Social Competence in Preschool Children: Replication of Results and Clarification of a Hierarchical Measurement Model

António J. Santos[1], Inês Peceguina[1], João R. Daniel[1], Nana Shin[2] and Brian E. Vaughn[2]

[1]*ISPA-Instituto Universitário de Ciências Psicológicas, Sociais e de Vida*
[2]*Auburn University*

Abstract

This study tested assumptions and conclusions reached in an earlier confirmatory factor analysis (CFA) study of the social competence (SC) construct for preschool children. Two samples (total N = 408; a new Portuguese sample and one from US samples that had participated in the original study) contributed data. Seven SC indicators were tested for mean differences across age, sex, and sample. Significant sex differences were found for peer acceptance (favoring girls) and for initiating affectively neutral interactions (boys had higher rates), and the sex by sample interaction also was significant for initiating interactions (i.e., effect significant only in the Portuguese sample). In CFAs, the hypothesized structure of SC fits the data and was invariant across sample and age within sample in both measurement and structural tests. The model was invariant at the measurement level for sex within sample tests, but not at the structural level. The results replicate and extend understandings of SC reported in the original study.

Keywords: social competence; preschool; peers/peer relations

Introduction

For many, perhaps most, young children living in contemporary urban societies, the day care or preschool group is a laboratory for the acquisition and practice of behavioral, social, cognitive, and affective skills required to negotiate relationships with children and adults outside the immediate family (e.g., Ladd & Price, 1987). Rose-Krasnor (1997) argued that these age-appropriate suites of skills are the foundation of social competence (SC) at any given age period. As such, it is important to study the behavior of children within these peer groups to document both normative growth and the emergence of individual differences in SC, and to consider status factors (e.g., age, sex, and culture) that may influence individual differences in the expression of SC.

Correspondence should be addressed to António J. Santos, ISPA-Instituto Universitário de Ciências Psicológicas, Sociais e da Vida, UIPCDE, 1149-041 Lisboa, Portugal. Email: *asantos@ispa.pt*

Although there is widespread consensus among developmental scientists concerning the value of becoming socially competent in early childhood, there has been little consensus about just what the full suite of skills might be for preschool children or what sorts of assessments best capture the range of their individual differences (see Rose-Krasnor, 1997; Waters & Sroufe, 1983 for extended discussions of these issues). Neither has consensus been achieved concerning the level of abstraction at which SC is best defined. Rose-Krasnor (1997) and Ladd (2005) both reviewed definitions of SC that ranged from concrete (e.g., specific behaviors tied to specific contexts) to abstract levels (e.g., the achievement of one's personal and social goals in social contexts). Ladd was non-committal about which kind of definition was most useful, but Rose-Krasnor as well as Waters and Sroufe (1983) concluded that the abstract levels of definition were more widely applicable across ages and contexts.

Waters and Sroufe (1983) argued that SC was an organizational construct relevant to many developmental periods. During infancy, they suggested that attachment was the domain in which competence was constructed and expressed, whereas during early childhood, competence was constructed and expressed in the context of peers. They assumed that individual differences in early attachment relationship quality would underlie (but not fully determine) SC in the peer groups of early childhood. They defined SC as the child's use of personal and interpersonal resources to obtain salient goals in social groups. Drawing on this conceptual framework, Vaughn and associates (e.g., Bost, Vaughn, Washington, Cielinski, & Bradbard, 1998; Shin et al., 2011a; Vaughn et al., 2009) proposed and supported a model of SC that assumes the construct is best defined in terms of success in attaining one's goals in social contexts. They argued that SC should be construed as a hierarchically organized latent trait integrating three broad domains of functioning for preschool-aged children: social engagement/motivation (SE/M), behavioral and personality/skill profiles, and peer acceptance (PA); and they further proposed measures for each of the three domains. They assumed that SE/M was fundamental for preschool children because unless a child actively engages peers, opportunities for learning and practicing effective means to achieve goals would be foregone, and because effective contact with peers would likely enhance their status with those peers. Vaughn et al. (2009) argued that a composite of assessments tapping these three domains would substantially overlap, although not necessarily exhaust, the meaning of SC for preschool children, as characterized by Waters and Sroufe.

Vaughn et al. (2009) tested their model using seven measured variables to index the three domains in a demographically diverse, multi-sample, and multi-national study, and reported that the hypothesized hierarchical structure fits the observed data well for all samples. Even so, the model was not invariant in some tests across age, sex, and sample. For example, although path weights in the measurement model (i.e., loadings of measured variables on first-order latent factors) were invariant across sex and sample, they differed for younger vs. older children (i.e., <48 months vs. ≥48 months at the beginning of the academic year), and the structural path weights varied for both sex and sample. Inspection of age by sex breakdowns of the correlation matrices suggested lower coherence for boys younger than 48 months of age at the start of the academic year (i.e., correlations among individual indicators and among first-order latent factor composites were lower than for other subgroups, and some were not significant). Furthermore, in one sample, correlations between the PA composite score and the other two composites tended to be lower than in the other samples.

Samples included in the Vaughn et al. (2009) study were diverse with regard to ethnic and socioeconomic status, sociocultural identity (including children from Portuguese and Dutch preschools), age ranges, and sample sizes so it is not surprising that measurement and/or structural features of the model would not be fully invariant. However, the source(s) of differences leading to measurement and structural variability could not be determined from their analyses. Again, given the implications of individual differences in SC after early childhood (see Rubin, Bukowski, & Parker, 2006), it seems important to attempt to disentangle these factors. Consequently, for this report, we reduced the dimensional diversity of the sample by examining the hierarchical model in two samples of equal size, with the same range of ages, and approximately matched for sex and social class. The data afford possibilities to test the invariance of the hierarchical model across sample (sociocultural group), sex, and age level. Finding invariance for measurement and structural models would suggest that the SC construct has a common meaning across these status categories and that subsequent studies examining individual differences and/or the development/growth of SC across developmental periods using this battery of measures could be meaningfully compared.

