

Social Competence: Development of an Indicator for Children and Adolescents

Stephen J. Blumberg · Adam C. Carle ·
Kathleen S. O'Connor · Kristin Anderson Moore ·
Laura H. Lippman

Accepted: 30 December 2007 / Published online: 23 January 2008
© Springer Science + Business Media B.V. 2007

Abstract We describe a new measure based on eight parent-reported items designed for use in large nationally representative surveys to assess social competence among children and adolescents 6 to 17 years of age. The measure's psychometric characteristics were evaluated using data ($N=67,405$) from the 2003 (U.S.) National Survey of Children's Health (NSCH), a broad cross-sectional random-digit-dial telephone survey sponsored by the U.S. Health Resources and Services Administration's Maternal and Child Health Bureau and conducted by the U.S. Centers for Disease Control and Prevention's National Center for Health Statistics. The examination of the dimensional structure suggested that two correlated factors (social skills, behavior problems) best explained the variability in the responses. The fit of a Rasch-family Graded Response Model with constrained discrimination parameters was confirmed for both factors, and no differential item functioning was noted for age or sex of the child. The NSCH Social Competence Scale is an internally valid and reliable survey measure for identifying and discriminating among children with below average social skills and/or above average frequency of behavior problems.

Keywords Social behavior · Indicators · Item response theory · Rasch models

1 Introduction

Societies change and evolve, and statistical indicators need to adapt. The basic facts about people in societies that have long been recorded in historical genealogies and

S. J. Blumberg (✉) · K. S. O'Connor
Centers for Disease Control and Prevention, National Center for Health Statistics,
3311 Toledo Road, Hyattsville, MD 20782, USA
e-mail: sblumberg@cdc.gov

A. C. Carle
University of North Florida, 4567 Saint Johns Bluff Road South, Jacksonville, FL 32224, USA

K. A. Moore · L. H. Lippman
Child Trends, 4301 Connecticut Avenue NW, Suite 350, Washington, DC 20008, USA

vital statistics registries include birth, marriage, childbearing, and death. As the primary concerns of societies moved “beyond survival” (Ben-Arieh and Wintersberger 1997), richer information about societal trends has also been collected. This information tends to focus on economic, health, and material issues such as unemployment, infant mortality, and poverty. New social indicators are being produced to monitor modern problem behaviors, such as substance abuse, unintended pregnancy, obesity, and dropping out of high school. Often these represent problems or issues that were not viewed as problematic in an earlier era.

Childhood has been recognized as a critical life cycle stage for the prevention of such morbidities (Andrews and Ben-Arieh 1999; Ahorn 1985), and calls have been heard for national and subnational indicators that further describe the development, circumstances, and well-being of children (Moore and Lippman 2005; Zill and Brim 1975). However, as many have noted (Moore and Halle 2001; Moore et al. 2001; Federal Interagency Forum on Child and Family Statistics 2003), available indicators generally focus on problems. They rarely address positive outcomes that might be fostered, but rather focus on problems that need to be reduced or eliminated (Moore et al. 2004). A lack of problems is not the same as flourishing, and positive indicators of social and psychological strengths have been particularly meager in the socioemotional domain (Moore and Lippman 2005).

Effective social and relationship skills are increasingly important predictors of success in school (NICHD Early Child Care Research Network 2006; Malecki and Elliot 2002; Wentzel 1993), successful work (Huston and Ripke 2006; Levy and Murnane 2001), and a fulfilling life in today’s information-based (rather than manufacturing-based) society (Rycken 2003). However, socioemotional skills and behaviors are almost completely missing from current indicator reports about children, such as *America’s Children* (www.childstats.gov) and *Kids Count* (www.aecf.org/kidscount). This dearth of positive outcome measures generally and measures of socioemotional skills specifically motivated the development of a measure of social competence among children that could be effectively implemented as part of large national and subnational surveys.

