in Sample 1. In the Head Start sample, only cross-rater/cross-sort agreement was available. Cross-rater/cross-Q-sort score correlations were .59 in the university-affiliated sample and .60 in the Head Start sample. Final scores were averages across observers for each of the criterion scores for the university-affiliated sample and the single scores for each criterion score from separate observers in the Head Start sample. Scores were standardized within classroom prior to further analysis (i.e., inferential analyses all use  $z$ -scored variables).

*Initiated interactions and visual attention.* Teams of observers (between two and six for any given classroom) who worked independently from the Q-sort observation teams collected the interaction and visual

attention data. 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 interacted. Codes for the initiator and affective valence (positive, neutral, negative) of the interaction event were recorded. (Criteria for assigning an interaction code of positive, negative, or neutral are not reproduced here but have been published in the report by Vaughn et al., 2009.) All children present in the classroom during a round of observation were watched for one 15-s interval before any child was watched twice. 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, 160–228 rounds of observation in a given classroom), the total scores were converted to rate scores (i.e., by dividing the total score by the number of observation rounds for which the target child was actually observed) and standardized within classroom. Children absent for 50% or more of the observational rounds in any classroom were treated as missing for these observations. Interaction data were available for 926 children in the cross-sectional sample and for 344 children in the longitudinal sample.

Observers were trained on the observation system prior to initiating direct observations in the classroom. For most classrooms, rater agreement was estimated as the Cronbach's alpha coefficient for individual rate scores across raters. That is, the vector of rate scores from the observations of one 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 .60 to .90 for the three interaction categories (median = .79) across classrooms. For 15 classrooms, raters also conducted separate joint observations and kappa coefficients were calculated. These ranged from .78 to 1.00 (median = .87) across the three categories of interaction. 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 also collected visual attention data. Observers were instructed to intersperse rounds of interaction and visual regard observations (e.g., five interaction rounds and visual attention rounds). An observer watched a given target child for a 6-s period and recorded the identity codes for all children who were looked at by the observation target during the interval. (No child was credited with receiving more than one unit of visual regard per 6-s interval although multiple children could each receive a unit of visual regard from a given target during the interval.) No child in the classroom was observed twice before all other peers present

were observed once. In each classroom, 2–4 observers collected approximately 200 observation rounds (range, 175–225 observation rounds across classrooms). Total scores were the sum of visual regard units received by a given child from all peers (adjusted for absences, as with the interaction data). These final rate scores were then standardized within classroom. Children who were not present in the classroom for 50% or more of the observation rounds were considered as having data missing for the visual regard observations. A total of 924 children had visual attention data in the cross-sectional samples, and 342 longitudinal cases had visual attention data in both waves of observation.

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 .64 to .90 (median = .85) across all classrooms, and kappa coefficients (based on joint observations in 15 classrooms) ranged from .74 to .89 across all rater pairs with joint observation data (median = .81).

*Sociometric acceptance.* Positive 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. Photos were turned over after they were selected. Positive-choice scores were derived on the basis of the three liked choices. Average values were calculated by dividing the total number of positive choices received by the number of children making choices. A total of 911 children in the cross-sectional sample and 300 in the longitudinal sample completed this sociometric task across the two waves of data collection.

Sociometric acceptance was also scored from a paired-comparisons task. For this task, cards (or computer images) were prepared for all pairs of children in the class. The order of presentation was such that no child was seen twice before all other children were seen once. Pairs 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 appeared to wander, the assistant stopped the task and continued it later. Positive-acceptance scores were the total number of times a child was chosen by peers. These scores were averaged by dividing the total by the number of children making choices and then standardized within classroom. Paired comparisons data were available for 915 children in the cross-sectional sample and for 325 (data in both annual waves of assessments) of the longitudinal cases.

There were several reasons why a child might not have complete sociometric data. The most common reason for missing data was a child's failure to be present in the classroom, because of an illness or a vacation with the family, when sociometric photos were taken. Some younger children also failed to complete one or the other task in their initial interview and declined to continue the interview later. For all children whose data were analyzed here, sociometric choice data were available from 75% to 100% of participating class peers.

### *Positive Adjustment*

*Peer ratings.* An Asher-type rating scale measure (Asher, Singleton, Tinsley, & Hymel, 1979) was also administered to each child. The child sorted photographs of participating classmates into one of three containers. One container was for children with whom the child liked to play "a lot," a second container was for children with whom the target "sort of liked to play," and the third was for children with whom the child "did not like to play." Schematic faces were attached to each container to help the child understand the choice of meanings (e.g., a smiling face for the container that children liked to play with a lot). Children were pretrained on the meanings of the three containers by asking them to rate food items (e.g., pancakes with syrup, a sandwich, cooked mushrooms). The scores were calculated by summing the peer ratings for a given child, dividing by the number of children providing ratings, and standardizing the result within classroom.

*Teacher ratings.* Teachers in Sample 1 rated children's social behavior, their social engagement tactics, and temperament/personality by using items adapted from widely used item sets: Child Characteristics Questionnaire (ChCQ; Bates, Freeland, & Lounsbury, 1979), 32 items; Social Competence and Behavior Evaluation Scale—Short Form (SCBE; LaFreniere & Dumas, 1996), 30 items; Interpersonal Competence Scale (ICS; Cairns, Leung, Gest, & Cairns, 1995), 18 items; and Teacher Rating of Social Skills (TRSS; Dodge & Somberg, 1987), 17 items. Typically, both the lead and associate teachers in a classroom completed ratings, which were averaged for each item. The results of previous analyses of these item sets in large samples of preschool children (Akers, 2006; Vaughn et al., 2009) have suggested that the dimensional structures of these items do not correspond well with the published accounts. Consequently, we identified a subset of items to capture two dimensions *peer acceptance/positive mood* and *classroom adjustment* (i.e., social or academic skills/accomplishments). Each item was standardized to adjust for differences among scale ranges and

averaged to create indices of *peer acceptance/positive mood* and of *classroom adjustment*.