To replicate the model, new data were collected in Portugal (N = 204). An equal number of children were selected randomly from the subsamples of three-, four-, and five-year-old children in the combined community and university-affiliated samples from the Vaughn et al. (2009) report. This sample allows for more precise tests of possible age, sex, and culture effects than was possible from the original report. Although both Portuguese and US cultures share a common western European philosophical, religious, and cultural heritage, there are some differences in practices regarding preschool children that could lead to differences in the meaning and/or organization of SC between the two groups. For example, in the US sample, it was a common practice to reshuffle rosters of children from one academic year to the other and to change simultaneously teachers and classrooms, whereas in the Portuguese sample, the practice was to keep groups intact across academic years and to have the teacher remain with the group in the same classroom. Consequently, children in the Portuguese sample had a more continuous experience of group life during their preschool years than did their counterparts in the US sample (e.g., friend dyads were less likely to be perturbed due to changing class rosters in the Portuguese sample). This practice may or may not reflect deeper cultural differences; nonetheless, it illustrates one way that experiences in one culture might lead to differences in SC across the two samples. Importantly, the conceptual framework assumes that the measurement and structure of SC should not differ across differences in sample demographics.

In addition to testing the latent structure of SC across age, sex, and sample, Vaughn et al. (2009) reported on mean differences in the seven indicator measures for the SC construct. They reported that girls tended to receive higher scores for PA scores from sociometric data, but this main effect was qualified by an interaction with samples (i.e., the differences between boys and girls were significant in only two of the five samples). They also reported a difference for two of three SE/M-measured variables (receiving visual attention and initiating neutral interactions; boys had higher scores), but one of these (visual attention) was also qualified by an interaction with sample (i.e., the sex difference was only significant in the Portuguese sample). It is not clear whether these differences were a function of sample size differences because the US samples were between three and four times larger than the European samples in their analyses, or whether they reflected reliable sex and/or national differences. Thus, a second aim for this report was to further examine possible age, sex, and sample mean differences for

measured SC in samples equated for overall size, age levels, and sex. We also tested for potential differences (and similarities) in the correlation patterns among the SC indicators by age, sex, and sample.

To summarize, our intention was to replicate the hierarchical model of SC implied by the conceptual framework proposed by Waters and Sroufe (1983) and described by Vaughn et al. (2009) in terms of its general structure, and whether the links between measured variables and first-order latent factors as well as links between first-order and second-order latent variables were invariant across age, sex, and sample when sample diversity was more constrained than had been true in the original study. We note that replication studies are not as common as they might be in the developmental sciences, but that such studies are requirements of a vital and growing discipline. Moreover, replications with clarifications constitute logical steps in an active research program. We hypothesized that the hierarchical structure previously obtained would be reproduced in these samples and that the links (paths) among measured and latent variables would prove invariant across sample, age, and sex. From this perspective, the model (and the theory underlying the model; Waters & Sroufe, 1983) is supported if no significant differences are observed across grouping variables of sample, age, and sex. In addition, mean difference and correlation pattern contrasts for measured variables across sample, age, and sex, while controlling for socioeconomic diversity and sample size, allowed us to determine the degree to which variability among these status categories (both mean differences and correlation patterns) may have influenced the measurement and structural invariances reported in the earlier study.

Method

Participants

Participants were 408 children (202 girls and 206 boys), ages 3–5 (104 observed as three-year-olds, 168 observed as four-year-olds, and 136 observed as five-year-olds). Half of these were recruited from private Portuguese preschools near Lisbon (N = 204). Both preschools were affiliated with elementary schools. The remaining children were randomly selected from the original American sample from the Vaughn et al. (2009) study (131 from the community sample and 73 from the university-affiliated sample). The samples were matched for size, sex (101 girls and 103 boys in each sample), and age (52 at the age of 3, 84 at the age of 4, and 68 at the age of 5 in each sample). The Portuguese sample was middle class, by the standards of the local community; consequently, we selected only middle-class cases from the original US sample. All classrooms were homogeneous with respect to children's age (i.e., each having either three-, four-, or five-year-olds). Before the data collection, letters describing the project were sent to schools and parents, and the child participation only took place after signed consents from parents were returned. Participation rates ranged from 90 to 100 percent in classrooms across both samples.

The Portuguese sample was recruited from two private preschool programs affiliated with primary schools (nine different classrooms). Most children from these preschools entered the program when they were about 36 months of age and remained there (usually with the same peer group) until the fourth-grade level.

The American sample was recruited from 11 childcare and early education centers in two communities from Alabama (two were accredited by the National Association for the Education of Young Children and administered by a major university, one was a non-profit center serving middle income families, and the remaining eight were

for-profit centers serving primarily middle and working class families). One center was located in a large urban area and the others in a smaller community in the east central region of the state.

Instruments and Procedures

Social Competence Assessment. SC was assessed using seven indicators, representing three broad dimensions, namely (1) SE/M, using direct observations (rate scores for visual attention received plus positive and neutral initiated interactions); (2) behavioral and psychological attributes (BPA), using Q-sort descriptions [California child Q-sort (CCQ); Block & Block, 1980; and preschool Q-sort (PQ), Bronson's adaptation of a Q-sort originally used by Baumrind, 1967]; and (3) PA, using sociometric interviews (acceptance scores for nominations and paired comparison sociometric tasks, McCandless & Marshall, 1957).

Social Motivation and Engagement. Receiving visual attention from peers has been associated with other SC indicators in prior research (e.g., Vaughn & Martino, 1988; Vaughn & Waters, 1981) and with initiating peer interactions (Vaughn et al., 2009). Initiated positive interactions are also associated with other SC indicators including PA from sociometric data (e.g., Denham & Holt, 1993; Denham, McKinley, Couchoud, & Holt, 1990; Hartup, Glazer, & Charlesworth, 1967; Masters & Furman, 1981). Prior to data collection, each observer spent about 2 hours in the classroom to become familiar with the names of the children and also to allow the children to become familiar with him/her.