Although the term *social competence* is widely used in the developmental literature, there exist a variety of published definitions (Rose-Krasnor 1997). We define social competence as a set of skills and behaviors necessary to get along with others and be well-liked. Our definition parallels research definitions used by other researchers who similarly center social success upon effective interaction (e.g., Anderson and Messick 1974; Cavell 1990; Dodge 1985; Gresham 1986; Rubin and Rose-Krasnor 1992). This set of skills and behaviors relies on certain cognitive abilities, such as perceptiveness of other’s feelings and perspectives, knowledge about social rules and social life, and insight in social situations. But these skills and behaviors are manifested as the general ability to interact successfully with other people, to be generous and thoughtful, and to use accepted social techniques when attempting to influence other’s behaviors.

Social competence is also indicated by the absence of problem behaviors, such as being highly argumentative, easily distracted, disobedient, or melancholic. However, Raver and Zigler (1997) argue that behaviors that interfere with social interactions (problem behaviors) should not substitute for measures of social skills but rather should complement measures of social skills. We agree that a measure of social

competence should ideally include measures of both positive and negative behaviors. To reiterate a point we made earlier, a lack of problems is not necessarily an indicator of flourishing. Thus, the absence of problem behaviors alone would not be considered social competence unless accompanied by the presence of social skills, and conversely, a measure of social competence that only assesses the presence of social skills, but does not measure the existence of problem behaviors, would be of limited scope and utility.

Below we describe several previous efforts to assess social competence. In subsequent sections, we describe a new measure that was designed for use in the 2003 National Survey of Children's Health, a large-scale nationally representative telephone survey of parents in the United States. We explore the substantive and psychometric characteristics of this brief, behavior-oriented measure and consider its appropriateness as a social indicator.

1.1 Previous Measures of Social Competence

Numerous measures of social competence have been employed in behavior assessments that are commonly used in school settings. In a study by Hosp et al. (2003), 14 commonly used behavior scales for assessing social skills and emotional functioning were rated to ascertain whether the scales were helpful when implementing or evaluating interventions. These scales included the: Walker-McConnell Scale of Social Competence and School Adjustment (Walker and McConnell 1995), Social Skills Rating System parent and teacher forms (Gresham and Elliott 1990), Child Behavior Checklist (Achenbach 1991) and Teacher's Report Form (Achenbach 2001), Revised Behavior Problem Checklist (Quay and Peterson 1987), Behavior Assessment System for Children (BASC) Monitor (Kamphaus and Reynolds 1998), BASC parent and teacher forms (Reynolds and Kamphaus 1992), Behavior Rating Profile parent and teacher forms (Brown and Hammill 1990), Burks' Behavior Rating Scales (Burks 1977), Social-Emotional Dimension Scale (Hutton and Roberts 1986), and the Behavior and Emotional Rating Scale (Epstein and Sharma 1998). As part of this review, the items in each scale were categorized as positive or negative actions or lack of actions. Items assessing behaviors of inactivity (e.g., "does not interrupt the classroom") were considered faulty because they could be successfully accomplished regardless of social skills (e.g., a quietly sleeping child might successfully avoid interrupting a classroom). The authors found that only 4 of the 14 scales used positive action questions for more than two-thirds of their items. Six of the scales examined did not include any positive action items. Each of these scales was also too lengthy to administer in large-scale telephone surveys: The average number of items for these scales was 77 items, ranging from the 30-item Behavior Rating Profile to the 148-item BASC teacher form.

A briefer assessment of social competence that relies predominately on positive actions is the 17-item "prosocial orientation" subscale of the Social Competence Inventory (SCI; Rydell et al. 1997). This subscale assesses a child's positive social interactions (e.g., "has the capacity for generosity" and "good at preventing conflicts"). It specifically excludes products of social functioning (e.g., "is admired by other children") and indicators of active social participation (e.g., "leads play activities" and "makes contact easily with unfamiliar children") that reveal initiative

in social situations but do not identify specific skills or behaviors. However, the documented validity of scores based on this instrument is limited to 8- and 9-year-old children and to paper-and-pencil administration (Rydell et al. 1997), and 17 items may still be too long for large-scale telephone surveys.

