



Published in final edited form as:

*Sch Psychol Q.* 2007 September ; 22(3): 380–406. doi:10.1037/1045-3830.22.3.380.

## An Examination of the Convergent and Discriminant Validity of the Strengths and Difficulties Questionnaire

Crystal R. Hill and Jan N. Hughes

Texas A&M University

### Abstract

This study examined the factor structure of the parent and teacher versions of the Strengths and Difficulties Questionnaire (SDQ; R. Goodman, 1997) with a sample of first-grade children at risk for educational failure. The 5-factor model previously found in studies using exploratory factor analysis was fit to the data for both parent and teacher questionnaires. Fit indices for both versions were marginally adequate. Model fit was comparable across gender and ethnic groups. Factor fit for the parent questionnaire was invariant across parent educational level. The examination of convergent and discriminant validity included peer evaluations of each SDQ construct. Thus, each of the five constructs was evaluated by three sources (parent, teacher, and child). On the basis of D. T. Campbell and D. W. Fiske's (1959) multitrait-multimethod approach as well as a confirmatory factor analysis using the correlated uniqueness model, the SDQ has good convergent validity but relatively poor discriminant validity.

### Keywords

Strengths and Difficulties Questionnaire; confirmatory factor analysis; correlated uniqueness model; multitrait-multimethod; validity

---

Several reports have documented the magnitude and range of risk conditions that children and youths confront in their daily lives (Annie E. Casey Foundation, 2004). Nearly one in five children in the United States live in poverty and experience the multiple, layered risks associated with poor homes and poor communities (Evans, 2004; U.S. Census Bureau, 2002). As a result of widespread risk conditions and low access to preventive or remedial mental health services, approximately 20% of children living in the United States have a diagnosable mental disorder, and 9% to 13% of children have severe emotional disorders that impair their functioning at home or school (Friedman, Katz-Leavy, Manderscheid, & Sondheimer, 1996). Consequently, at least 25% of children in the United States are at serious risk of not achieving "productive adulthood" (National Research Council, 1993).

These statistics on population-level risk among youth, combined with a rapidly expanding array of empirically supported mental health prevention programs, have led to renewed calls for an increased emphasis on deploying empirically supported prevention programs into children's natural settings. The Surgeon General's National Action Agenda for Children's Mental Health (U.S. Public Health Service, 2002) has recommended that prevention programs be embedded into schools and other child-serving institutions, and the National Institute of Mental Health (NIMH) has targeted significant resources toward efforts to disseminate such programs into

schools and other community settings. As school psychologists increasingly move into prevention and public health-oriented roles (Cummings et al., 2004; Durlak, 1995; Greenberg, Domitrovic, & Bumbarger, 2001; Hoagwood & Johnson, 2003), they experience the need for brief, reliable, and valid measures of behavioral and emotional adjustment that can be used in epidemiological and prevention research.

This study provides evidence of the psychometric properties of the Strengths and Difficulties Questionnaire (SDQ; Goodman, 2000), a measure recently included by NIMH as its epidemiological screener in the National Health Interview Survey (NHIS; NIMH, 2002, 2003, 2004). A brief screening instrument that teachers and parents can complete on all students would be of particular value to school psychologists who adopt a prevention-oriented, public health orientation to practice (Hoagwood & Johnson, 2003).

## CHARACTERISTICS IMPORTANT TO SCREENING MEASURES

Questionnaires offer an economical and valid method of obtaining information on children's behavior across settings (Dillman, 2000). When a substantial number of respondents fail to complete and return questionnaires, the utility and generalizability of the obtained data are compromised, as nonrespondents are likely to differ from respondents in ways that influence questionnaire responses (Barton et al., 1980; Srole et al., 1964). Thus, there is a need for teacher and parent questionnaires that have high user acceptability.

Some researchers have suggested that the problem of low response rate is a result, in part, of questionnaires being perceived by parents and teachers as intrusive, negative, or even insulting (Fantuzzo, McWayne, & Bulotsky, 2003). For example, parents who have consented to participate in a prevention research study for academically at-risk children may feel their privacy is invaded when asked to complete a questionnaire that requests information on their child's sexual behavior or toileting. Questionnaire length is another reason for low response rates, as shorter questionnaires are more likely to be returned (Bean & Roszkowski, 1995; Dillman, Sinclair, & Clark, 1993; Edwards et al., 2002).

The level of reading comprehension at which commonly used questionnaires are written may also contribute to low response rates. Although the average reading level for adults living in the United States is approximately eighth to ninth grade, approximately one out of five adults reads at or below the fifth-grade level (Doak, Doak, & Root, 1996). Thus, a significant number of parents may have difficulty comprehending the questionnaire items. Furthermore, because most questionnaires have been written and standardized in English only, language is also an issue. Finally, questionnaire items often require a level of inference that parents or teachers might find difficult. That is, parents or teachers do not have a reference for evaluating whether a statement (e.g., "feels too guilty") is not true, sometimes true, or very true of their child.

## STRENGTHS AND DIFFICULTIES QUESTIONNAIRE

The SDQ (Goodman, 1997) is a promising measure developed recently to screen for intervention and has more recently been used in epidemiological and prevention research with school-age youths that addresses these limitations. Indeed, the SDQ is widely used throughout the world and is increasingly being used in the United States (for articles on the SDQ go to [www.sdqinfo.com](http://www.sdqinfo.com)). Although NIMH provides normative data for the parent-completed SDQ, no normative data have been published for the teacher-completed SDQ. Two Web sites have been established providing information for this instrument, <http://www.sdqinfo.com> and <http://www.youthinmind.com>. The latter allows parents, teachers, and youths to complete the SDQ online (in English or Spanish) and receive instant feedback without charge.

The SDQ has parallel forms for parents and teachers. Because the SDQ is brief (25 items) and focuses on both positive and negative attributes of child behavior, it may be more acceptable to parents and teachers, resulting in higher response rates. The SDQ is written at a fifth-grade reading level, has no negatively worded items, and describes readily observable behaviors (see Table 2). Therefore, parents who are less well educated may have fewer difficulties completing the instrument, and the level of inference required to answer questions is low. These promising characteristics account for the fact that the SDQ has been translated into more than 30 different languages (Goodman & Scott, 1999).

A caveat must be noted about the limitations of brief screeners and thus of the SDQ. These measures are intended to be used to screen children for more comprehensive assessment and are not intended for diagnostic use. Therefore, the SDQ's sensitivity to detect specific disorders may be poor. Indeed, Goodman and Scott (1999) concluded that although the SDQ is a suitable screening measure for community studies, its limited coverage of different types of problems makes it unsuitable for studies or clinical assessments that require coverage of a broad range of childhood psychopathology. Furthermore, screening measures are typically shorter than scales designed to uncover specific psychopathologies. Shortening scales results in decreased reliability and decreased overall validity (Streiner & Norman, 1989). An example of this can be seen with the Conners' Rating Scales—Revised (Conners, 2001, p. 113). All of the shorter forms (Parent, Teacher, Self) of the Conners' Rating Scales have lower reliability than the respective longer forms (Conners, 2001).

Even though shorter screening instruments such as the Aggression, Moody, Learning Problems (Durlak, Stein, & Mannarino, 1980) have been created, most only provide a total problem score. However, the SDQ consists of five subscales, each consisting of five items: Prosocial Behaviour, Hyperactivity–Inattention, Emotional Symptoms, Conduct Problems, and Peer Relationship Problems. The SDQ, similar to the Rutter questionnaire (Rutter, 1967; Rutter, Tizard, & Whitmore, 1970) from which several of the SDQ items are derived, categorizes children as likely psychiatric “cases” or “noncases,” according to whether their total deviance score is equal to or greater than a standard cut-off. The SDQ yields a total problem score that can be used as the basis for standard cut-offs for normal, borderline, and abnormal ranges. Using these classifications, 80% of all children are considered normal, 10% of all children are considered borderline, and 10% of all children are considered abnormal or “cases.” Classifications above the 90th percentile predict a substantially higher probability of independently diagnosed psychiatric disorders (Goodman, 2001). Using these criteria, a number of studies have indicated that the SDQ is sensitive to “psychiatric caseness” (Fombonne, Simmons, Ford, Meltzer, & Goodman, 2003; Goodman, 1999; Goodman, Renfrew, & Mullick, 2000; Thabet, Stretch, & Vostanis, 2000).