For rates of interactions initiated to peers, each child was observed during 15-second intervals. At the end of the interval, the observer registered the identification codes of all the children with whom the focal child interacted (200 observations per child over 8–10 persons/day). Additionally, the affective tone of the exchange was recorded (i.e., the interaction was characterized as a positive, neutral, or negative based on expressed affect of one or both children). To be categorized as positive, (1) one or both children had to express positive affect during the social exchange (e.g., smiles, laughs, gestures, or vocalizations indicative of positive emotions); or (2) the positive affect expression was not accompanied or followed by negative affect from the interactive partner (e.g., crying, distress, pain, and intense irritability). To be coded as negative, (1) one or both children had to clearly evidence negative affect, during the social exchange (e.g., anger, distress, fear, and sadness), whether through vocal, gestural, or facial means; or (2) the negative affect expression did not occur in the context of pretend/fantasy play [e.g., at the doll house, a child, pretending to be a mother, uses an angry tone of voice with her (pretend) husband because he was late for dinner]. Social interactions not coded as either positive or negative were considered neutral and included all the verbal and non-verbal exchanges that did not contain affect expression.

For our purposes, only positive and neutral interactions were considered indicators of SC. Initiation of negatively toned interactions were not included for two reasons. Firstly, expressed negative affect tends to be negatively associated with the PA (Shin, Vaughn, Akers, et al., 2011) and, secondly, initiated negative interactions tend to be observed infrequently (in these samples, observed in less than 5 percent of observation intervals). Scores for each child were the number of times the child was observed to initiate interaction in each of the three categories. These were divided by the number

of observation rounds for which the child was present, to adjust for absences from the class, and then were standardized within classroom group prior to substantive analysis.

Past research using this procedure has shown that observers rapidly attain agreement rates of 80 percent and above with only limited training periods (e.g., Waters, Garber, Gornall, & Vaughn, 1983). Two or more observers collected these data in each classroom. Observers watched children according to randomized class rosters with the restriction that all children present in class on a given day were observed once before any child was observed twice. A total of 200 rounds of the class lists were made in each classroom (approximately 100 observations per child per observer, when two observers completed the rounds). Cross-rater reliability coefficients (alphas) ranged from .43 to .90 across all interaction categories for the US sample. For the Portuguese sample, mean reliability estimates (alphas) were .67 for positive interactions and .77 for neutral interactions scores.

Visual Regard. Rates of visual attention received from peers were collected using randomized class lists. Observers were the same as for the interaction observations and worked independently from each other. Each watched a given child for a 6-second observation interval. At the end of the interval, the codes identifying the children who received attention from the focal child were recorded as a unit of visual attention given, namely a look (i.e., orientation of the head and/or eyes in the direction of another person for a period of 2 seconds or more) or a glance (i.e., orientation of the head and/or eyes for less than 2 seconds). If the focal child looked at a group in which a specific child could not be designated as the recipient, the orientation was registered as a doubtful occurrence (i.e., '?'). If the child looked at an object held by a peer and not directly at him/her, the orientation was also recorded as a doubtful occurrence. Questionable instances were not considered in the child total received scores.

For each round, a focal child was observed when his/her name appeared on the class roster, and no child was observed twice before; all children present in the classroom were observed once. The total score for visual attention is the sum of looks and glances that each child received from peers. The sum of visual attention units received was divided by the number of rounds that the child was present, to adjust the total score for absences. Approximately 200 rounds of visual attention observations were collected in each classroom. Final scores were standardized within classroom group prior to substantive analysis. As for the interactions, previous research has demonstrated that observers reach agreement rates of 80 percent or better (exact agreement) after a short period of observing live interactions. The alpha reliability estimate for visual regard received total scores in the Portuguese sample averaged .85 across all classrooms, and ranged from .53 to .90 (median .85) across all classrooms in the two US samples.

Behavioral and Psychological Attribute Profiles. This measurement family was assessed using the CCQ (Block & Block, 1980), a 100-item set consisting of personality- and behavior-descriptive items designed specifically for the description of young children, and the PQ (Bronson's adaptation of a Q-sort originally used by Baumrind, 1967), a 72-item set consisting of items explicitly related to social behaviors and oriented toward observable interactions. Q-sorting is an ipsative procedure for assigning scores to the items in a standard personality- or behavior-descriptive item pool (i.e., a Q-set). Both CCQ and PQ sets have been widely used to describe children's behavioral and personality characteristics (e.g., Buss, Block, & Block, 1980; Vaughn & Waters, 1981; Waters, Noyes, Vaughn, & Ricks, 1985), reflecting the child's ability to

establish and maintain positive social interactions, to manage emotionally demanding situations as well as the child's independence and self-esteem. The CCQ has also been used to assess emotional regulation and autonomy (e.g., Mendez, Fantuzzo, & Cicchetti, 2002; Shields & Cicchetti, 1997) as well as inhibition and aggressiveness (Asendorpf, Denissen, & van Aken, 2008).

Q-sort descriptions of each child were used to derive SC scores according to the criteria published by Waters et al. (1985). Q-sort observers (two in each classroom) each spent 20 hours observing the children in a variety of activity settings (e.g., small groups, meal times, free play indoors, and outdoor play). When observations were completed, assistants described the children with both CCQ-set and PQ-set. The Q-items were sorted into nine categories, with a rectangular distribution of 11 items per category for the CCQ (with the exception of the middle category, which receives 12 items) and eight items per category for the PQ. In the Portuguese sample, each observer used the CCQ to describe half of the children and the PQ for the other half, in each classroom, due to time and personnel limitations. For children included in the US samples, each observer completed both Q-sorts for every child. Vaughn et al. (2009) reported an average of cross-rater agreements of .59 and .62 for the CCQ and PQ SC criterion scores, respectively. For the Portuguese sample, average cross-rater agreement was .78 for the CCQ and .71 for the PQ. Differences between samples for rater agreement levels may have arisen because observers tended to visit the classrooms together for the Portuguese sample, whereas in the US sample, observers typically visited classrooms on different days (thus obtaining different samples of information on which to base their Q-sorts). Observers were instructed to complete sorts for all children using one Q-set before describing them using the other Q-set.