To our knowledge, only two scales about social behaviors and including positively worded items have been previously used in representative surveys in the United States and have documented psychometric properties: The Positive Behavior Scale was fielded in 1997 as part of the Panel Study of Income Dynamics (PSID), and the Strengths and Difficulties Questionnaire (SDQ) was included in the 2001, 2003, and 2004 National Health Interview Surveys (NHIS). Selected items assessing social competence have been included in other surveys—for example, the National Longitudinal Survey of Youth (NLSY), the National Survey of Youth and Religion, the California Healthy Kids Survey—but psychometric analyses have not been published for these items.

The Positive Behavior Scale, a 25-item scale developed for use with young children from low-income families (Quint et al. 1997) and validated primarily with samples of such children (Epps et al. 2005), was abbreviated to 10 items for use in the PSID. Six of those items comprise a social competence subscale. Data from the PSID revealed that the subscale has strong internal consistency and scores are related to teacher's ratings of social skills for children 1–12 years of age, but such results are the only published psychometric properties from a representative sample of U.S. children (Epps et al. 2005).

The SDQ is a 25-item measure that has been validated as a clinical assessment and screening tool for psychopathology in children 5–15 years of age, and representative data from a British mental health survey suggest that the SDQ has strong psychometric properties (Goodman 2001). Five of the items, however, are not used when generating total difficulties scores from the SDQ; these five items are positively worded (“strengths”) and assess behaviors intended to benefit others (e.g., “kind to younger children,” “often offers to help others”, “shares readily”). These prosocial behaviors may reveal social competence, but these somewhat altruistic items are not explicitly assessing social competence (Eisenberg and Fabes 1998). In addition, studies have suggested that the parents' responses to items in the prosocial behavior scale may reflect their general willingness to attribute positive qualities to their children rather than their observations of behavior (Dickey and Blumberg 2004; Goodman 1994).

In short, existing measures of social competence among children are too lengthy for administration as part of large broad national surveys, focus on problems rather than actions with positive consequences, focus too narrowly on prosocial behaviors, or their psychometric properties have not been sufficiently examined with nationally representative samples of children of varying ages. A new measure of social competence was needed.

1.2 National Survey of Children's Health Social Competence Scale

The (U.S.) National Survey of Children's Health (NSCH) is a broad comprehensive cross-sectional survey that examines a variety of physical, emotional, and behavioral health and well-being indicators for children and measures of children's experiences with the health care system. This random-digit-dial telephone survey is sponsored by

the U.S. Health Resources and Services Administration, Maternal and Child Health Bureau, and is conducted using the State and Local Area Integrated Telephone Survey (SLAITS) mechanism by the U.S. Centers for Disease Control and Prevention, National Center for Health Statistics. Estimates based on the survey generalize to the noninstitutionalized population of children in each state and across the United States. In 2003, the parents or guardians of more than 100,000 randomly selected children less than 18 years of age responded to the survey. It is being repeated again with a slightly smaller sample in 2007 and 2008.

The specific content of the survey was recommended by a panel of experts consisting of state and federal maternal and child health program directors, representatives of family advocacy organizations, child and health services researchers, and survey design experts (Blumberg et al. 2005). Their decisions, based on epidemiological and policy importance, included the recommendation to create a brief scale on social competence for children 6–17 years of age. Scores derived from the desired scale were to be valid and reliable for both boys and girls and for both younger children and adolescents. A subset of the panel was assigned the responsibility of developing the specific items; two of this manuscript's authors (SJB, KAM) were members of that technical panel.