International studies reporting results of exploratory factor analyses support the five-scale structure of the SDQ. Exploratory factor analysis using a nationwide epidemiological sample of 10,438 British 5- to 14-year-olds supported the five-factor structure (Goodman, 2001). Ratings were obtained from 9,998 parents and 7,313 teachers. Thabet et al. (2000) identified a similar structure in an Arabian sample. Finally, Smedje, Broman, Hetta, and von Knorring (1999) also identified the five-factor structure with a Swedish sample, but showed significant gender differences in the item-factor loadings, especially on the Prosocial Behaviour and Hyperactivity–Inattention subscales. Two previous factor analyses have been published with a U.S. sample for the parent version of the SDQ. Dickey and Blumberg (2004) conducted both a principal-components analysis as well as a confirmatory factor analysis with data provided by the parents or guardians of a national probability sample of 9,574 children and adolescents from 4 to 17 years of age. The predicted five-component structure (emotional, hyperactivity, prosocial, peer, conduct) was not entirely confirmed; many items that were intended to assess conduct problems were closely correlated to hyperactivity, and some items intended to assess

peer problems were more strongly correlated to emotional or prosocial problems. Thus a more stable three-factor model consisting of internalizing problems, externalizing problems, and positive behavior was found. In addition, Bourdon, Goodman, Rae, Simpson, and Koretz (2005) reported results that indicated good acceptability and internal consistency. However, no published study has reported results of exploratory factor analyses for the teacher version with U.S. samples, nor has the teacher version's structure been identified with confirmatory factor analysis. Cole (1987) indicated that confirmatory factor analysis can determine whether factors are redundant with one another and can address both convergent and discriminant validity, whereas exploratory factor analysis only addresses convergent validity. *Convergent validity* refers to the extent to which the same trait is measured by different methods, whereas *discriminant validity* is defined as the extent to which traits are distinct (Carmines & Zeller, 1979).

This study further examined the psychometric properties of the SDQ with an ethnically diverse sample of low-achieving first-grade students in the United States. Specifically, exploratory and confirmatory factor analyses tested the hypothesized five-factor model. Invariance of fit across gender and ethnic groups was also determined. Evidence of convergent and discriminant validity was obtained through multitrait-multimethod (MTMM) approaches and confirmatory factor analysis using a correlated trait, correlated method (CTCM) model and a correlated uniqueness (CU) model in which both trait and source factors are modeled. In addition, these methods allow for a comparison between the teacher and parent SDQ forms as well as a comparable peer report.

## METHOD

### Participants

Participants were drawn from a larger study involving 784 children attending one of three school districts in southeastern Texas participating in a longitudinal study of school achievement. School districts were chosen on the basis of their representation of the ethnic breakdown of urban areas of Texas. The student enrollment for the three schools was 42% White, 25% African American, 27% Hispanic, and 6% other. The participants were recruited across two sequential first-grade cohorts in Fall 2001 and Fall 2002. Eligibility for participation in the longitudinal study was based on a child's scoring below the school district median score on a state-approved, district-administered measure of literacy administered at the end of kindergarten or the beginning of first grade. Because children who begin kindergarten and first grade with relatively low academic readiness skills are at increased risk for emotional and behavioral problems (Braymen & Piersel, 1987; Breznitz & Teltsch, 1989; Janson, 1974), they represent a population of great concern to prevention-oriented school psychologists.

There were 1,374 children eligible for participation on the basis of scoring below the district median. Because teachers distributed consent forms to parents of eligible children via children's weekly folders, the exact number of parents who received the consent forms cannot be determined. Incentives in the form of small gifts to children and the opportunity to win a larger prize in a lottery were instrumental in obtaining 1,200 returned consent forms, of which 784 (65%) provided consent.

Analyses on a broad array of archival variables—including performance on the district-administered test of literacy (standardized within district, owing to differences in the test used), age, gender, ethnicity, eligibility for free or reduced-price lunch, bilingual class placement, cohort, and school context variables (i.e., the percentage of those who were an ethnic or racial minority; the percentage of those who were economically disadvantaged)—did not indicate any differences between children with and without consent. Although we cannot rule out differences between consenters and nonconsenters on variables not included in our data, we

can conclude that the resulting sample of 784 participants closely resembles the population from which it was drawn on demographic and literacy variables relevant to students' educational performance. Our sample consisted of 265 (34%) Caucasians, 185 (23%) African Americans, 287 (37%) Hispanics, and 47 (6%) others. A breakdown of other demographic elements of this sample appears in Table 1.

Teacher questionnaires were returned for 676 (86%) of these children. Children with and without teacher data did not differ statistically for child gender, ethnicity, economic disadvantage status, or child's limited English proficiency status.

Parent questionnaires were returned for 505 (64%) of these children. Children with and without parent data did not have any statistically significant differences for child gender, economic disadvantage status, or limited English proficiency status. However, statistically significant differences were found for ethnicity. More minority parents returned the questionnaire than did nonminority parents.

Parent permission for participation in the peer assessments was sought for all children in classes containing at least 1 child participating in the longitudinal study. Of these 6,031 children, parental permission for sociometric participation was obtained for 4,131 children (68%). Terry (1999) indicated that reliable and valid sociometric data can be collected using the unlimited nomination approach as long as at least 40% of children in a classroom participate. Therefore, sociometric scores were computed only for the 602 (77%) children in the total sample who were in classrooms in which a minimum of 40% of children participated. The mean rate of classroom participation in the sociometric administrations was .64 ( $SD = .14$ , range = .41–.95). The mean number of children participating in each classroom was 11.33 ( $SD = 2.78$ , range = 5–19 students). Children with and without peer data did not have any statistically significant differences for child gender, ethnicity, economic disadvantage status, or limited English proficiency status.

Complete parent, teacher, and peer data were available for 374 (48%) of all study participants. Children with and without complete data did not differ across child gender, economic disadvantage status, or limited English proficiency status. However, statistically significant differences were found for ethnicity. More minority children had complete data than did Caucasian children.

## Procedures

The parent version of the SDQ was mailed to parents as part of an assessment battery with a postage-paid business-reply envelope. Parents of children whose ethnicity was classified as Hispanic or Latino at school received both an English and a Spanish version of the SDQ in their mailing. Parents were paid \$25 for completing the assessment packet. The teacher version of the SDQ was given to teachers as part of an assessment battery with a business-reply envelope. Teachers were paid \$25 for completing and mailing back the packet. In individual sociometric interviews conducted at school by a member of the research team, children were asked to nominate classmates who met descriptions similar to SDQ items.

## Instruments

**SDQ**—The SDQ was originally created by modifying the Rutter Parent Questionnaire by including extra items on children's strengths (Rutter, 1967). These modifications were made as a result of Goodman's realization that many parents and teachers found the focus of the Rutter items disconcerting because all probes in the Rutter questionnaires are about undesirable traits, and respondents indicated to the researchers that they wanted to identify the child's strengths as well as weaknesses (Goodman, 1994). Therefore, to increase the acceptability of

the measure and the response rates in their epidemiological study of childhood hemiplegia, Goodman and his colleagues added to the scale extra items from the Prosocial Behavior Questionnaire (Weir & Duveen, 1981) on children's strengths as well as items that were atypical of psychopathology.