Observer Q-sorts for a given child were subsequently correlated with the profile of a hypothetical child at the extreme for SC that had been generated by aggregating descriptions provided by social development experts (Waters et al., 1985). Pearson's correlation between a Q-sort for a given child and the 'criterion' sort for the construct becomes his or her 'score' for that construct. Q-sort scores were standardized within classroom prior to subsequent analyses. For the 387 participants with CCQ descriptions, SC scores averaged .02 (range = −2.49 to 2.53), and for the 390 children with PQ descriptions, the mean SC score was .01 (range = −2.39 to 2.02) (samples combined).

Peer Acceptance. The PA dimension of SC was assessed using two sociometric interviews—peer nominations (McCandless & Marshall, 1957) and a paired comparison task (e.g., Vaughn & Waters, 1981). As with our other indicators, sociometric scores have the quality of being broadband indicators of SC (i.e., they do not assess a particular skill or social ability, reflecting instead an array of behaviors, interactions, and attributes in the context of peer relations that affect likeability). Interviews took place outside of the classroom, in a quiet area of the preschool.

For the nominations task, children were presented with photographs of all classmates and asked to name a peer with whom they especially liked to play with. The request was repeated two more times and after that; the children were asked to identify a peer with whom they did not especially like to play with (repeated again twice). As the child named a peer, the photograph of that child was turned face down. The PA score was the number of times a child was one of his/her peers' first three choices divided by the number of children making choices.

For the paired comparison task, photographs of all the possible pairs in each classroom (i.e., (N*(N−1))/2) were presented to the child being interviewed, who was

asked to choose for each pair the peer with whom she/he especially liked to play with. The pairs were randomly organized, and no child was seen twice before all classmates were seen once. Each child's photograph appeared the same number of times on the left- and right-hand sides of the picture file. The PA score for this measure was the total number of choices received from peers divided by the number of classmates who completed the task. Following conventions for classroom sociometric data, both scores were standardized within the classroom prior to substantive analysis.

Results

Initial analyses tested mean differences for the seven SC indicators using three-factor (country/sample, sex, and age) analyses of variance (ANOVA). Because each sample included cases with missing data and the patterns of 'missingness' differed across samples, and the Little's missing completely at random test suggested that 'missingness' was not completely random, we decided against using an omnibus multivariate ANOVA as the initial step. To protect against type 1 error, we adjusted the significance level for main effects in the ANOVAs to $p < .01$ (for interaction effects, we used the less conservative $p < .05$ criterion for significance). Tests for mean differences were followed by correlation analyses among the indicators separately by sample, age, and sex to determine if correlation patterns were similar across these breakdowns of the full sample (using Fisher's tests for r to z transformed values). These analyses are followed by tests of the hypothesized hierarchical model, with increasing constraints on the invariance of measurement and structural models for sample, age within sample, and sex within sample.

Variable-level Analyses

ANOVA Results. Means and standard deviations for the standardized SC indicator variables are presented in Table 1, for sample, age, and sex. Significant main effects of sex were obtained in two of the seven ANOVAs. Girls received higher sociometric scores than boys for the nominations PA measure, $F(1, 358) = 8.74$, $p < .01$, partial $\eta^2 = .02$, and boys received higher scores for initiating neutral interactions from the SE/M family, $F(1, 385) = 9.86$, $p < .01$, $\eta^2 = .03$. The main effect for initiating neutral interactions was qualified by a significant interaction with sample, $F(2, 385) = 3.25$, $p < .05$. Follow-up tests indicated that the sex difference was significant only in the Portuguese sample. No other main effects or interactions reached significance. Consistent with our expectations, these analyses suggest that sample, sex, and age mean differences for the seven SC indicators are small even when they are significant, when sample sizes and socioeconomic status are equated.

Correlation Analyses. Within-sample correlations are presented in Table 2. For both samples, all values were positive and significant, and correlations among indicators for a given dimension (e.g., PA) tended to be equal to or higher than their correlations with indicators from the other dimensions (e.g., SE/M). A single counterexample was found in the US correlation matrix; the association between visual attention received, and the SC criterion score from the 72-item Q-sort was slightly higher ($r = .51$) than the highest within-family association for visual attention ($r = .49$). Cross-sample contrasts for homologous correlation magnitudes (after r to z transformation) yielded no significant differences.

Table 1. Means for the Seven Social Competence Indicators by Sample and Sex

	Sex	VAR 3	VAR 4	VAR 5	PII 3	PII 4	PII 5	NII 3	NII 4	NII 5	PNR 3	PNR 4	PNR 5	PC 3	PC 4	PC 5	CCQ 3	CCQ 4	CCQ 5	PQ 3	PQ 4	PQ 5
PT sample	F	-.31	-.05	.05	.01	-.05	-.37	-.32	-.13	.03	.03	.06	.02	.16	.02	.11	.35	-.07	.09	-.04	-.16	-.13
		.89	.94	.98	1.20	1.06	.72	.89	1.00	1.05	1.32	1.24	1.12	.93	.81	1.02	1.02	.93	.97	.93	.99	.99
	M	.42	.05	.11	-.01	.04	.31	.42	.12	.30	-.13	-.22	-.14	-.05	-.01	-.17	-.06	-.14	-.06	.08	-.00	.10
		1.07	.95	.81	.79	.90	1.12	1.00	.96	.87	.65	.83	.83	.94	1.05	.94	.90	1.01	.77	.96	.93	.71
US sample	F	.07	.00	-.03	.12	.01	.10	.30	-.13	-.35	.50	-.00	.35	.48	-.05	-.01	.39	-.02	.00	.25	-.12	-.18
		1.01	1.18	.74	.83	.89	.77	.93	.84	.93	.84	.98	1.07	.80	1.03	.88	.72	1.11	1.01	.88	1.03	1.01
	M	-.06	.32	-.21	-.09	.12	.02	.26	.22	-.07	-.28	-.08	-.05	-.33	-.05	-.04	.02	.09	-.14	.14	.28	-.14
		.90	1.03	1.11	.84	.87	1.09	1.07	.97	1.01	.74	.91	1.15	.72	1.02	.96	.83	.91	1.15	.87	.89	1.07

Notes: Standard deviations are in second row for each cell. VAR = rates of visual attention received; PII = rates of positive interactions initiated; NII = rates of neutral interactions initiated; PNR = positive nominations received; PC = paired comparisons; CCQ = California Q-set; PQ = preschool Q-set. PT = Portuguese sample (N = 184–204); US = American sample (N = 183–195).