The technical panel initially reviewed the Positive Behavior Scale and the Behavior Problems Index to identify candidate items. The Behavior Problems Index is a 28-item rating scale based on the Child Behavior Checklist (Achenbach and Edelbrock 1981) and the Rutter questionnaires (Rutter 1967); this index was used in the 1981 and 1988 NHIS and the 1986 and 1988 NLSY. Based on face validity and exploratory factor analyses of PSID and NHIS data, 6 items from the Positive Behavior Scale and 11 items from the Behavior Problems Index were selected for further study. An additional 6 items identifying positive behaviors were developed by two of this manuscript's authors (KAM, LHL). These 23 items were included in a pretest of the NSCH questionnaire. Exploratory factor analyses with data from 661 parents were used to identify items that best represented each of the two constructs (social skills, behavior problems) within each of two age groups (6–11 years, 12–17 years). Four items were selected by the technical panel for each construct based on three criteria: (1) face validity, (2) policy relevance, and (3) strength of relationship with the relevant factor/construct for both age groups.

These 8 items were included in the 2003 NSCH and are listed in the [Appendix](#). All 4 problem behavior items were from the Behavior Problems Index. One of the social skills items was from the Positive Behavior Scale (“gets along well with other children”). These items comprise a scale named the NSCH Social Competence Scale. The current study uses data from the 2003 NSCH to examine the psychometric properties of this scale.

2 Method

2.1 Sample

As noted earlier, the NSCH is a random-digit-dial telephone survey of a representative sample of children less than 18 years of age in each state. From January 29,

2003 to July 1, 2004, nearly two million telephone lines were called to identify households with children. When such households were identified, one child was randomly selected to be the subject of a 29-minute interview that focused on demographics, health and functional status, health insurance coverage, health care access and utilization, health care quality, family functioning, parental health, and neighborhood characteristics. The NSCH Social Competence Scale was included approximately three-quarters of the way through the interview. The interview was conducted with the parent or guardian living in the household who reportedly knew the most about the health and health care of the child. Additional details about the NSCH data collection are available elsewhere (Blumberg et al. 2005).

Interviews were completed for 69,031 children aged 6–17 years; state-specific sample sizes ranged from approximately 900 to 1,500 children. 87% of the interviews were completed in 2003. The child's mother was usually the respondent (78.6%); fathers (17.3%) and grandparents (2.8%) comprise most of the remaining respondents; for simplicity, all will be referred to as "parents" in this manuscript. The overall weighted response rate, based on American Association for Public Opinion Research guidelines, was 55.3%.

2.2 Other Measures

Evidence for construct validity can be obtained by examining the relationships between scores derived from the new scale and other variables that are believed to be related to the construct purportedly measured by the scale. Other measures administered as part of the NSCH were used to examine the construct validity of scores derived from the NSCH Social Competence Scale. For example, parents were asked whether their child currently had "any kind of emotional, developmental, or behavioral problem for which he/she needs treatment or counseling," and whether this problem "has lasted or is expected to last 12 months or longer." They were asked whether a doctor had ever told them that their child had "attention deficit disorder or attention deficit hyperactive disorder" or "behavioral or conduct problems." More generally, parents were asked if they think their child "has difficulties with one or more of the following areas: emotions, concentration, behavior, or being able to get along with other people." Affirmative answers were followed with questions asking if these difficulties would be described as "minor, moderate, or severe" and whether the child's "mental or emotional health puts a burden on your family a great deal, a medium amount, a little, or not at all." The NSCH also ascertained the number of times that the child's "school contacted you or another adult in your household about any problems he/she is having with school," the proportion of the child's friends that the parent has met, and whether the child was "on a sports team or took sports lessons," "participated in any clubs or organizations after school or on weekends," and/or was "involved in any type of community service or volunteer work at school, church, or in the community." Parents were asked to rate the closeness of their relationship with the child ("very close, somewhat close, not very close, or not close at all"). Finally, aggravation in parenting was determined using three questions about the extent to which, in the past month, the parent felt the child was "much harder to care for than most children his/her age," the child did "things that really bother you a lot," and the parent "felt angry with him/her." A

response of “usually” or “always” to any one of these questions was considered an indicator of high aggravation.