### *Analysis Plan*

Scores for each of the seven SC indicators had been standardized within classroom before analysis. It is a convention in developmental studies to standardize sociometric data within group to adjust for effects of class size. Furthermore, because the observation data come from different studies and many different observers collected these data, it is prudent to standardize the other indicators, as well. We are not so much interested in mean differences across classrooms or across studies as we are in the pattern of associations among the indicators and the concordance of these indicators across time, and introducing multilevel means (i.e., classroom, study sample) might obscure these associations and concordances. Prior to analysis, missing cases for these seven SC indicators were imputed by using the expectation maximization (EM) estimation method. Preliminary analyses are presented in three parts. First, we examined mean differences for standardized scores across sample, sex, and age to detect potential interactions among these variables. Longitudinal cases were included in these analyses at either Time 1 or Time 2 (so as to have approximately equal numbers of younger and older children from each sample in the analysis). Second, we calculated correlations among the seven SC indicators to determine within-year and cross-year associations. Finally, we tested the measurement and structural models for the cross-sectional sample by using a multigroup CFA model. The model tests hypotheses that (a) each measured variable is explained by a single first-order latent variable (i.e., Q-sort SC Scores, Social Engagement–Motivation, Peer Acceptance), and (b) the three first-order latent variables are explained by a single second-order latent factor (Social Competence). In the first model tested, error variances and disturbances were free to vary. Additional tests for this model increased the equivalence constraints. Model 2 constrained paths between measured variables and first-order latent variables to equality (measurement or metric equivalence), and Model 3 added equivalence constraints on the paths from the second-order to first-order latent variables (structural invariance).

The primary analyses are presented in three parts. First, we tested the longitudinal invariance of the SC construct using structural equations comparing the measurement models across both time points. The chi-square difference and change in the Comparative Fit Index (CFI) were used to test whether there was a significant loss of fit for the model after constraint as compared to the previous (less restrictive) model (for the rationale

concerning CFI as an index of measurement equivalence, see Cheung & Rensvold, 2002). Following the logic of our analyses for the cross-sectional data, we tested and compared three models that imposed successively more restrictive constraints on the corresponding model parameters across the two waves of data collection to detect invariance in the latent constructs over time. Model 1 tested the equality of the overall structure. In Model 2, first-order factor loadings were constrained to be invariant across time. In Model 3, second-order factor loadings were also constrained to be equal. Second, changes in the mean levels of the second-order factor across time and samples were evaluated by using the latent mean structure model and nested-model comparisons. In the final set of analyses, we assessed the degree to which peer SC at Time 1 predicts changes in *Positive Adjustment* between Time 1 and Time 2 (i.e., Time 2 residual scores). The hypothesized SEM includes two components. The Time 1 second-order SC model serves as the predictor. The second component is Time 2 Positive Adjustment, a latent variable with three indicators based on teacher (Peer Acceptance-Positive Mood and Classroom Adjustment) and peer-rated likeability. Each Time 2 indicator is a residual (controlling for the level of that indicator at Time 1) representing the changes in the indicator over time. The path from Time 1 SC to the Time 2 change score tests whether increases in Positive Adjustment can be attributed to the Time 1 SC. All SEMs analyzed in this report were estimated and tested with the AMOS 6 program (FIML option for missing data) (Arbuckle, 2005).

## Results

### *Preliminary Analyses*

Univariate ANOVAs tested sample, age, and sex effects, as well as their interactions, on scores for the seven SC indicators. A significant main effect of sex was obtained for the nominations sociometric task,  $F(1, 961) = 26.81, p < .001, \eta_p^2 = .03$ ; the paired comparisons task,  $F(1, 961) = 29.64, p < .001, \eta_p^2 = .03$ ; and the SC criterion score from the CCQ,  $F(1, 961) = 17.28, p < .001, \eta_p^2 = .02$ . Following Tabachnick and Fidell's (1989) suggestion, the partial eta-squared scores ( $\eta_p^2$ ) were used as a measure of effect size. Girls had higher sociometric acceptance scores and CCQ criterion scores. As anticipated (because standardized variables were used in the analyses), no main effects of sample or age were observed. However, a significant three-way interaction of sample, age, and sex was found for the paired comparisons sociometric acceptance score,  $F(1, 961) = 4.65, p < .05, \eta_p^2 = .01$ . A separate analysis for the university-affiliated sample and

Head Start sample revealed that a significant age by sex interaction was observed in only the university-affiliated sample,  $F(1, 490) = 9.05, p < .01, \eta_p^2 = .02$ . Effect sizes in these tests were small; nevertheless, to adjust for potential sex effects in subsequent analyses, the three indicators with main effect sex differences were residualized for use in subsequent analyses.