Table 2. Within-sample Pearson Correlations Among the Seven Social Competence Indicators

	1	2	3	4	5	6	7
1. Visual attention received		.39**	.46**	.36**	.36**	.38**	.37**
2. Positive interactions	.49**		.31**	.21*	.21*	.31**	.31**
3. Neutral interactions	.45**	.36**		.24**	.28*	.21*	.21*
4. CCQ criterion score	.44**	.30**	.26**		.63**	.27**	.29**
5. PQ criterion score	.51**	.36**	.33*	.77**		.25**	.25**
6. Nominations acceptance	.23*	.31**	.17*	.31**	.23*		.60**
7. Paired comparisons acc.	.30**	.36**	.25**	.26*	.27*	.63**	

Notes: Correlations for Portuguese sample are above the diagonal; correlations for the US sample are below the diagonal. No pairwise comparison of homologous correlations was significant.
* $p < .01$; ** $p < .001$.

Similar patterns of associations were obtained when the sample was split by sex (i.e., all values positive and significant); however, four of 21 contrasts on homologous correlation values reached significance. All four significant correlation differences involved the PA and the SE/M variable sets; in three tests, associations within the male set were lower than in the female set, and in one test, the association was larger in the male set. The table is not presented but is available from the corresponding author. The general pattern was also observed when the sample was split by age level. That is, all correlations were positive, and for the most part, correlations within measurement family were greater than cross-family correlations; however, not all values were significant in the reduced sample sizes when the sample was split into three subsets, and values were lower with more non-significant associations in the three-year-old subset than in either of the other age subset. The table of correlations is available from the corresponding author. In general, the pattern of associations among the SC indicators does not appear to vary substantially as a function of sample/culture or sex in this study, although age-level analyses suggest the possibility that cross-domain coherence among variables in the indicator set may be lower for the youngest children in the sample. The full sample CFAs explicitly tests each of these conclusions.

Model-level Analyses

Replication Analyses for the Hierarchical Model. The next series of analyses are intended to test the replicability of results reported by Vaughn et al. (2009). Mplus 5 software was used to test both the measurement and the structural components of the models, providing standard goodness-of-fit estimates between the hypothesized model and the observed data. Missing data were treated using the full information maximum likelihood (FIML) estimation procedure (Kline, 2010; Muthén & Muthén, 1998–2006). Firstly, the model was tested using CFA for the two country samples without constraints on equivalence of path coefficients for either measurement or structural components (unconstrained model). Model 1 imposed equality constraints on loadings for measured variables with the first-order latent factors, and model 2 imposes equality
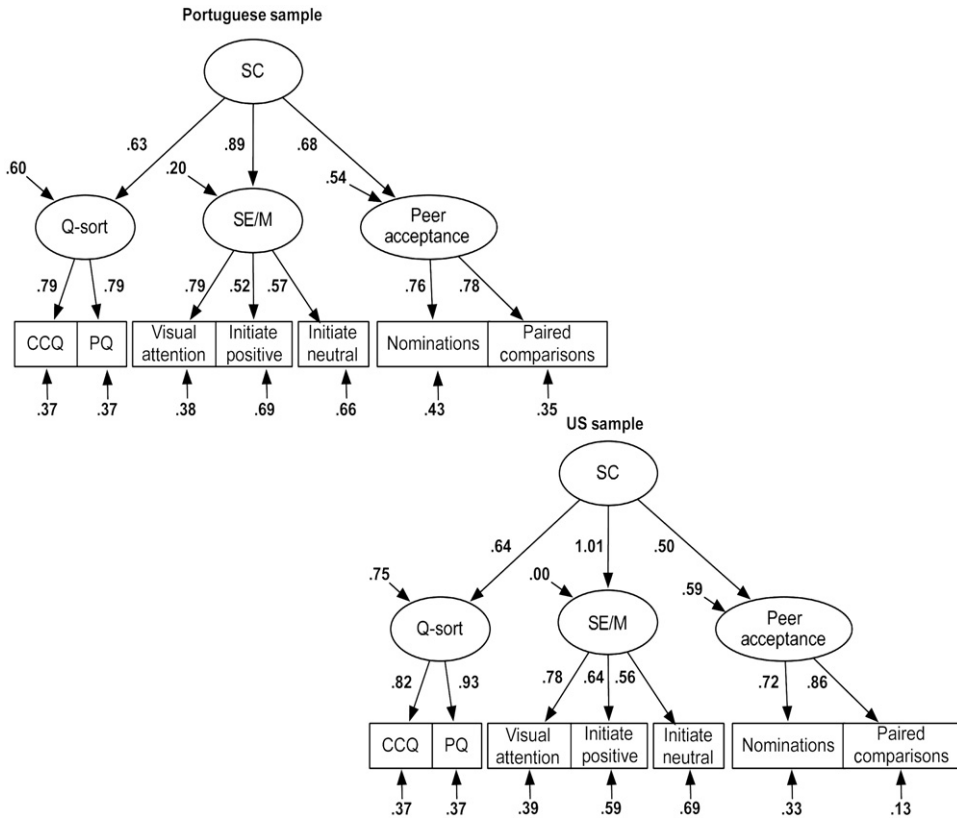
Figure 1. Hierarchical Model of Factorial Structure for Social Competence for Portuguese and American Samples With no Constraints on Path Coefficients.

constraints on both measurement structural factor loadings across samples. Figure 1 presents the measurement and structural path weights for the unconstrained model, and Table 3 presents the relative chi-square statistics and other fit indexes associated for the three models. Model fit statistics were acceptable for all tests, and the change in chi-squared index plus the Akaike information criterion (AIC) suggests that the most highly constrained model is the simplest required to explain the data. These results constitute a replication of primary findings reported by Vaughn et al., and they extend their findings by demonstrating equivalence across cultural groups for both measurement and structural components of the model. We conclude that cultural differences were not the source of diversity leading to variability in the structural components of the hierarchical model in the original report.