Although no item from the SDQ is identical to a Rutter item or Prosocial Behavior Questionnaire item, 5 items are similarly worded. The initial choice of items was guided by the previous factor analyses and frequency distributions from the expanded Rutter questionnaire that Goodman and his colleagues created (Goodman, 1994). However, after a succession of informal trials, Goodman subsequently modified the SDQ into its current 25-item form (Goodman, 1997). Approximately half of the items cover strengths. Each item is coded 0, 1, or 2, corresponding to *not true*, *somewhat true*, or *certainly true*. All items are reproduced in Table 2. The five items in italics are reversed scored.

Goodman (1997) stated that the questionnaire consists of five scales, each of which consists of five items: Hyperactivity–Inattention, Emotional Symptoms, Conduct Problems, Peer Relationship Problems, and Prosocial Behaviour. The scores for Hyperactivity–Inattention, Emotional Symptoms, Conduct Problems, and Peer Relationship Problems are summed to generate a Total Difficulties score. Goodman indicated that the Prosocial Behaviour score is not incorporated into the Total Difficulties score because the absence of prosocial behaviors is seen as conceptually different from the presence of psychological difficulties.

**Sociometric Assessment**—A modified version of the Class Play (Masten, Morison, & Pelligrini, 1985) and a roster rating of liking for classmates were used to obtain children's evaluations of classmates' behavioral characteristics on dimensions similar to those assessed by the SDQ. Research assistants individually interviewed children at school. Children were asked to nominate as few or as many classmates as they wished who could best play each of the following parts in a class play: trouble ("Some kids get into trouble a lot"), aggression ("Some kids start fights, say mean things, or hit others"), hyperactivity ("Some kids do strange things and make a lot of noise. They bother people who are trying to work"), prosocial ("Some kids help others, play fair, and share"), and emotional symptoms ("Some kids cry a lot and look sad"). The trouble and aggression items were averaged to create a conduct problems score. Peer problems were measured by social preference scores (Coie, Dodge, & Coppotelli, 1982). Social preference scores were computed as the standardized liked-most nomination score (using unlimited nominations) minus the standardized liked-least scores. To avoid asking children to nominate disliked children, children were provided a class roster and asked to rate how much they liked each child on a 5-point scale (1 = *don't like at all* and 5 = *like very much*). A rating of 1 was considered equivalent to a liked-least nomination score (Asher & Dodge, 1986). All sociometric scores were standardized within classrooms.

Elementary school children's peer nomination scores derived from procedures similar to those used in this study have been found to be stable over periods from 6 weeks to 4 years and to be associated with concurrent and future behavior and adjustment (for a review, see Hughes, 1990).

## RESULTS

### Overview of Analyses

Analyses were conducted in two phases. The first series of analyses examined the factor structure of the SDQ using confirmatory factor analysis. Second, multigroup comparison analyses were conducted using latent variable structural equation modeling with AMOS version 4.0 (Arbuckle & Wothke, 1999) to evaluate the potential moderating effect of gender and ethnic group on the teacher questionnaire as well as gender, ethnic group, and parental

level of education on the parent questionnaire. Next, the convergent and discriminant validities of the scales were examined using a MTMM approach (Campbell & Fiske, 1959) as well as confirmatory factor analysis (Marsh & Grayson, 1995). These analyses included a peer evaluation of each SDQ construct. Thus, each of the five constructs was evaluated by three sources (parent, teacher, and child). The MTMM matrix allows a visual examination of the variance that is due to traits, variance that is due to methods, and unique or error variance (Kenny & Kashy, 1992). This analysis is enhanced by a confirmatory factor analysis that permits less ambiguous conclusions about trait and method variances.

### Preliminary Investigation Comparing Our SDQ Scores With Normative Data

In addition, scores for the SDQ Parent Questionnaire in this study can be compared with the means and standard deviations of a U.S. sample found at <http://www.sdqinfo.com>. Our results are represented in Table 3. None of these scores is within a 95% confidence interval of the sample presented online. As expected given the at-risk status of our sample, our sample exhibited more problem behaviors and less prosocial behavior than the U.S. normative sample.

### Measurement Characteristics and Factor Structure of the SDQ

**Teacher Questionnaire**—For the 676 participants with teacher SDQ data, structural equation modeling with AMOS (Arbuckle & Wothke, 1999) tested whether the SDQ's factor structure confirmed the five-factor model found in previous exploratory factor analyses: Hyperactivity–Inattention, Emotional Symptoms, Conduct Problems, Peer Relationship Problems, and Prosocial Behaviour. The model converged, and all estimates were within bounds. Model fit was evaluated with multiple indicators of model fit. Hu and Bentler (1999) suggested that comparative fit index (CFI) and Tucker-Lewis Index (TLI) values above .95 and root mean square errors of approximation (RMSEA) values less than .08 represent acceptable fit. Fit indices all approached these levels:  $\chi^2(265, N = ) = 1291.99, p < .001$ , TLI = .84; CFI = .87 (Bentler, 1990), RMSEA = .08 (Browne & Cudek, 1993). An examination of modification indices suggested that fit could be improved if several errors were free to correlate, but did not suggest that any specific item be moved to another factor. Specifically, the modification indices indicated that a better fit would be obtained if the errors between Items 15 and 25, Items 21 and 25, and Items 9 and 20 were correlated. When these changes were implemented, fit was improved:  $\Delta\chi^2(3) = 215.04, p < .001$ ;  $\chi^2(262, N = ?) = 1,076.95, p \leq .001$ , TLI = .87, CFI = .89, RMSEA = .07. Table 2 provides the standardized factor loadings. Internal consistency reliabilities using this model for the five scales ranged from barely acceptable (.64, Peer Relationship Problems) to acceptable (.89, Hyperactivity–Inattention). The internal consistency reliabilities of the other scales were as follows: Conduct Problems,  $\alpha = .84$ ; Prosocial Behaviour,  $\alpha = .84$ ; and Emotional Symptoms,  $\alpha = .74$ . Table 4 presents descriptive data for items and scale scores. For all scales, skewness and kurtosis values were less than 1.5, indicating sufficient normality for multivariate tests (Stevens, 2002).

**Gender Differences**—In light of extensive research documenting gender differences in social behaviors (Maccoby, 2002), we next examined whether gender affected the structural relations among SDQ items. We used multigroup comparison analyses within Amos. A chi-square difference test,  $\Delta\chi^2(21) = 33.05, p = .05$ , revealed that the unconstrained model (in which structural paths are allowed to differ for boys and girls) did not provide a significant increment in fit,  $\chi^2(524, N = ?) = 1,405.09, p \leq .001$ , TLI = .86, CFI = .88, RMSEA = .05, over the constrained model (in which structural paths are constrained to be the same for boys and girls),  $\chi^2(545, N = ?) = 1,438.15, p < .001$ , TLI = .86, CFI = .88, RMSEA = .05. Thus, the hypothesized model is an equally good fit for both genders.

**Ethnic Differences**—We also examined whether ethnicity (majority vs. minority) moderated model fit. Owing to the relatively small number of African American children, we

combined African American and Hispanic children into a “minority” group and tested fit invariance between majority and minority children. A chi-square difference test,  $\Delta\chi^2(21) = 31.48, p = .06$ , revealed that the unconstrained model,  $\chi^2(524, N = ?) = 1,400.40, p < .001$ , TLI = .87, CFI = .88, RMSEA = .05, did not provide a significant increment in fit over the constrained model,  $\chi^2(545, N = ?) = 1,431.88, p < .001$ , TLI = .87, CFI = .88, RMSEA = .05. Thus, the hypothesized model is an equally good fit for both majority and minority children.