Correlations among the SC index variables within and across years are presented in Table 1. Within-year correlations were calculated for older and younger children, and cross-year correlations were calculated for the longitudinal cases (presented on the diagonal in Table 1). All within-year correlations reached statistical significance for both older and younger children. For the most part (with the exception of the visual regard score), correlations within a given subdomain of SC tended to be of greater magnitude than between subdomain correlations. For example, the nominations and paired-comparisons sociometric acceptance scores were more highly correlated with each other than either was with the Q-sort criterion scores or visual regard–interaction scores at both Time 1 and Time 2. Correlations between variables tended to be somewhat higher for older children, and these correlations were significantly different for 10 of 21 correlation pairs (by using  $z$  tests for  $r$  to  $z$  transformed values). All longitudinal correlations were statistically significant, suggesting modest to moderate stability of the SC indicators across consecutive preschool years. Paired sample  $t$  tests with longitudinal cases for the seven indicators revealed significant cross-year changes in standardized scores for the paired comparisons acceptance scores:  $M_{\text{time1}} = .02, M_{\text{time2}} = .14, t(344) = 2.14, p < .05, d = .12$ ; the CCQ,  $M_{\text{time1}} = -.07, M_{\text{time2}} = .13, t(344) = 3.64, p < .001, d = .22$ ; the PQ,  $M_{\text{time1}} = -.09, M_{\text{time2}} = .14, t(344) = 4.45, p < .001, d = .26$ ; visual attention received from peers,  $M_{\text{time1}} = -.07, M_{\text{time2}} = .16, t(344) = 4.29, p < .001, d = .25$ ; initiated positive interactions,  $M_{\text{time1}} = -.01, M_{\text{time2}} = .14, t(344) = 2.25, p < .05, d = .15$ ; and initiated neutral interactions,  $M_{\text{time1}} = .01, M_{\text{time2}} = .13, t(344) = 2.00, p < .05, d = .12$ . These results suggest positive growth with respect to SC indicators over time. We return to this point in the discussion.

To test the hypothesized second-order factor structure for SC, we specified a multigroup CFA, grouping cases by age level. *Social engagement/motivation* indicators were visual attention that received visual attention from peers, initiated positive, and initiated neutral interactions. The Q-sort SC criterion scores were indicators for *Profiles of socially competent attributes*, and the two sociometric scores were indicators for *Peer acceptance*. Because the  $\chi^2$  statistic and  $\chi^2/df$  ratio are sensitive to sample and model sizes (e.g., models with more estimated parameters tend to have larger chi-square values), overall model fit was assessed by using several goodness-of-fit statistics representing different classes of indices (Bollen & Long, 1993).

**Table 1.** Correlations Among the Measured Variables Within and Across the Year

	1	2	3	4	5	6	7	$M_{\text{older}}$	$SD_{\text{older}}$
1. Positive nomination	.19***	.63***	.30***	.31***	.24***	.32***	.17***	0.01	1.02
2. Paired comparison	.46***	.41***	.35***	.34***	.27***	.35***	.23***	0.02	1.01
3. CCQ criterion score	.20***	.22***	.38***	.73***	.59***	.47***	.43***	0.02	1.01
4. PQ criterion score	.23***	.26***	.76***	.41***	.63***	.47***	.45***	-0.06	0.93
5. Visual regard	.14**	.17***	.43***	.49***	.42***	.54***	.54***	-0.05	0.95
6. Positive interactions	.15**	.20***	.38***	.36***	.41***	.26***	.50***	-0.05	0.93
7. Neutral interactions	.10*	.16***	.32***	.36***	.51***	.31***	.38***	-0.04	0.96
$M_{\text{younger}}$	-0.01	-0.02	-0.02	-0.04	-0.00	-0.07	0.03		
$SD_{\text{younger}}$	0.98	0.99	0.99	0.86	0.98	0.94	1.00		

*Note.* Correlations for younger children ( $n = 493$ ) are below the diagonal, and correlations for older children ( $n = 468$ ) are above the diagonal. Cross-year correlations ( $n = 345$ ) are on the diagonal (bold font). Differences in correlation coefficients between older and younger children were tested using the Fisher  $r$  to  $z$  transformation and significant differences (10 of 21) are italicized. CCQ = California Child Q-sort; PQ = Preschool Q-set.

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

These include the Normed Fit Index (NFI), the Comparative Fit Index (CFI; Bentler, 1990), and the root mean square error of approximation (RMSEA; Steiger, 1990). Models are considered adequate if the  $\chi^2/df$  ratio is less than 3 (Kline, 1998), other fit indices are greater than .90 (Hu & Bentler, 1995), and RMSEA is less than .05 (Browne & Cudeck, 1993). Using these criteria, the two-group, second-order factor model fit the data,  $\chi^2(22\ df) = 43.96$ ,  $\chi^2/df = 2.00$ , NFI = .98, CFI = .99, and RMSEA = .03 (90% CI = .02, .05), suggesting that the model meets the criteria for configural equivalence (see Cheung, 2008). We further tested the equivalence of the model at the level of measurement and found that the two samples were invariant in terms of first order factor loadings,  $\chi^2(26\ df) = 47.27$ ,  $\chi^2/df = 1.82$ ,  $\Delta\chi^2 = 3.31$ , *ns*, NFI = .98, CFI = .99, and RMSEA = .03 (90% CI = .02, .04), using both the  $\Delta\chi^2$  and change in CFI criteria suggested by Chen (2007). Finding configural and measurement equivalence reproduces the Vaughn et al. (2009) findings. We further tested the model for equivalence at the structural level (i.e., relations between first-order and second-order factors) and found evidence of invariance across age level, as well,  $\chi^2(28\ df) = 48.17$ ,  $\chi^2/df = 1.72$ ,  $\Delta\chi^2 = 4.21$ , *ns*, NFI = .98, CFI = .99, and RMSEA = .03 (90% CI = .01, .03). Again, both the  $\Delta\chi^2$  and (lack of) change in CFI criteria suggested that the model was equivalent for younger and older children. These results indicate that our hypothesized hierarchical model of SC is consistent with the observed data for these children and suggest that the second-order SC variable is assessed satisfactorily for preschool children by using our test battery. This result differs from findings reported by Vaughn et al. (2009).