We next tested a model that explicitly considered the possibility of age differences in the two samples (six groups). Results are presented in Table 4. As for models presented in Table 3 above, model fit statistics were within accepted ranges, and the AIC was reduced with each successive imposition of constraint. In these analyses, both measurement and structural paths proved invariant when constrained to equivalence, suggesting that neither the cultural sample nor age of child contributed to diversity in the characterization of SC, when overall sample sizes are equal and socioeconomic status is controlled. The last set of models tested address the possibility that sex

Table 3. Goodness-of-fit Statistics for Hierarchical Model Confirmatory Factor Analysis (CFA) Comparing Portuguese and US Samples

	χ^2	df	χ^2/df	NNFI	CFI	RMSEA	SRMRR	$\Delta\chi^2$	AIC
Unconstrained model	19.84	22	.90	1.01	1	0	.04		6804.97
Model 1. Measurement constrained	20.98	26	.81	1.01	1	0	.03	1.14	6798.10
Model 2. Structural constrained	24.42	28	.87	1.01	1	0	.03	3.44	6797.55

Notes: Model 1—unconstrained model; model 1—invariance of measurement factor loadings; model 2—invariance of structural factor loadings; NNFI = non-normed fit Index; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; AIC = Akaike Information Criterion.

Table 4. Testing the Effects of Age and Sex Within Sociocultural Samples

	χ^2	df	χ^2/df	NNFI	CFI	RMSEA	SRMR	$\Delta\chi^2$	AIC
Age effects									
Unconstrained model	88.26	67*	1.32	.95	.98	.06	.05		6882.62
Model 1. measurement constrained	107.75	86	1.25	.96	.97	.06	.06	19.49	6864.11
Model 2. structural constrained	121.39	96	1.26	.96	.97	.06	.08	13.64	6857.75
Sex effects									
Unconstrained model	48.45	44	1.10	.99	1.00	.03	.04		6777.14
Model 1. measurement constrained	66.52	56	1.19	.98	.99	.04	.05	18.07	6771.21
Model 2. structural constrained	79.83	62	1.29	.97	.98	.05	.07	13.31*	6772.52

Note: Model 1—unconstrained model; model 1—invariance of measurement factor loadings; model 2—invariance of structural factor loadings; NNFI = non-normed fit index; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; AIC = Akaike Information Criterion.
* $p < .05$.

differences within samples (four groups) contribute to model diversity. The results of these analyses are also presented in Table 4. As in previous analyses, model fit statistics were within acceptable ranges for the measurement constraint; however, the AIC increased when the structural component was constrained to equality, suggesting that path weights between the first-order latent factors and the *SC* latent factor were not equivalent across boys and girls within sample. The 12-group model (age × sex within sample simultaneously) was not evaluated because the cell sizes would not be large enough to meet the recommended five cases per parameter estimated for the model.

Discussion

This study was planned as a replication and extension of the study reported by Vaughn et al. (2009) regarding the definition and measurement of the SC construct. Following suggestions made by Waters and Sroufe (1983), Vaughn and associates had defined SC in terms of success in achieving goals in social contexts and used three broadband measures as indicators of the construct. Their analyses indicated that the measures conformed to a hierarchical structure such that the higher order construct (*SC*) could be interpreted as the cause of differences in three first-order constructs (*SE/M, behavioral and personality profiles* characteristic of socially competent children, and *PA*). Each of the three first-order constructs was measured using multiple indicators. Further analyses indicated that the structure was a good fit in the five moderate- to large-sized samples they used; however, they were not able to show measurement and/or structural invariance in the model at the level of samples, sex within sample, and age within sample. Because the samples were diverse across many parameters, including size, nationality, ethnicity, and socioeconomic status, it was not possible to determine which aspect of diversity might be responsible for failing to find invariance in all tests. Thus, despite finding that the SC construct had the same general form across samples, it remained possible that core meaning(s) of SC might differ across the range of status categories included in the larger sample.

We addressed these issues by controlling for potentially critical features, including sample size, distribution by sex and age, and socioeconomic status (by selecting only more middle-class participants). Within this more homogeneous sample, we still identified a practice that distinguished the subgroups at the national level, namely the fact that classroom groups tend to be more stable over time in the Portuguese than in the US preschool groups, which might lead to observable differences in the expression and/or organization of SC. These possibilities were tested at the level of individual variables (i.e., mean and pattern differences across sample, sex, and age) as well as at the level of the model itself.

Our ANOVA results address mean difference questions. We found neither sample nor age main effects for the SC indicators, but this could be anticipated by the fact that all SC indicators were standardized within classroom. We did, however, find a significant main effect of sex for one indicator of PA and for initiation neutrally toned peer interactions. The second effect was qualified by a significant sample by sex interaction. In both samples, girls received significantly higher scores on the PA score from sociometric nominations (a similar pattern that did not reach our criterion for significance was observed for the paired comparisons acceptance score as well) and lower scores on an interaction variable from the SE/M indicator family. A *post hoc* analysis of the sex difference for initiated neutral interactions indicated that the effect was significant only in the Portuguese sample. The main effects found in this sample were also observed in the Vaughn et al. (2009) sample, but the significant sex by sample interaction was not. On the whole, and consistent with conclusions from the original report, our analyses suggest that demographic status factors do not strongly influence the levels of our SC indicators.