**Parent Questionnaire**—When tested with CFA in AMOS, the hypothesized five-factor parent model ( $N = 505$ ) converged and all estimates were within bounds. As was true for the teacher questionnaire results, the fit indices for the five-factor model for the parent questionnaire approached but did not meet conventional levels for good model fit,  $\chi^2(265, N = ?) = 739.76, p \leq .001$ , TLI = .80, CFI = .82, RMSEA = .07.

An examination of the modification indices indicated that fit could be improved by correlating the errors between Items 15 and 25 as well as the errors between Items 21 and 25. Furthermore, an examination of regression weight modification indices and implied moments suggested moving the items “Has at least one good friend” and “Generally liked by other children” from the Peer Relationship Problems scale to the Prosocial Behaviours scale. When the alternative model suggested by the modification indices and implied moments was subjected to confirmatory factor analysis, the fit indices were as follows:  $\chi^2(262, N = ?) = 616.66, p \leq .001$ , TLI = .86, CFI = .87, RMSEA = .05. A chi-square difference test,  $\Delta\chi^2(3) = 123.10, p < .001$ , indicated that the revised model provided a better fit for these data. Table 2 provides the standardized factor loadings for this model. Using the modified model, internal consistency reliabilities on the parent questionnaire ranged from .59 (Peer Relationship Problems) to .81 (Hyperactivity–Inattention). The internal consistency reliabilities of the other scales were as follows: Conduct Problems,  $\alpha = .71$ ; Prosocial Behaviour,  $\alpha = .72$ ; and Emotional Symptoms,  $\alpha = .74$ . Table 4 presents descriptive data for items and scale scores. For all scales, skewness and kurtosis were within acceptable limits.

**Parent Educational Level**—Suspecting that parents’ reading comprehension might influence item response and the structural relations among items, we next divided the parents into high educational level (greater than high school education;  $N = 228$ ) and low educational level (equal to or less than a high school education;  $N = 202$ ). A chi-square difference test,  $\Delta\chi^2(24) = 25.22, p = .39$ , revealed that the unconstrained model for the parent questionnaire,  $\chi^2(522, N = ?) = 949.67, p < .001$ , TLI = .84, CFI = .86, RMSEA = .04, did not provide a significant increment in fit over the constrained model for the parent questionnaire,  $\chi^2(546, N = ?) = 974.89, p < .001$ , TLI = .84, CFI = .86, RMSEA = .04, indicating that model fit is invariant across parent educational level.

**Gender Differences**—We next examined whether gender affected the structural relations among SDQ items for the parent data. A chi-square difference test,  $\Delta\chi^2(24) = 50.74, p = .001$ , revealed that the unconstrained model,  $\chi^2(522, N = ?) = 927.20, p < .001$ , TLI = .84, CFI = .86, RMSEA = .04, provided a significant increment in fit over the constrained model,  $\chi^2(546, N = ?) = 977.94, p < .001$ , TLI = .83, CFI = .85, RMSEA = .04. However, none of the critical ratios for differences between structural parameters or path covariances were statistically significant. Next we ran the model separately for boys and girls. Differences in fit indices for the boys,  $\chi^2(261, N = ?) = 500.89, p < .001$ , TLI = .83, CFI = .85, RMSEA = .06, and for the girls,  $\chi^2(261, N = ?) = 426.31, p < .001$ , TLI = .83, CFI = .85, RMSEA = .06, were negligible. None of the relative fit indices differed more than .05 across groups, indicating that the difference in fit across groups is negligible (Little, 1997). Thus, we concluded that the model was an equally acceptable, but not identical, fit for both genders.



**Ethnic Differences**—We also examined whether ethnicity (majority vs. minority) moderated model fit. A chi-square difference test,  $\Delta\chi^2(24) = 103.39, p < .001$ , revealed that the unconstrained model for the parent questionnaire,  $\chi^2(522, N = ?) = 899.17, p < .001$ , TLI = .85, CFI = .87, RMSEA = .04, provided a statistically significant increment in fit over the constrained model for the parent questionnaire,  $\chi^2(546, N = ?) = 1,002.56, p < .001$ , TLI = .82, CFI = .84, RMSEA = .04. An examination of critical ratios for differences between structural parameters and covariances found three covariances that differed across majority and minority groups. In each case, the observed difference was in the strength of the association, not in the direction. Specifically, the covariances between emotional symptoms and conduct problems ( $r_s = .177$  and  $.436$  for majority and minority, respectively; critical ratio =  $-3.353$ ), conduct problems and hyperactivity ( $r_s = .477$  and  $.585$  for majority and minority, respectively; critical ratio =  $3.217$ ), and emotional symptoms and peer problems ( $r_s = .245$  and  $.494$  for majority and minority, respectively; critical ratio =  $2.835$ ) differed between our two groups.

Differences in fit indices for the majority group,  $\chi^2(261, N = ?) = 474.03, p < .001$ , TLI = .86, CFI = .87, RMSEA = .06, and for the minority group,  $\chi^2(261, N = ?) = 425.07, p < .001$ , TLI = .85, CFI = .87, RMSEA = .06, were negligible. Neither of the relative fit indices differed more than .05 across groups, indicating that the difference in fit across groups is negligible (Little, 1997). Thus, we concluded that the model was an equally acceptable, but not identical, fit for both ethnicities.

### Convergent and Discriminant Validity of the SDQ

The criterion-related validity of the SDQ was tested by evaluating the relative contributions of the trait and method factors to parent, teacher, and peer measures of hyperactivity, emotional symptoms, conduct problems, peer problems, and prosocial behavior. In these analyses, the parent scale scores were based on the revised factor structure. First, correlations were obtained among the teacher-rated scale scores, parent-rated scale scores, and the  $z$  scores of the peer ratings. The MTMM matrix provides a framework for a rigorous examination of questions concerning convergent and discriminant validity (Campbell & Fiske, 1959). Trait effects can be disentangled from potential method effects by examining the relationships among different traits across different measures. Thus, the convergence of a trait across different assessment methods strengthens the interpretation of that trait, and the divergence of different traits, assessed with the same or different methods, is used to demonstrate that traits represent separate constructs. A MTMM matrix was produced and is reflected in Table 5. Monotrait–heteromethod (same trait, different informants) correlations were in the low to moderate range (mean  $r = .28$ ), with the highest convergence occurring between teachers and parents and between teachers and peers and the lowest convergence between peers and parents. The heterotrait–monomethod (different traits, same informant) correlations varied across traits but were, on average, somewhat higher than the monotrait–heteromethod correlations (mean  $r = .34$ ). The heterotrait–heteromethod (different traits, different informants) correlations (mean  $r = .20$ ) were lower than both the monotrait–heteromethod and the heterotrait–monomethod correlations. The findings of low to moderate monotrait–heteromethod and low heterotrait–heteromethod correlations support the convergent validity, but not the discriminant validity, of the SDQ.

Although the MTMM approach provides one of the best available tests of validity (Cole, 1987), confirmatory factor analysis is the preferred method for evaluating method and trait variance for several reasons. First, confirmatory factor analysis reproduces the original theoretical formulation of Campbell and Fiske's (1959) MTMM correlation matrix (Lance, Noble, & Scullen, 2002). Furthermore, confirmatory factor analysis can determine whether expected models are consistent with the data (Cole, 1987). Finally, confirmatory factor analysis allows the factor loadings, variances, and relationships among the latent traits to be seen at one

glance, aiding in the determination of the degree of convergent and discriminant validity. Convergent validity is then evidenced through the size of the trait factor loadings. Discriminant validity is evidenced by small factor covariances (Kenny & Kashy, 1992).

### CTCM Model

Of the confirmatory factor analytic models, the CTCM model is preferred over other models because it reproduces the Campbell and Fiske (1959) MTMM correlation matrix most faithfully (Cole, 1987; Lance et al., 2002). Furthermore, this model assumes that only one latent factor underlies each trait and each method. However, research has indicated that this model often fails to converge or gives improper solutions (Eid, Lischetzke, Nussbeck, & Trierweiler, 2003; Lance et al., 2002). Indeed, our CTCM model did not converge and produced five implied covariance matrices that were not positive definite.