### *Analysis of Longitudinal Invariance*

We tested the longitudinal invariance of the factor structure of SC by fitting Year 1 and Year 2 data simultaneously in a longitudinal model that tested increasing levels of model invariance in sequential tests ( $N = 345$ ). In this model, covariances of error and disturbance terms for the same indicators across time were freely estimated. A series of hierarchically nested models was computed in order to test whether the second-order factor structure of SC replicated across the years. Starting with an unconstrained model, each successive model added increasingly stringent levels of invariance to the previous model solution. The first model (Model 1) was the baseline model without constraints. In Model 2, first-order factor loadings (measurement loadings) were constrained to equality across time. Model 3 tested the invariance of structural factor loadings, constraining all three second-order factor loadings to equality across time. These models and their corresponding fit statistics are displayed in Table 2.

**Table 2.** Summary Goodness-of-Fit Indices for Longitudinal Model Comparisons

Model descriptions	$\chi^2(\Delta\chi^2)^a$	$df(\Delta df)^a$	$\chi^2/df$	NFI	CFI	RMSEA (90% CI)
1. Unconstrained model	106.11***	60 (4)	1.77	.94	.97	.05 (.03, .06)
2. Invariance of measurement loadings	113.53*** (7.42)	64 (4)	1.77	.94	.97	.05 (.03, .06)
3. Invariance of structural loadings	115.29*** (1.76)	66 (2)	1.75	.94	.97	.05 (.03, .06)

Note. NFI = Normed Fit Index; CFI = Comparative Fit Index; RSEA = root mean squared error of approximation.

<sup>a</sup>Only for Model 1, the entry in this column is the  $\chi^2$  statistics and  $df$ ; for all others, the entries are  $\chi^2(\Delta\chi^2)$  (i.e., the increase of  $\chi^2$  statistic relative to the base model because of the additional invariance constraints) and  $df(\Delta df)$  ( $df$  differences between the two models).

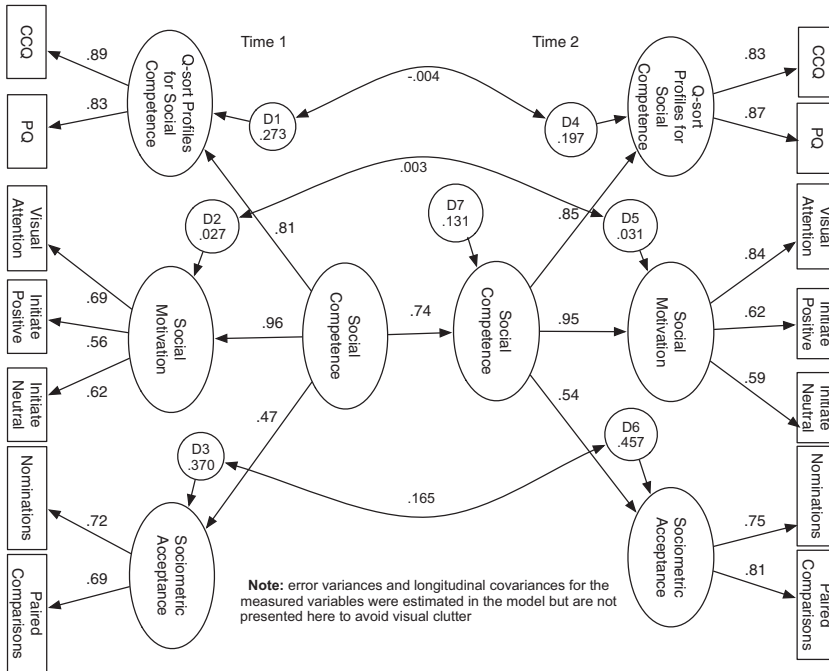
\*\*\* $p < .001$ .

All three models fit the data well; the chi-square values divided by its degree of freedom were less than 2.50, and fit indices were within the recommended cut points (see Bentler & Bonett, 1980; Browne & Cudeck, 1993; Marsh, Balla, & McDonald, 1988). The model and factor loadings for the unconstrained model are presented in Figure 1. In comparing Model 2 to Model 1, the chi-square difference test was not significant, suggesting that first-order factor loadings are invariant across time. In Model 3, both measurement and structural paths were constrained, and the chi-square difference test between Model 3 and Model 2 also was not significant, providing support for the invariance of structural factor loadings. Overall, results supported the longitudinal invariance of the second-order factor model. The stability coefficient for the second-order factor (i.e., *Social Competence* latent variable) was  $\beta = .74$  ( $p < .001$ ), indicating high rank-order stability of SC at the trait level over consecutive years.

### *Testing for Latent Mean Structure*

Latent mean structure analysis was used to test for latent mean differences across time and samples (i.e., Group 1 = Time 1 university-affiliated sample, Group 2 = Time 1 Head Start sample, Group 3 = Time 2 university-affiliated sample, Group 4 = Time 2 Head Start sample). A two-part strategy suggested by Sörbom (1974) was adopted to identify mean structures of second-order CFA models analyzed across these four groups. In the first step, the measurement model without the mean structure was simultaneously estimated across four groups. Comparisons of three models (i.e., the unconstrained model, a model with measurement factor loadings constrained across groups, and a model with measurement and structural factor loadings constrained across groups) revealed that both measurement and structural loadings were invariant across groups. The fit of the final model was good,  $\chi^2(62) = 112.3$ ,  $\chi^2/df = 1.81$ , NFI = .94, CFI = .97, and RMSEA = .03 (90% CI = .02, .05).