Correlation patterns across category breakdowns (i.e., sample, sex, and age) also suggest more similarity than differences across categories. All cross-sample and cross-sex correlations were positive and significant. No significant differences in correlation magnitudes for homologous correlations were observed in the cross-sample comparison, and only four (of 21) correlations were observed in the sex by sample breakdown.

Analyses of these patterns across age yielded different results. For the four- and five-year-old cases, the general pattern of positive, significant associations was repeated, and no two homologous correlations were significantly different from each other. For the three-year-old sample, correlations were also positive, but they tended to be modest (only 10 of 21 reached significance) and seven (of 42) tests on homologous correlations involving the three-year-old subgroup were significant. This pattern of results suggests that SC is not as well consolidated for younger preschool children perhaps because for some but not others, this is the age at which they enter preschool programs and are only beginning to assemble the behavioral, cognitive, and affective skill sets that ground SC during early childhood. Even so, within individual measurement families (i.e., Q-sorts, PA, and SE/M), indicators were significantly correlated. New research following children over time will be useful for documenting the consolidation SC that is implied by our model.

The results of our CFA analyses both replicate and extend the interpretations offered in the Vaughn et al. (2009) study. The hierarchical model proved invariant for both measurement and structural components in this sample, whereas in the original study, only the observed measurement model was invariant across groups. We attribute this difference to our participant selection decisions and equating of sample sizes. Subsequent analyses indicated that both measurement and structural components were also invariant over age levels, whereas neither was found to be invariant in the earlier study. This difference may be due to the fact that the previous study combined both the four- and the five-year age levels into a single group ('older' preschool children) rather than segregating them into unique groups, as we did. Finally, our analyses examining sex within sample replicate findings reported in the original study (i.e., measurement model invariance, but non-equivalence of the structural model). These results suggest that there may be a meaningful difference between boys and girls in the organization of the dimensions underlying peer SC. The nature of this difference is not, however, clear in these data. The significant mean differences found in our cross-sex analyses (i.e., one PA indicator, one SE/M indicator) were quite small, and only four (of 21) homologous correlations differed significantly for girls and boys (all involving the PA and SE/M variables; three of these were higher for girls). A *post hoc* breakdown of sex by sample (tables available from corresponding author) revealed somewhat different patterns by sex. For girls in both samples, the PA indicators cluster with indicators from both of the other two indicator sets, but this convergence is less apparent in the samples of boys, especially boys from the US sample. For boys in the US sample, correlations between indicators from the PA family were relatively lower than was observed for girls (both cultural groups), and six (of 10 possible) correlations are not significant. New research with different samples will be required to determine if this difference replicates and accounts for the sex differences regarding variance in the structural component of the hierarchical model.

With respect to possible influences of culture on the expression and organization of SC, our data suggest similarities rather than differences across the two sociocultural samples studied here. This is so despite differences in cultural expectations for training of preschool teachers and differences in practice regarding group integrity through time. That is, in the Portuguese education system, lead preschool teachers are required to complete a formal (five-year) course of study ending with a degree as preparation for their profession, whereas in the USA, a BA or BS is rarely a required credential for preschool teaching. Although many of the lead teachers for classrooms included in the US sample did have college degrees, most were not trained in an early childhood

education or development program. We also noted that the Portuguese preschools traditionally maintain the class roster (with the same teacher) over consecutive years of preschool and that it is a common practice to move classes *in toto* from preschool to primary school classrooms. This practice is very different from most US preschools and was not found for any classroom in our US sample. This practice means that group entrainment and group structures are likely to be preserved over longer time frames in the Portuguese than in the US samples. Given that our data suggest a common meaning and organization of SC in these two samples, we can only conclude that the cultural differences distinguishing these groups do not strongly impact individual differences in children's capacity to attain their personal and social goals within the peer group, the definition of SC suggested by Waters and Sroufe (1983). Additional research will be required to determine whether the actual goals children attempt to achieve differ across cultural boundaries and/or whether the specific tactics used in goal attainment are different (or similar). We also acknowledge that social class differences in strategies for goal achievement may differ across culture, and our sampling decisions (i.e., restricting samples to children from middle-class status families) limit the generality of our results and their interpretations.

The notion that SC is a multidimensional developmental construct relevant to children's adaptation in social groups has a long history (e.g., Howes, 1987, 1988; Rose-Krasnor, 1997; Waters & Sroufe, 1983). Our model was proposed as an attempt to instantiate this developmental construct for young children. Studies to date have provided validity, stability, and generality data for this hierarchical model of SC. The present study supports the theoretical premise that the model (and the theory from which it derives) transcends the differences in socialization practices that may distinguish our Portuguese and US samples. We expect now that the model should be useful for planning and organizing new research on the development and implications of SC for children's positive adaptation within and across cultural groups.

Developmentally, we expect that antecedents to peer SC in early childhood should be found in social relationships with adults and peers from infancy and toddlerhood (e.g., Sroufe, Egeland, Carlson, & Collins, 2005) and that individual differences for SC during the preschool years will carry implications for both social and academic adjustment as children transition to school (e.g., Ladd, 1989, 2005; Ladd, Kochenderfer, & Coleman, 1997). Shin, Vaughn, Kim, et al. (2011) showed that individual differences on the second-order SC factor were stable from three to four years of age and also were predictive of teachers' positive evaluations of children over time. Future studies should examine longitudinal stability over longer periods to determine whether additional or different dimensions are needed to characterize SC for older children and whether individual differences in SC remain stable, as anticipated by Waters and Sroufe (1983). Samples such as the Portuguese and US samples we reported in this study should be very informative in such research given the differences in classroom integrity already discussed above.

We also anticipate that the indicator dimensions we have included in this model (e.g., SE/M, PA, and BPA) will themselves have somewhat unique age-related trajectories and may assume different roles in the model and in the SC construct itself across developmental periods. For example, it seems likely that for preschool children, the SE/M dimension is especially critical because peer interaction is an essential context for building and broadening the social skills required for successfully negotiating the peer group. Children with lower SE/M may miss opportunities to tune and expand their skill sets and, as a consequence, fail to achieve age-appropriate advances in these skills.