### CU Model

The CU model is often recommended when CTMM does not converge. An advantage of the CU model is that it does not assume equal method effects across traits (Kenny & Kashy, 1992). A unique factor for a given measure can be considered the combination of error plus a method factor. A disadvantage of the CU model is that trait variance may be inflated over that of CTCM because of this model's inability to disentangle method variance and error variance (Kenny & Kashy, 1992). In the CU model, method variance is represented as error, and errors for items sharing the same method are correlated. Therefore, the correlated error terms provide an index of method effects. A finding of larger unique covariances among the error terms is interpreted as greater method effects.

The CU model converged, and all estimates were within bounds. The CU model fit the data well,  $\chi^2(50, N = 374) = 117.93, p \leq .001, TLI = .91, CFI = .96, RMSEA = .07$ . The standardized parameter estimates for the CU model are presented in Table 6. The unstandardized parameter estimates are presented in Figure 1. The regression weights for peer reports are fixed at 1.00 to establish the scale of measurement for the latent variables and to estimate the variances of the latent trait factors.

**Convergent Validity**—The factor loadings in Figure 1 and Table 6 are moderate and statistically significant, providing evidence for convergence among the teacher, parent, and peer ratings of the SDQ constructs. Standardized factor loadings are reported in Figure 1 and also in Table 6. However, to compare factor loadings by different raters, we examined the unstandardized factor loadings. For parents, there was a range from .27 to .57 and a mean of .41; for teachers, there was a range from .07 to 1.31 and a mean of .64; and for peers, the unstandardized factor loadings were set a priori to 1.00 and therefore had no variance. These results indicate that parent ratings contain less trait variance than either teacher or peer ratings.

**Method Effects**—Evidence for method effects comes from the unique covariances (i.e., covariances between scales within a single method), all of which were statistically significant but smaller than the factor loadings. To determine the relative method variance of the three raters, we tested three alternative models against the full model. In each model, the unique error covariances for parent, teacher, or peer were unconstrained. The chi-square difference for the model in which the parent covariances were left to vary,  $\Delta\chi^2(10) = 97.01, p < .001$ , was greater than the model for teacher,  $\Delta\chi^2(10) = 34.54, p < .001$ , or peer  $\Delta\chi^2(10) = 82.41, p < .001$ . These results indicate that method effects are strongest for parent ratings.

**Discriminant Validity**—The CU model provides relatively poor evidence of discriminant validity. Evidence of discriminant validity comes from the factor covariances, all of which are

small but statistically significant. This indicates relatively poor factor distinctiveness (i.e., discriminant validity).

## DISCUSSION

This study examined the psychometric properties of the SDQ with an ethnically diverse sample of relatively low-achieving first-grade students. Using confirmatory factor analysis, we demonstrated that the five-factor solution reported in previous exploratory factor analyses with non-U.S. populations provides a marginally adequate fit for both the teacher and the parent versions of the SDQ. For both versions, data fit a five-factor model; however, the factors consisted of different items. On the teacher questionnaire, the five-item by five-factor model as hypothesized by Goodman (1997) did indeed fit our data better than any other model tested, using the modification indices as a guide. However, on the parent questionnaire, fit was improved when items on the Peer Relationship Problems factor and the Prosocial Behaviour factors were moved. Owing to difference in factor structure, differences across parent and teacher scales may reflect differing conceptions of the underlying dimensions, as well as differences in setting and normative expectations. Both versions of the SDQ also appear to have a comparable fit, but a somewhat different fit for minority and majority groups. Finally, the parent version of the SDQ was an equally good fit for parents differing in educational level.

This study provides evidence of convergent and discriminant validity through the use of the MTMM and confirmatory factor analysis methods for both the teacher and the parent versions of the SDQ. Measured convergent validity appears to be stronger than measured discriminant validity. Thus, these results are consistent with results of the MTMM approach in demonstrating the presence of both trait and method variance and better convergent validity than discriminant validity. That is, the SDQ provides valid measures of each construct; however, the measures of different constructs overlap substantially. Problems tend to co-occur (Arcelus & Vostanis, 2005; Nottelmann & Jensen, 1995); thus, it is not surprising that the scales overlap. The implication of the relatively low level of factor distinctiveness is that the SDQ is not appropriate for individual diagnosis.

The teacher–peer convergent and discriminant validities were in the moderate range. Parent ratings showed less convergent and discriminant validity. Peer associations may be constrained by the fact that peer data were collected by means of a different methodology from that used to obtain parent and teacher evaluations. Furthermore, each of the five constructs on the peer measures was assessed with a single item (completed by multiple raters). A single item does not provide coverage of symptoms that are covered in both the parent and the teacher SDQ. Despite this constraint, it appears that teachers and peers shared more agreement than did parents and peers or parents and teachers. This may be due in part to setting differences, as teacher and peers observed children in the same setting. Furthermore, peers may have used information based on their observations of teacher–student interactions to form judgments about classmates’ social competencies (Hughes, Cavell, & Willson, 2001).

An examination of correlations among scales for different raters reveals that teacher and peer ratings of conduct problems, hyperactivity, and peer problems were each strongly and inversely associated with teacher and peer ratings of prosocial behaviors. However, this was not the case for parents. Parents who rated their child as having emotional problems tended to rate their child as having other problem behaviors but not as being less prosocial. Thus, teachers and peers viewed externalizing problems, but not emotional problems, as related to prosocial behavior. Conversely, parents viewed emotional problems as indicative of other problems and did not see emotional or externalizing problems as implicated in prosocial behavior.

The SDQ shows adequate factorial validity and modest convergent validity, but its discriminant validity is relatively poor. The SDQ does not appear to discriminate well among the constructs of emotional symptoms, conduct problems, hyperactivity, peer problems, and prosocial behaviors. The low to moderate discriminant validity underscores the fact that the SDQ is not appropriate for clinical diagnoses or treatment decisions. These findings are consistent with Goodman's (1997) intended use of the SDQ as a screening versus a diagnostic measure. These findings further emphasize the need to obtain data from multiple sources to ensure reliable measures of constructs. Goodman noted that the best strategy for researchers using the SDQ is to choose cutoffs according to the prevalence rates of the disorder in the sample being studied and according to the relative importance for that study of false positives and negatives. He noted that it may be necessary to also adjust cut-offs for screening on the basis of age and gender (Goodman, 1997). The low correspondence between parent and teacher report increases the chance of false positives. To address this issue, Frick, Bodin, and Barry (2000) have suggested scoring the higher item score of the parent or the teacher rating. Further research is needed to establish the validity of the SDQ when scored in this manner—criterion-related validity with such instruments such as the Behavior Assessment System for Children (Reynolds & Kamphaus, 1992) or the Child Behavior Checklist (Achenbach, 1991), as well as the SDQ's test-retest reliability and its utility with children of differing ages and ethnicities.

The SDQ shows that it has the potential to meet the need for a brief, psychometrically sound screening measure of children's behavioral and emotional adjustment for community-based, public health research. This instrument would be of particular value to school psychologists involved in assessing the effectiveness of universal prevention programs in improving mental health and decreasing symptoms of emotional and behavioral problems. However, the SPQ's brevity comes at the expense of its being able to discriminate among different types of problems. The problem of low discriminant validity has been encountered with other screening instruments such as the Acting Out, Moody, Learning Disability (Durlak et al., 1980). With the exception of the Peer Relationship Problems scale on the parent questionnaire, the SDQ scales evidenced acceptable internal consistency reliability.