In the second step, the mean structure was added to the final model from the first step. The latent mean of the second-order factor was fixed to zero at Year 1 for the university-affiliated sample (i.e., reference group) and freely estimated at Year 1 in the Head Start sample (i.e., comparison group) and at Year 2 in both samples. Statistical significance of differences in latent means was determined by the critical ratios (CRs) associated with the estimate of the latent mean. An absolute value of the CR for the latent mean above 1.96 is sufficient to reject the hypothesis that there is no significant difference in mean scores between groups (Byrne, 2001). The overall fit of the model was good,  $\chi^2(80) = 142.82$ ,  $\chi^2/df = 1.79$ , NFI = .92, CFI = .96,



**Figure 1.** Longitudinal stability of the second-order factor model in preschool children. CCQ = California Child Q-sort; PQ = Preschool Q-set.

and RMSEA = .03 (90% CI = .02, .05). Given that the latent mean of the university-affiliated sample at Year 1 was fixed to zero, the estimated latent means for other groups represent latent mean differences between the reference and comparison groups. With our data, these values were  $-.02$  for the Head Start sample at Year 1,  $CR = -.30$ ;  $.13$  for the university-affiliated sample at Year 2,  $CR = 2.23$ ; and  $.14$  for the Head Start sample at Year 2,  $CR = 1.89$ . These findings suggest that the Year 2 data for the university-affiliated sample differed from the Year 1 data for this sample. A separate analysis tested the mean differences in the SC construct across years for the Head Start Sample (i.e., by using the Head Start sample at Time 1 as the reference group). The latent mean difference was not significant (estimate of the mean =  $.16$ ,  $CR = 1.78$ ). Finally, the latent mean differences in the SC construct between the samples were not statistically significant at either Time 1 or Time 2.

To distinguish longitudinal and sample effects, differences in latent means were further analyzed by comparing CFA models separately across

sample and across time to test main effects of sample and time. In the first step, analyses indicated that both measurement and structural loadings were invariant across sample,  $\chi^2(28) = 67.04$ ,  $\chi^2/df = 2.39$ , NFI = .96, CFI = .98, and RMSEA = .05 (90% CI = .03, .06); and across time,  $\chi^2(28) = 69.13$ ,  $\chi^2/df = 2.47$ , NFI = .96, CFI = .97, and RMSEA = .05 (90% CI = .03, .06). In the second step, the relative mean difference was  $-.02$  for the Head Start sample (latent SC mean for the university-affiliated sample = 0; reference sample, CR =  $-.45$ , *ns*) and  $.13$  for Year 2 (latent SC mean for Year 1 = 0; reference sample, CR =  $2.74$ ;  $p < .01$ ). These results support the findings from the latent mean structure analysis, suggesting that there is significant growth with respect to SC over time, especially in the university-affiliated sample.

We noted earlier that children in the longitudinal sample did not differ from age peers in their first year of participation in this study (i.e., as “younger” children); however, as “older” children, longitudinal participants *did* differ significantly from their age peers on six of the seven SC indicators. To probe these results further, we contrasted the older cross-sectional participants with the full sample of younger participants by using raw, rather than standardized, data. The older cross-sectional children had higher scores than younger children on six of the seven indicators, and three of these six indicators were statistically different in one-tailed tests (i.e., paired-comparisons sociometric; CCQ criterion score; initiating neutral interactions;  $t$  values 2.09, 1.72, 3.31;  $ps < .05$ ,  $.05$ ,  $.001$ , respectively). These directional changes are consistent with the notion that maturation should contribute to increases in SC. However, the differences between longitudinal and cross-sectional participants suggest that other factors may contribute to these changes. We return to this point later in the discussion.

### *Social Competence and Positive Adjustment*

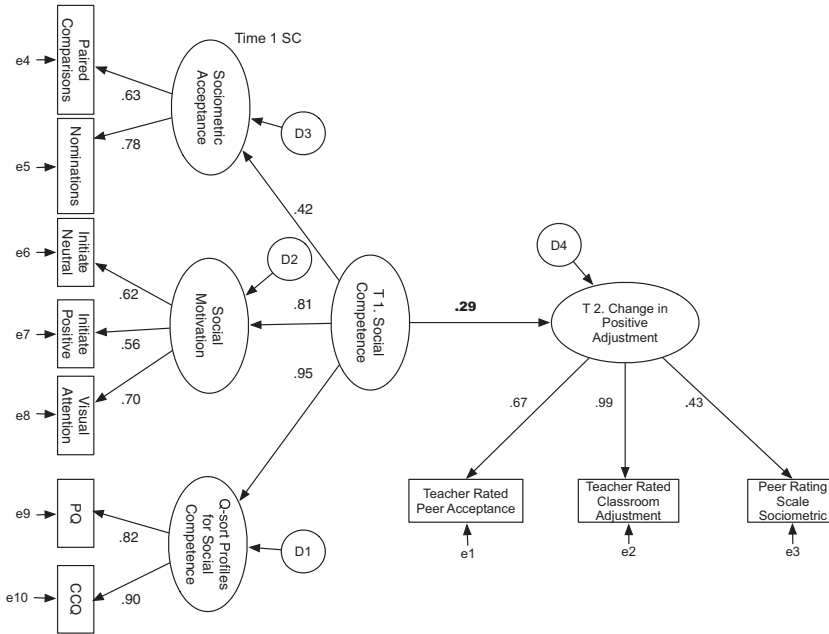
For these analyses, we examined the degree to which our model of peer SC (from Time 1) could predict scores for competence-relevant variables (from Time 2) that were not included in the model, by using the scores derived from items rated by classroom teachers (i.e., *peer acceptance*, *classroom adjustment*) and from sociometric ratings completed by children at Time 2. For each of these three competence-relevant variables, we created change indicators by residualizing Time 2 scores on Time 1 scores. Correlations among the change scores were  $.28$  ( $p < .01$ , average ratings  $\times$  peer acceptance),  $.47$  ( $p < .001$ , average ratings  $\times$  classroom adjustment), and  $.67$  ( $p < .001$ , peer acceptance  $\times$  classroom adjustment). Because of the positive associations among these variables, a first-order factor was constructed explaining these three variables and named *Positive Adjustment*.