Furthermore, because peer interactions are a principal basis for children's social preferences and PA, children less motivated to engage socially may not provide peers with sufficient information to choose them as preferred playmates or close friends (Parker & Asher, 1987). These possibilities have been explored by Rubin, Coplan and their associates (see Rubin, Coplan, & Bowker, 2009 for an extensive review) for withdrawn and inhibited children and could be studied for young children from the perspective of this model using the SE/M indicators to index low social engagement.

A final suggestion for future research concerns the means through which more and less socially competent children approach and attain (or fail to attain) their goals in social contexts. We know rather little about how children's SC facilitates their actualization of choices for preferred playmates or preferred activities in the group. Some of our SC indicators (e.g., peer sociometric acceptance and frequencies of receiving visual attention from peers) are aggregated over the entire social group. They serve well to identify socially competent children but are not themselves socially competent 'acts'. It will be important in new research to document the specific behavioral and cognitive means used by more and less socially competent children as they act on their social choices to achieve goals (e.g., engage in a preferred activity) in particular settings with specific peers and adults.

References

Asendorpf, J. B., Denissen, J. J. A., & van Aken, M. A. G. (2008). Inhibited and aggressive preschool children at 23 years of age: Personality and social transitions into adulthood. *Developmental Psychology*, *44*, 997–1011.

Baumrind, D. (1967). Child-care practices anteceding three patterns of preschool behavior. *Genetic Psychology Monographs*, *75*, 43–88.

Block, J., & Block, J. H. (1980). *The California child Q-set*. Palo Alto, CA: Consulting Psychologists 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.

Buss, D. M., Block, J. H., & Block, J. (1980). Preschool activity level: Personality correlates and developmental implications. *Child Development*, *51*, 401–408.

Denham, S. A., & Holt, R. W. (1993). Preschooler's likeability as cause or consequence of their social behavior. *Developmental Psychology*, *29*, 271–275.

Denham, S. A., McKinley, M., Couchoud, E. A., & Holt, R. (1990). Emotional and behavioral predictors of peer status in young preschoolers. *Child Development*, *61*, 1145–1152.

Hartup, W. W., Glazer, J., & Charlesworth, R. (1967). Peer reinforcement and sociometric status. *Child Development*, *38*, 1017–1024.

Howes, C. (1987). Social competence with peers in young children: Developmental sequences. *Developmental Review*, *7*, 252–272.

Howes, C. (1988). Peer interaction of young children. *Monographs of the Society for Research in Child Development, 53*. (1, Serial No. 217).

Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.

Ladd, G. W. (1989). Children's social competence and social support: Precursors of early school adjustment? In B. Schneider, G. Attili, J. Nadel, & R. Weissberg (Eds.), *Social competence in developmental perspective* (pp. 277–292). Dordrecht: Kluwer.

Ladd, G. W. (2005). *Children's peer relations and social competence: A century of progress*. New Haven, CT: Yale University Press.

Ladd, G. W., Kochenderfer, B. J., & Coleman, C. C. (1997). Classroom peer acceptance, friendship, and victimization: Distinct relational systems that contribute uniquely to children's social adjustment. *Child Development*, *68*, 1181–1197.

Ladd, J. M., & Price, J. M. (1987). Predicting children's social and school adjustment following the transition from preschool to kindergarten. *Child Development*, *58*, 1168–1189.

Masters, J. C., & Furman, W. (1981). Popularity, individual friendship, and specific peer interaction among children. *Developmental Psychology*, *17*, 344–350.

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.

Mendez, J. L., Fantuzzo, J., & Cicchetti, D. (2002). Profiles of social competence among low-income African American preschool children. *Child Development*, *4*, 1085–1100.

Muthén, L., & Muthén, B. (1998–2006). *Mplus users guide* (4th ed.). Los Angeles, CA: Muthén & Muthén.

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

Rose-Krasnor, L. (1997). The nature of social competence: A theoretical review. *Social Development*, *6*, 111–135.

Rubin, K. H., Bukowski, W., & Parker, J. G. (2006). Peer interactions, relationships, and groups. In N. Eisenberg (Ed.), *Handbook of child psychology*: *Vol. 3. Social, emotional, and personality development* (pp. 571–654). New York: Wiley.

Rubin, K. H., Coplan, R. J., & Bowker, L. (2009). Social withdrawal in childhood. *Annual Review of Psychology*, *60*, 141–171. Retrieved May 2010, from http://www.psych.annualreviews.org

Shields, A., & Cicchetti, D. (1997). Emotion regulation among school-age children: The development and validation of new criterion Q-sort scale. *Developmental Psychology*, *33*, 906–916.

Shin, N., Vaughn, B. E., Akers, V., Kim, M., Stevens, S., Krzysik, L., et al. (2011). Are happy children socially successful? Testing a central premise of positive psychology in a sample of preschool children. *Journal of Positive Psychology*, *6*, 366–367.

Shin, N., Vaughn, B. E., Kim, M., Kryzsik, L., Bost, K. K., McBride, B., et al. (2011). Longitudinal analyses of a hierarchical model of peer social competence for preschool children: Structural fidelity and external correlates. *Merrill Palmer Quarterly*, *57*, 73–103.

Sroufe, L. A., Egeland, B., Carlson, E. A., & Collins, W. A. (2005). *The development of the person: The Minnesota study of risk and adaptation from birth to adulthood*. New York: Guilford Press.

Vaughn, B. E., & Martino, D. G. (1988). Q-sort correlates of visual regard in groups of young preschool children. *Developmental Psychology*, *24*, 589–594.

Vaughn, B. E., Shin, N., Kim, M., Coppola, G., Kryzsik, 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., Gornall, M., & Vaughn, B. E. (1983). Q-sort correlates of visual regard among preschool peers: Validation of a behavioral index of social competence. *Developmental Psychology*, *19*, 550–560.

Waters, E., Noyes, D. M., 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*, *21*, 508–522.

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

Acknowledgments