2.3 Statistical Analysis

Analyses in this report are limited to the 67,405 children with nonmissing responses for the item set. Factor analysis and Item Response Theory (IRT) methods were used to explore the psychometric properties of the scale. First, given that previous research had not established the dimensionality for the item set, analyses explored the dimensional structure and fit of factor analytic models in a sample of 5,000 children randomly selected from the full sample. Second, based on the exploratory findings, the fit of a Rasch family Graded Response Model with constrained discrimination parameters was validated with the same sample of 5,000 children and then cross-validated in the full sample. Third, after establishing the fit of the model in the general sample, analyses explored the possibility of differential item functioning (DIF) across age and sex. Fourth, using the final IRT measurement model, scale scores were estimated for all children in the sample and then compared to raw scale scores (derived by summing the integer responses to the scale items: *never* = 1, *sometimes* = 2, *usually* = 3, and *always* = 4). Finally, logistic regression analyses with the raw scores were used to evaluate the construct validity of the measure.

The NSCH uses a complex sampling design with stratification of the sample at the state level and unequal probabilities of selection for each child (e.g., because of the random selection of children within households). To account for the complex sampling design, Mplus (Muthén and Muthén 2006) was used for all factor analytic and IRT analyses, and SUDAAN (Research Triangle Institute 2004) was used for analyses of construct validity. Fit assessment focused on close-fit indices and followed guidelines suggested by Hu and Bentler (1999), Muthén and Muthén (2006), Steiger (1998), and Cheung and Rensvold (2002). Misfit was considered when the majority of indices suggested poor fit. In these analyses, the weighted least square means and variance adjusted (WLMSMV) estimator was used. The WLMSMV is a robust estimator appropriate for categorical data (Muthén and Muthén 2006) and is a refinement of the weighted least square estimator used in the analysis of ordinal data. Consistent with arguments for more stringent error control in model tests of this type (Green and Babyak 1997; Thissen et al. 1993), an alpha level of 0.01 was adopted for significance testing.

3 Results

3.1 Dimensionality

To establish dimensionality, confirmatory factor analysis for ordered-categorical measures (CFA-OCM), an extension of the more familiar CFA for continuous measures, was used. CFA-OCM indicates a set of equations to describe the relations among a set of ordered-categorical items, suggesting that individuals' item responses

are determined by their value on an underlying factor or factors and by several measurement parameters (e.g., intercepts, loadings, thresholds, and uniquenesses). The parameters from the CFA-OCM model can then be transformed into IRT parameters (Millsap and Yun-Tein 2004; Takane and de Leeuw 1987).

Several models were tested before arriving at the final IRT model, including a single factor model and various two factor models that allowed different items to load on different factors and that allowed the two factors to be correlated. Space constraints limit the presentation here of results for each model. The one-factor model was rejected when the fit indices failed to meet established criteria for good fit (i.e., the root mean square error of approximation [RMSEA] was greater than 0.05, and the standardized root mean square residual [SRMR] was greater than 0.08). Testing continued with various two-factor models that were consistent with the scale development strategy of identifying (1) skills associated with social competence and (2) related behavior problems. Ultimately, analyses examined the fit of a factor model constrained to allow each of the four items that loaded most highly on each factor to load only on that factor. This model allowed correlation between the two factors (i.e., the correlation was not constrained to zero). As might be expected, the four positively worded items identifying social skills were the four items loading most highly on one factor, and the four negatively worded items identifying related behavior problems were the four items loading most highly on the other factor. Fit indices for this model were acceptable and indicated good fit. The correlation between the factors was moderately strong ($r=-0.68$).