Further, when the path loading from the first year SC latent construct to *Positive Adjustment* was added in the model (see Figure 2), it fit the data well,  $\chi^2(31) = 40.66$ ,  $\chi^2/df = 1.31$ , NFI = .95, CFI = .99, and RMSEA = .03 (90% CI = .00, .05). The standardized path coefficient from Time 1 SC latent variable to *Positive Adjustment* was significant (.29,  $p < .001$ ), suggesting that Time 1 SC predicted increases for *Positive Adjustment* from Time 1 to Time 2.

## Discussion

We noted at the outset that developmental scientists have not reached consensus on the definition of social competence (SC). Waters and Sroufe (1983) and Rose-Krasnor (1997) have suggested that a primary reason for the failure to achieve consensus was that different researchers worked at different levels of abstraction with respect to children's social behavior and SC, and that a comprehensive characterization of SC could be attained at only the highest level of abstraction (i.e., SC as an organizational construct). Our data lend themselves to the task of testing this conjecture because our assessment strategy specified broadband measures of social engagement/motivation, profiles of skills and attributes supporting SC, and peer acceptance. For each subdomain, measured variables cohere as first-order latent factors, the three subdomains themselves cohere at the level of the second-order latent factor, and the second-order latent factor is quite stable across consecutive years. Although other possible subdomains for SC might have also been assessed (e.g., the capacity to regulate behavior and emotion has been suggested as a candidate subdomain by Denham, 1998; the quality of peer play was emphasized by Howes, 1987b, in press), we are satisfied that the three classes of variables we used capture a substantial portion of the variance of the SC construct.

We include engagement/motivation (assuming that high rates of initiated interactions imply the motivation to be socially engaged) as a central feature of SC for preschool children because, without engagement, skills cannot be deployed (or assessed) and goals are not likely to be achieved. The Q-sort scores summarize a wide range of interaction skills, emotion/behavioral regulation capacity, and personality traits, as these characterize socially competent preschool children (see Waters et al., 1985). Finally, peer acceptance is a group-level variable with a wide range of established, competence-relevant predictors (see Ladd, 2005). In addition to probing the structure of SC for older and younger preschoolers, our analyses also constitute the first tests of longitudinal stability and predictive validity of SC, from the perspective of the hierarchical model of Bost et al. (1998) rather than at the level of individual indicators (Vaughn, 2001).



**Figure 2.** Predicting change in Positive Adjustment from Time 1 social competence (SC). CCQ = California Child Q-sort; PQ = Preschool Q-set.

Our analyses show that each of the subdomains is saturated with SC trait variance, although not necessarily to the same degree. Our studies (Bost et al., 1998; Vaughn, 2001; Vaughn et al., 2009) have consistently found that the *peer acceptance* latent variable has a path weight from the second-order latent variable that is less substantial (albeit significant) than for the other two latent variables in the model. This may be due to method differences (e.g., direct observations summarized as frequency/rate variables or as Q-sort scores vs. child interviews about social preferences). Alternatively, this may be due to the well-documented associations between sociometric acceptance scores and personal attributes that are not a priori indicators of SC (for examples, see Hartup, 1970, 1983). If our *peer acceptance* latent variable was influenced by these factors, that could account for its lower saturation with variance from the second-order SC latent factor.

Our findings support the hierarchical model of SC described by Bost et al. (1998) and Vaughn et al. (2009). We found that the general model fit the data for both younger and older preschool age children, and that the fit of the model was not reduced significantly when all measurement

and structural pathways were constrained to equality. We note that Vaughn et al. (2009) did not find structural invariance over age levels in their five-sample, multigroup study of SC. This may be because the range of ages among “older” children in that study was larger (from 48 to 72 months of age compared to 48 to 60 months here) and that one of Vaughn et al.’s samples did not include any younger children. Our findings indicate that invariance at both measurement and structural invariance can be obtained with samples somewhat more homogeneous with respect to age (and perhaps sociocultural context).

The longitudinal analyses also yielded evidence of invariance at both the measurement and the structural levels of the model. Furthermore, our longitudinal data provide evidence of longitudinal rank-order stability across consecutive years in preschool. That is, the longitudinal path coefficient from the SC latent variable at Time 1 to SC at Time 2 was .74, meaning that approximately 55% of the trait-level variance in SC at Time 2 was predictable from Time 1 latent scores. This is a substantial association, and it suggests that children maintain their rank-order position with respect to peers in their groups across time, even though that the group of peers differs from one year to the next (i.e., as class rosters are shuffled and teachers are changed when children make the transition from one year to the next in preschool). We would anticipate finding even greater stability in child-care settings that did not shuffle class rosters in consecutive years. These conclusions are very similar to those reached by Howes (1987a) in her longitudinal study of peer SC, even though her work adopted a different view on SC and she used different indicators to measure the SC construct.

We also found that individual differences for the SC latent factor at Time 1 predicted *increases* in the *Positive Adjustment* latent factor at Time 2. Of course, there are significant overlaps between the SC indicator variables and the scales used to construct the *Positive Adjustment* latent variable (e.g., the peer rating scale sociometric acceptance score would be expected to correlate with the other peer sociometric scores; items selected from the rating scales completed by teachers were similar to some items from the Q-sorts), and this could account for within-year associations between SC and *Positive Adjustment*. However, teachers at Time 2 had no direct knowledge of the child’s behavior at Time 1 and had no access to the observation and interview data collected at Time 1. Even so, Time 1 SC significantly predicted positive change (i.e., children with higher Time 1 SC scores had larger increases in the Time 2 *Positive Adjustment* score than did children with lower Time 1 SC scores). We interpret this result as evidence for Waters and Sroufe’s (1983) assertion that SC is an organizational construct for the preschool period.