In summary, the SDQ is promising but needs further development to continue to improve model fit and factor structure. It should not be used as a basis for diagnosis or for selection into different treatment conditions. However, the SDQ could be used as the first step in a multigating, school-wide assessment system (Feil, Walker, Severson, & Ball, 2000) and to collect information on the prevalence of behavioral and emotional adjustment in a school for purposes of assessing the need for prevention programs and for evaluating schoolwide prevention interventions.

Results need to be considered in light of study limitations. First, the relatively small number of African American children did not permit testing fit of the model for African American and Hispanic students separately. Second, the SDQ needs to be tested using a wider age group. Our study only addressed first graders with a mean age of 6 years, 2 months. Furthermore, the sample was restricted to students scoring below the district median score on a state-approved, district-administered measure of first-grade literacy. Thus, results may not generalize to a sample of children representative of the full spectrum of literacy levels. Specifically, the restricted nature of the sample renders it inappropriate for establishing norms. However, our purposes were to determine whether the SDQ evinces equivalent factor structure with ethnically diverse samples and to test its convergent and discriminant validity. Because minority children are more likely to have low literacy skills (U.S. Department of Education, 2005) and because children with relatively low literacy skills are more likely to develop mental health problems (Adams, Snowling, Hennessy, & Kind, 1999), our sample was an appropriate and relevant one for the study's purposes.

## Acknowledgements

This research was supported in part by National Institute of Child Health and Human Development Grant 5 R01 HD39367-02 to Jan N. Hughes.

## References

- Achenbach, TM. Child behavior checklist. Burlington, VT: Author; 1991.
- Adams JW, Snowling MJ, Hennessy SM, Kind P. Problems of behaviour, reading, and arithmetic: Assessments of comorbidity using the Strengths and Difficulties Questionnaire. *British Journal of Educational Psychology* 1999;69:571–585. [PubMed: 10665170]
- Annie E. Casey Foundation. Kids Count 2004 data book online. 2004. Retrieved April 7, 2005, from <http://www.kidscount.org/sld/databook.jsp>
- Arbuckle, JL.; Wothke, WW. Amos 4.0 user's guide. Chicago: SmallWaters; 1999.
- Arcelus J, Vostanis P. Psychiatric comorbidity in children and adolescents. *Current Opinion in Psychiatry* 2005;18:429–434. [PubMed: 16639137]
- Asher SR, Dodge KA. Identifying children who are rejected by their peers. *Developmental Psychology* 1986;22:444–449.
- Barton J, Bain C, Hennekens CH, Rosner B, Belanger C, Roth A, Speizer FE. Characteristics of respondents and non-respondents to a mailed questionnaire. *American Journal of Public Health* 1980;70:823–825. [PubMed: 7416342]
- Bean AG, Roszkowski MJ. The long and short of it. *Marketing Research* 1995;7:20–26.
- Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin* 1990;107:238–246. [PubMed: 2320703]
- Bourdon KH, Goodman R, Rae DS, Simpson G, Koretz DS. The Strengths and Difficulties Questionnaire: U.S. normative data and psychometric properties. *Journal of the American Academy of Child & Adolescent Psychiatry* 2005;44:557–564. [PubMed: 15908838]
- Braymen RK, Piersel WC. The early entrance option: Academic and social/emotional outcomes. *Psychology in the Schools* 1987;24(2):179–189.
- Breznitz Z, Teltsch T. The effect of school entrance age on academic achievement and social-emotional adjustment of children: Follow-up study of fourth graders. *Psychology in the Schools* 1989;26(1): 62–68.
- Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. *Testing structural equation models*. Newbury Park, CA: Sage; 1993. p. 247–261.
- Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* 1959;56:81–105. [PubMed: 13634291]
- Carmines, EG.; Zeller, RA. *Reliability and validity assessment*. Beverly Hills, CA: Sage.; 1979.
- Coie JD, Dodge KA, Coppotelli HA. Dimensions and types of social status: A cross-age perspective. *Developmental Psychology* 1982;20:941–952.
- Cole DA. Utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology* 1987;55:584–594. [PubMed: 3624616]
- Conners, CK. *Conners' Rating Scales—Revised: Technical manual*. North Tonawanda, NY: Multi-Health Systems; 2001.
- Cummings JA, Harrison PL, Dawson MM, Short RJ, Gorin S, Palomares RS. The 2002 conference on the future of school psychology: Implications for consultation, intervention, and prevention services. *Journal of Educational & Psychological Consultation* 2004;15:239–256.
- Dickey WC, Blumberg SJ. Revisiting the factor structure of the Strengths and Difficulties Questionnaire: United States, 2001. *Journal of the American Academy of Child & Adolescent Psychiatry* 2004;43:1159–1167. [PubMed: 15322420]
- Dillman, DA. *Mail and Internet surveys: The total design method*. 2. New York: Wiley; 2000.
- Dillman DA, Sinclair MD, Clark JR. Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupant-addressed census mail surveys. *Public Opinion Quarterly* 1993;57:289–304.

- Doak, CC.; Doak, LG.; Root, JH. Teaching patients with low literacy skills. Philadelphia: J. B. Lippincott; 1996.
- Durlak, JA. School-based prevention programs for children and adolescents. Thousand Oaks, CA: Sage; 1995.
- Durlak JA, Stein MA, Mannarino AP. Behavioral validity of a brief teacher rating scale (the AML) in identifying high-risk acting-out school children. *American Journal of Community Psychology* 1980;8:101–115.
- Edwards P, Roberts I, Clarke M, DiGuseppi C, Pratap S, Wentz R, Kwan I. Increasing response rates to postal questionnaires: Systematic review. *British Medical Journal* 2002;324:1183–1185. [PubMed: 12016181]
- Eid M, Lischetzke T, Nussbeck FW, Trierweiler LI. Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C (M-1) model. *Psychological Methods* 2003;8:38–60. [PubMed: 12741672]
- Evans GW. The environment of childhood poverty. *American Psychologist* 2004;59:77–92. [PubMed: 14992634]
- Fantuzzo J, McWayne C, Bulotsky R. Forging strategic partnerships to advance mental health science and practice for vulnerable children. *School Psychology Review* 2003;32(1):17–37.
- Feil EG, Walker H, Severson H, Ball A. Proactive screening for emotional/behavioral concerns in Head Start preschools: Promising practices and challenges in applied research. *Behavioral Disorders* 2000;26:13–25.
- Fombonne E, Simmons H, Ford T, Meltzer H, Goodman R. Prevalence of pervasive developmental disorders in the British nationwide survey of child mental health. *International Review of Psychiatry* 2003;15:158–165. [PubMed: 12745327]
- Frick PJ, Bodin SD, Barry CT. Psychopathic traits and conduct problems in community and clinic-referred samples of children: Further development of the psychopathy screening device. *Psychological Assessment* 2000;12:382–393. [PubMed: 11147105]
- Friedman, RM.; Katz-Leavy, JW.; Manderscheid, RW.; Sondheimer, DL. Prevalence of serious emotional disturbance in children and adolescents. In: Manderscheid, RW.; Sonnenschein, MA., editors. *Mental health: United States*. Rockville, MD: Substance Abuse & Mental Health Services Administration; 1996. p. 71-89.
- Goodman R. A modified version of the Rutter parent questionnaire including extra items on children's strengths: A research note. *Journal of Child Psychology and Psychiatry* 1994;35:1483–1494. [PubMed: 7868642]
- Goodman R. The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry* 1997;38:581–586. [PubMed: 9255702]
- Goodman R. The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of Child Psychology and Psychiatry* 1999;40:791–799. [PubMed: 10433412]
- Goodman R. Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of Child Psychology and Psychiatry* 2001;40:791–799. [PubMed: 10433412]
- Goodman R, Renfrew D, Mullick M. Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European Child and Adolescent Psychiatry* 2000;9:129–134. [PubMed: 10926063]
- Goodman R, Scott S. Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful? *Journal of Abnormal Child Psychology* 1999;27:17–24. [PubMed: 10197403]
- Greenberg MT, Domitrovic C, Bumbarger B. The prevention of mental disorders in school-aged children: Current state of the field. *Prevention & Treatment* 2001;4(Article 1)
- Hoagwood K, Johnson J. School psychology: A public health framework: I. From evidence-based practices to evidence-based policies. *Journal of School Psychology* 2003;41:3–22.
- Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 1999;6:1–55.

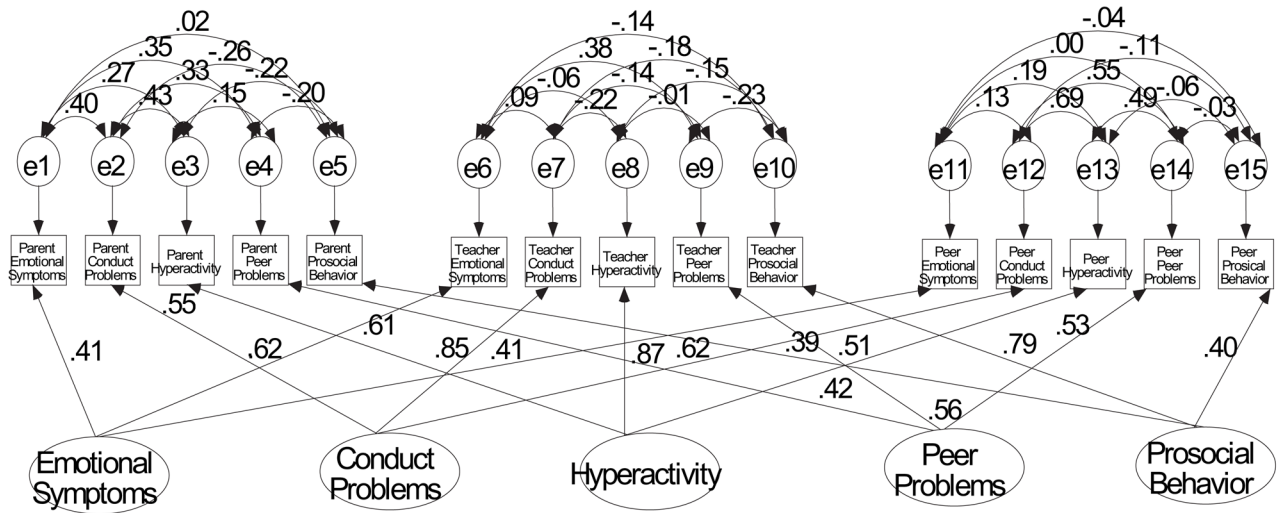
- Hughes, J. Assessment of children's social competence. In: Reynolds, CR.; Kamphaus, R., editors. Handbook of psychological and educational assessment of children. New York: Guilford Press; 1990. p. 423-444.
- Hughes JN, Cavell TA, Willson V. Further evidence of the developmental significance of the teacher-student relationship. *Journal of School Psychology* 2001;39:289-302.
- Kenny DA, Kashy DA. Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin* 1992;112:165-172.
- Lance CE, Noble CL, Scullen SE. A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods* 2002;7:228-244. [PubMed: 12090412]
- Little TD. Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research* 1997;32:53-76.
- Maccoby, EE. Gender and social exchange: A developmental perspective. In: Laursen, B.; Graziano, WG., editors. *Social exchange in development: New directions for child and adolescent development*. San Francisco: Jossey-Bass; 2002. p. 87-105.
- Marsh, HW.; Grayson, D. Latent variable models of multitrait-multimethod data. In: Hoyle, RH., editor. *Structure equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: Sage; 1995. p. 177-198.
- Masten AS, Morison P, Pellegrini DS. A revised class play method of peer assessment. *Developmental Psychology* 1985;21:523-533.
- Nottelmann ED, Jensen PS. Comorbidity of disorders in children and adolescents: Developmental perspectives. *Advances in Clinical Child Psychology* 1995;17:109-155.
- Reynolds, CR.; Kamphaus, RW. *Behavior Assessment System for Children manual*. Circle Pines, MN: American Guidance Service; 1992.
- Rutter M. A children's behavior questionnaire for completion by teachers: Preliminary findings. *Journal of Child Psychology and Psychiatry* 1967;8:1-11. [PubMed: 6033260]
- Rutter, M.; Tizard, J.; Whitmore, K., editors. *Education, health and behavior*. London: Longman; 1970.
- Smedje H, Broman JE, Hetta J, von Knorring AL. Psychometric properties of a Swedish version of the "Strengths and Difficulties Questionnaire". *European Child and Adolescent Psychiatry* 1999;8:63-70. [PubMed: 10435454]
- Srole, L.; Langner, TS.; Michael, ST.; Opler, MK.; Rennie, TAC. *Mental health in the metropolis—The midtown Manhattan study*. 1. New York: McGraw-Hill; 1964.
- Stevens, J. *Applied multivariate statistics for the social sciences*. 4. Mahwah, NJ: Erlbaum; 2002.
- Streiner, DL.; Norman, GR. *Health measurements scales*. Oxford, England: Oxford University Press; 1989.
- Terry, R. Measurement and scaling issues in sociometry: A latent trait approach. Paper presented at the biennial meeting of the Society for Research in Child Development; Albuquerque, NM. 1999 Apr.
- Thabet AA, Stretch D, Vostanis P. Child mental health problems in Arab children: Application of the Strengths and Difficulties Questionnaire. *International Journal of Social Psychology* 2000;46(4): 266-280.
- U.S. Census Bureau. American community survey. Percent of people below poverty level. 2002. Retrieved August 14, 2004, from <http://www.census.gov/acs/www/Products/Ranking/2002/R01T050.htm>
- U.S. Department of Education, National Center for Education Statistics. *Education Digest*. 2006. Retrieved June 2, 2007, from <http://nces.ed.gov/programs/digest/>
- U.S. Public Health Service. Report of the Surgeon General's Conference on Children's Mental Health: A national action agenda. Washington, DC: Author; 2002.
- Weir K, Duveen G. Further development and validation of the prosocial behaviour questionnaire for use by teachers. *Journal of Child Psychology and Psychiatry* 1981;22:357-374. [PubMed: 7287845]

## Biographies

Crystal René Hill is a recent graduate of the School Psychology Program of the Department of Educational Psychology, College of Education and Human Development, Texas A&M University, which is accredited by the American Psychological Association (APA) and the National Association of School Psychologists (NASP). She completed an APA-accredited internship in professional psychology with the Learning Support Center for Child Psychology (LSCCP) at Texas Children's Hospital. The LSCCP is an independent clinical department of Texas Children's Hospital as well as the child psychology academic department at Baylor College of Medicine within the School of Allied Health Sciences. She has served in a national leadership role as the American Psychological Association of Graduate Students liaison to the Board of Educational Affairs of the APA. Her current research interests include the development and assessment of children's social, emotional, and behavioral disorders, prevention science, and the family functioning of children diagnosed with neurodevelopmental and genetic disorders.

Jan N. Hughes is professor of educational psychology at Texas A&M University, where she chairs the Interdisciplinary Research Faculty for Children, Youth, and Families. A distinguished research fellow in the College of Education and Human Development, her research focuses on the development, assessment, and treatment of children's social, emotional, and behavioral disorders and prevention science. She has published more than 100 articles, book chapters, and books. Her research has been supported by the National Institute of Child Health and Human Development, the National Institute on Drug Abuse, and the Department of Education have supported her research. A fellow of the APA and a licensed psychologist, she has served in many national leadership roles including president of the APA Division of School Psychology (Div. 16) and member of the APA Committee on Accreditation. In 1998, she received the Distinguished Service Award from Division 16.