Beyond the finding of rank-order stability, our data suggest the possibility of growth for SC across consecutive years in preschool. When compared to the reference group (i.e., university affiliated, Time 1), children in the university-affiliated sample had significantly higher scores at Time 2, and children in the Head Start sample also showed an increase (although this did not reach significance) at Time 2. Paired-sample *t* tests indicated that six of the seven indicators increased across years. These results are especially impressive because the measured variables had been standardized within classroom each year, which might have obscured cross-time changes. Our findings are consistent with the notion that children tend to increase their repertoires of skilled behavior and also become more expert at using the skills available with increasing age. However, in the case of children in the longitudinal samples studied here, children advanced in terms of both age and experience in the larger group-care setting across years. When compared to class peers as 3-year-olds, longitudinal cases did not differ in terms of any SC indicator. However, in their second year of participation, these same children *did* differ from class peers on six of the seven indicator variables even though they did not differ from those peers with respect to age and these cross-sectional age-peers did tend to have higher scores on our SC indicators than did younger children. We had not anticipated this result and did not collect information about prior child-care histories for cross-sectional children entering participating programs as “older” children. Overall, the results are consistent with the idea that both maturation and experiences in peer groups contributes to growth of SC (for a similar conclusion, see Howes, 1987a). It is also plausible that children who enter programs after 48 months of age are perceived as being less skilled by parents and this is why they show differences from age peers with more experience. Resolving this issue will require additional research that specifically targets children entering group care at different age points and includes information about prior experiences in group care.

Examining this question should be of interest to policy makers, practitioners, and parents who are making decisions about early education and care for their young children. If children in group-care settings benefit from multiple years of participation in a stable child-care program in ways that go beyond preparation for the academic challenges of formal schooling, this could justify increasing support for pre-K programs and extending them to younger children. We were able to document advances in SC for our longitudinal cases even though their class peer group changed by up to 50% for children in one sample and up to 70% in the other sample across consecutive years of participation. These findings contrast somewhat with reports using the National Institute of Child Health & Human Development

Early Child Care (NICHD ECC) data set (e.g., Belsky, 2001; Belsky et al., 2007; Dmitrieva, Steinberg, & Belsky, 2007; NICHD ECC Research Network, 2003) that have emphasized negative effects of child-care history on teachers' ratings of children's aggression and other externalizing problem behaviors.

Because we focus on competence and success in the peer group rather than on deviance and deficits, we cannot compare results directly, but it is clear that SC increased for children observed in consecutive years in preschool and they are distinguishable from children entering these programs as older preschoolers. In addition, it seems likely that program quality for the centers included in our study is higher than for the NICHD ECC study because children from the university-affiliated sample were recruited from NAEYC-accredited programs and children in the other sample were recruited from Head Start programs (which tend to receive average or above scores on standard program quality measures [Administration for Children and Families, 2005]). To the extent that program quality may be associated with externalizing problem behaviors, the range of program quality in our study may have been insufficient to detect this association. Finally, our study relies primarily on direct observations of and interviews with the participants while they are in care and only secondarily on teachers' ratings of child behavior and adjustment, which also distinguishes these data from those of the NICHD ECC study (in which the bulk of observational data concern functioning in the family and not in the child-care setting). In prior analyses (Vaughn et al., 2009), we have found that teachers' ratings of problem behavior tend to have only modest, and often nonsignificant, associations with our behaviorally based competence measures, and this may also have led us to somewhat different conclusions about the effects of child care than those emphasized by the NICHD ECC network.

Although our findings support the hypotheses about SC that motivated the study, there are limitations to be acknowledged. First, although our sample is large and diverse with respect to socioeconomic status and ethnicity, it is a convenience sample and does not represent the diversity of ethnic groups or child-care program quality at the national level. We believe that our findings would generalize in representative samples, but our confidence is based on speculation from a limited data set. We would also prefer to have additional SC indicators in each of the subdomains. Two of these have only the minimum number of variables required to specify a latent variable (*Q-sort profiles*, *peer acceptance*). It is not economically feasible to add new Q-sorts to the protocol, but it may be possible to subdivide the Q-sets, for which multiple items and dimensions of behavior and personality are summarized in a single profile score, to find homogeneous

subsets of items relevant to the SC construct that could serve as separate indicators for this measurement family. This would be analogous to the identification of item parcels (for a discussion of the use of parcels, see Little, Cunningham, Shahar, & Widaman, 2002) in CFA studies of personality and adaptation (e.g., Hawley, Little, & Card, 2008). We have also considered adding the sociometric rating scale task (used here as an outcome indicator at Time 2), but this poses difficulties for the younger children in the age range we study. Hymel (1983) argued that children under 4 years of age should not be assessed with the rating scale measure, and the results of analyses for the Head Start sample have suggested that rating scale scores were not significantly correlated with the other sociometric acceptance scores in the younger age group (Krzysik, 1996).

Also, we do not have access to data relevant to SC or child-care histories for children before they were recruited to our study. This is unfortunate because many children in the university-affiliated sample had been enrolled in care since their infancy or toddlerhood, and effects of their early care histories could be determined. Future research will correct this oversight. Child-care histories for the children in the Head Start sample may be of less concern because this sample was recruited and assessed in the early 1990s before Early Head Start was initiated, and most of the children had been cared for at home by their mothers or other relatives until they were eligible for Head Start enrollment. If a new sample were recruited from these programs, it would be important to evaluate participation in Early Head Start programs. Finally, our decision to dichotomize the sample according to age at the start of the academic year (i.e., “younger” and “older”) rather than treat age as a continuous variable is an additional limitation because some of the “younger” children were “older” (i.e.,  $\geq 48$  months of age) by the time they were actually observed/interviewed. This dichotomy was dictated in part by the necessity of making observations in different sites over the academic year (because of constraints on staff size) and in part by privacy concerns of some Head Start center directors (who were not comfortable providing identifying information for individuals, although average ages by classroom were provided).