**Figure 1.** Standardized correlated uniquenesses model. Note: Correlations among the Emotional Symptoms, Conduct Problems, Hyperactivity–Inattention, Peer Relationship Problems, and Prosocial Behavior factors were allowed. These correlations can be seen in Table 6.

Table 1

Selected Participant Characteristics

Characteristic	Parent (N = 505)		Teacher (N = 676)		Peer (N = 602)		MTMM sample (N = 374)		Total sample (N = 784)	
	n	%	n	%	n	%	n	%	n	%
Female	236	47	326	48	283	47	179	48	412	53
Male	269	53	350	52	319	53	195	52	372	47
African American	110	22	156	23	127	21	76	20	182	23
Caucasian	197	39	235	35	215	36	144	39	267	34
Hispanic	167	33	248	37	230	38	131	35	292	37
Other	31	6	37	5	30	5	23	6	43	5
Bilingual, limited English proficient, or English as a second language	97	19	155	23	146	24%	75	20	185	24
Economic disadvantage	305	60	424	63	369	61%	227	61	484	62
Single-parent home	94	19	99	15	78	13%	62	17%	437	56
	M	SD	M	SD	M	SD	M	SD	M	SD
Age	6.11	0.65	6.12	0.65	6.08	0.64	6.06	0.64	6.14	0.65
Unit score	93.95	14.22	93.13	14.27	93.50	14.59	94.37	13.82	92.91	14.62
Years of parental education	14.15	3.77	14.20	3.73	14.19	3.83	14.26	3.90	14.15	3.75

Note. MTMM = multitrait-multimethod.

Table 2  
Standardized Factor Loadings on the Predicted Five Factors for Teachers ( $N = 676$ ) and Parents ( $N = 505$ )

Predicted factor and questionnaire item	Standardized factor loadings				
	1	2	3	4	5
<b>Emotional Symptoms</b>					
3. Often complains of headaches, stomachaches, or sickness	.37/.43				
8. Many worries or often seems worried	.70/.57				
13. Often unhappy, downhearted, or tearful	.66/.59				
16. Nervous or clingy in new situations, easily loses confidence	.59/.59				
24. Many fears, easily scared	.68/.62				
<b>Conduct Problems</b>					
5. Often loses temper		.64/.58			
7. <i>Generally well behaved, usually does what adults request</i>		-.80/-.50			
12. Often fights with other children or bullies them		.79/.75			
18. Often lies or cheats		.69/.64			
22. Steals from home, school, or elsewhere		.58/.45			
<b>Hyperactivity–Inattention</b>					
2. Restless, overactive, cannot stay still for long			.89/.74		
10. Constantly fidgeting or squirming			.87/.73		
15. Easily distracted, concentration wanders			.74/.75		
21. <i>Thinks things out before acting</i>			-.66/-.49		
25. <i>Good attention span, sees work through to the end</i>			-.70/-.63		
<b>Peer Relationship Problems</b>					
6. Rather solitary, prefers to play alone				.28/.44	
11. <i>Has at least one good friend</i>				-.70	.34 <sup>a</sup>
14. <i>Generally liked by other children</i>				-.85	.45 <sup>a</sup>
19. Picked on or bullied by other children				.41/.67	
23. Gets along better with adults than with other children				.17/.49	
<b>Prosocial Behaviour</b>					
1. Considerate of other people's feelings					.80/.56
4. Shares readily with other children, for example toys, treats, pencils					.75/.49
9. Helpful if someone is hurt, upset, or feeling ill					.70/.64
17. Kind to younger children					.64/.45
20. Often offers to help others (parents, teacher, other children)					.62/.57

Note. Bold values are the standardized factor loadings for teachers. All others are the standardized factor loadings for parents.

<sup>a</sup>The standardized factor loadings for Items 11 and 14 on the parent questionnaire are reported on the Prosocial Behavior scale.

**Table 3**

Comparison of Means and Standard Errors for Study Sample and Normative Sample

Factor scored according to original SDQ	Normative SDQ data from the United States ( <i>N</i> = 2,779)		Our sample ( <i>N</i> = 505)	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Emotional Symptoms	1.5	.13	2.27	.102
Conduct Problems	1.4	.03	2.03	.079
Hyperactivity–Inattention	3.2	.05	4.40	.105
Peer Relationship Problems	1.3	.03	2.05	.065
Prosocial Behaviour	8.4	.04	7.91	.091
Total Difficulties	7.4	.10	10.74	.265

*Note.* Strengths and Difficulties Questionnaire (SDQ) normative data reported at [www.sdqinfo.com](http://www.sdqinfo.com) on November 13, 2004.

**Table 4**  
Means and Standard Deviations for Items and the Five Factors for Parents ( $N = 505$ ) and Teachers ( $N = 676$ )

Questionnaire item	Parents		Teachers	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1. Considerate of other people's feelings	1.60	.542	1.43	.647
2. Restless, overactive, cannot stay still for long	0.85	.782	0.70	.821
3. Often complains of headaches, stomach-aches, or sickness	0.45	.679	0.39	.648
4. Shares readily with other children, for example toys, treats, pencils	1.46	.602	1.38	.660
5. Often loses temper	0.68	.725	0.34	.625
6. Rather solitary, prefers to play alone	0.37	.624	0.32	.591
7. <i>Generally well behaved, usually does what adults request</i>	1.45	.601	1.47	.689
8. Many worries or often seems worried	0.43	.632	0.44	.615
9. Helpful if someone is hurt, upset, or feeling ill	1.59	.569	1.37	.661
10. Constantly fidgeting or squirming	0.68	.768	0.66	.808
11. <i>Has at least one good friend</i>	1.69	.567	1.54	.682
12. Often fights with other children or bullies them	0.26	.549	0.39	.659
13. Often unhappy, downhearted or tearful	0.33	.619	0.36	.610
14. <i>Generally liked by other children</i>	1.66	.519	1.51	.664
15. Easily distracted, concentration wanders	0.92	.729	0.96	.825
16. Nervous or clingy in new situations, easily loses confidence	0.61	.719	0.51	.688
17. Kind to younger children	1.71	.517	1.51	.617
18. Often lies or cheats	0.41	.608	0.39	.643
19. Picked on or bullied by other children	0.49	.659	0.22	.499
20. Often offers to help others (parents, teacher, other children)	1.56	.566	1.36	.675
21. <i>Thinks things out before acting</i>	0.99	.603	0.97	.694
22. Steals from home, school, or elsewhere	0.13	.418	0.18	.497
23. Gets along better with adults than with other children	0.53	.688	0.33	.567
24. Many fears, easily scared	0.46	.645	0.26	.485
25. <i>Good attention span, sees work through to the end</i>	1.07	.686	1.00	.788
Factor Score According to Original				
Strengths and Difficulties				
Questionnaire				
Emotional Symptoms	2.27	2.26	1.95	2.12
Conduct Problems	2.03	1.96	1.83	2.42
Hyperactivity–Inattention	4.40	2.69	4.35	3.26
Peer Relationship Problems	2.05	1.88	1.83	1.91
Prosocial Behaviour	7.91	1.94	7.05	2.53
Total Difficulties	10.74	6.57	9.96	7.28

Note. Italicized items are reversed scored.



**Table 6**

Standardized Estimated Parameters for the Strengths and Difficulties Questionnaire Data: Correlated Uniqueness Model

Factor	Factor loadings			
	Emotional Symptoms	Conduct Problems	Hyperactivity	Peer Problems
Conduct Problems	.46			
Hyperactivity–Inattention	.62	.88		
Peer Relationship Problems	.60	.93	.81	
Prosocial Behaviour	–.26	–.86	–.76	–.82

*Note.* Values of 0.00 and 1.00 are fixed a priori ( $N = 284$ ). All  $ps < .05$ .