In conclusion, we assembled a large sample of preschool age children and tested the utility of a hierarchical model of social competence (SC) by using longitudinal data. Analyses suggest that SC is a coherent and stable construct that is interpretable for both younger and older preschool children. Furthermore, individual differences assessed when children were younger tended to be stable over consecutive years. The findings are interpreted as support for the hierarchical model proposed by Bost et al. (1998) and for the conceptual model of SC described by

Waters and Sroufe (1983). Finally, our data suggest that children attending a given program in consecutive preschool years show increases in SC over time. Our findings are consistent with the notion that early childhood education can be valuable for promoting positive development in social as well as academic domains.

## References

- Administration for Children and Families. (2005). *Head Start impact study: First year findings*. Washington, DC: US Department of Health and Human Services.
- 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.
- Asher, S. R., Singleton, L. C., Tinsley, B. R., & Hymel, S. (1979). A reliable sociometric measure for preschool children. *Developmental Psychology, 15*, 443–444.
- 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 [Monograph]. *Genetic Psychology Monographs, 75*, 43–48.
- Belsky, J. (2001). Emanuel Miller lecture: Developmental risks (still) associated with early child care. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42*, 845–859.
- Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, A., McCartney, K., Owen, M. T., & the NICHD Early Child Care Research Network. (2007). Are there long-term effects of early child care? *Child Development, 78*, 681–701.
- Bentler, P. M. (1990). Fit indexes, Lagrange multipliers, constraint changes and incomplete data in structural models. *Multivariate Behavioral Research, 25*, 163–172.
- 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. (Original work published 1961.)
- 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.

- Bollen, K. A., & Long, J. S. (Eds.). (1993). *Testing structural equation models*. Newbury Park, CA: Sage.
- 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.
- Bowlby, J. (1973). *Attachment and loss: Vol. 2. Separation, anxiety, and anger*. New York: Basic Books.
- 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.
- Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.
- 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.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464–504.
- Cheung, G. W. (2008). Testing equivalence in structure, means, and variances of higher-order constructs with structural equation modeling. *Organizational Research Methods, 11*, 593–613.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indices for testing measurement equivalence. *Structural Equation Modeling, 9*, 233–255.
- Denham, S. A. (1998). *Emotional development in young children*. New York: Guilford Press.
- 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.
- Dmitrieva, J., Steinberg, L., & Belsky, J. (2007). Child-care history, classroom composition, and children's functioning in kindergarten. *Psychological Science, 18*, 1032–1039.
- 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.
- Fabes, R. A., Eisenberg, N., Jones, S., Smith, I., Guthrie, R., Poulin, S. S., et al. (1999). Regulation, emotionality, and preschoolers' socially competent peer interactions. *Child Development, 70*, 432–442.
- Hartup, W. W. (1970). Peer interaction and social organization. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology* (3rd ed., Vol. 2, pp. 361–456). New York: Wiley.

- Hartup, W. W. (1983). Peer relations. In E. M. Hetherington (Ed.) & P. H. Mussen (Series Ed.), *Handbook of child psychology: Vol. 4. Socialization, personality, and social development* (pp. 103–198). New York: Wiley.
- Hawley, P. H., Little, T. D., & Card, N. A. (2008). The myth of the alpha male: A new look at dominance-related beliefs and behaviors among adolescent males and females. *International Journal of Behavioral Development, 32*, 76–88.
- Howes, C. (1987a). Peer interaction of young children [Monograph]. *Monographs of the Society for Research in Child Development, 53* (1, Serial No. 217).
- Howes, C. (1987b). Social competence with peers in young children: Developmental sequences. *Developmental Review, 7*, 252–272.
- Howes, C. (in press). Social play of children with adults and peers. In A. Pellegrini (Ed.), *Oxford Handbook of the Development of Play*. New York: Oxford University Press.
- Hu, L.-T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Thousand Oaks, CA: Sage.
- Hymel, S. (1983). Preschool children's peer relations: Issues in sociometric assessment. *Merrill-Palmer Quarterly, 29*, 237–260.
- Kline, R. B. (1998). Principles and practice of structural equation modeling. New York: Guilford Press.
- Krzysik, L. (1996). Sociometric assessments of very young preschool children: Cross-measure convergence and correlations with social competence. Unpublished master's thesis, Auburn University, Auburn, AL.
- Ladd, G. W. (2005). *Children's peer relations and social competence: A century of progress*. New Haven, CT: Yale University Press.
- 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.
- Little, T. A., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9*, 151–173.
- 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.
- NICHD ECC Research Network. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten? *Child Development, 74*, 976–1005.

- Rose-Krasnor, L. (1997). The nature of social competence: A theoretical review. *Social Development, 6*, 111–135.
- 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 Press.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structures between groups. *British Journal of Mathematical and Statistical Psychology, 27*, 229–239.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173–190.
- Tabachnick, B., & Fidell, L. (1989). *Using multivariate statistics*. New York: Harper & Row.
- 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., Shin, N., Kim, M., Coppola, G., Krzysik, L., Santos, A. J., et al. (2009). Hierarchical models of social competence in preschool children: A multisite, multinational study. *Child Development, 80*, 1775–1796.
- 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.
- 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.

Copyright of Merrill-Palmer Quarterly is the property of Wayne State University Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.