3.2 Rasch Modified Graded Response Model

The four items for each factor were treated as separate scales, and the fit of a Rasch family modified Graded Response Model (Samejima 1969) was examined. The alteration constrained the item discrimination parameters to equality for each of the four-item sets and allowed the model to inherit familiar Rasch model properties (Embretson and Reise 2000). Specifically, the item discrimination parameters for the social skills scale were constrained to equality, and the item discrimination parameters for the behavior problems scale were constrained to equality, but the parameters were not constrained to equality across the scales. No other additional constraints were added to the model. This model continued to provide good fit (RMSEA=0.07, comparative fit index [CFI]=0.93, Tucker–Lewis Index [TLI]=0.96, gamma hat=0.97, and McDonald’s Non-Centrality Index=0.94). In the full sample, fit indices demonstrated increased fit (RMSEA=0.04, CFI=0.92, TLI=0.95, gamma hat=0.99, and McDonald’s Non-Centrality Index=0.98). The modified two-factor Rasch graded response model (GRM) that allowed the factors to correlate was accepted as the general IRT measurement model.

3.3 Differential Item Functioning

Following establishment of the general IRT model, analyses examined the possibility of differential item functioning (DIF) across age and across sex in the full sample. DIF is present if two groups of children differ in the response to a scale item when, in fact, the groups have equivalent levels of the underlying trait or if the two groups

have the same response to a scale item when, in fact, the groups have different levels of the underlying trait. DIF may be revealed if the measurement parameters of the IRT model are different for each group. To examine equivalence of measurement parameters across age and across sex, a series of hierarchically nested models was tested (Millsap and Yun-Tein 2004). The method started with the least restricted measurement model across groups, and then cross-group equivalence constraints in the measurement parameters were added in a stepwise fashion in later models. The tenability of the equivalence constraints in a given set of measurement parameters was evaluated at each step. When fit indices suggested the constraints were not tenable (Cheung and Rensvold 2002), bias was considered present.

To test for DIF across age groups (defined as 6–11 years and 12–17 years), the configural invariance of the Rasch modified GRM was tested first. The loadings were constrained to equality across the groups, the third threshold was constrained to equality across the groups for the four positively worded items identifying social skills, the first threshold was constrained to equality across the groups for the four negatively worded items identifying related behavior problems, the uniquenesses were constrained to a value of one in each group, the factor variance in the younger group was constrained to one, and the factor means were fixed at zero for the younger group. This model fit the data well (RMSEA=0.04, CFI=0.92, TLI=0.96, γ hat=0.99, and McDonald's Non-Centrality Index=0.98). Subsequent analyses examined the equivalence of the thresholds across the age groups. This model retained the previously described constraints and forced the thresholds to equality across the age groups, resulting in a fully invariant measurement model. Although the χ^2 difference test was significant ($\Delta\chi^2=406.83$, $p<0.01$), this test is sensitive to small deviations of fit in large samples (Cheung and Rensvold 2002). Changes in the close fit indices (Cheung and Rensvold 2002) did not suggest that the constraints were untenable (RMSEA=0.04, CFI=0.91, TLI=0.96, γ hat=0.99, $\Delta\gamma$ hat=-0.001, McDonald's Non-Centrality Index=0.98, and Δ McDonald's Non-Centrality Index=0.00). Therefore, the hypothesis of a fully invariant measurement model was not rejected and analyses examined the possibility of DIF across sex.

The method for examining DIF across sex groups was the same as the method for examining DIF across age groups. The model assuming configural invariance of the Rasch modified GRM fit the data well (RMSEA=0.04, CFI=0.92, TLI=0.96, γ hat=0.99, and McDonald's Non-Centrality Index=0.98). When testing the equivalence of the thresholds across the groups, the χ^2 difference test was significant ($\Delta\chi^2=185.48$, $p<0.01$), but an examination of the close fit indices (Cheung and Rensvold 2002) did not suggest that the constraints were untenable (RMSEA=0.04, CFI=0.92, TLI=0.97, γ hat=0.99, $\Delta\gamma$ hat=0.000, McDonald's Non-Centrality Index=0.98, and Δ McDonald's Non-Centrality Index=0.00). Again, the hypothesis of a fully invariant measurement model was not rejected.

Although the hypothesis of measurement invariance was not rejected, some small deviations of fit were observed (as indicated by significant χ^2 difference tests), and these small deviations may affect scale score estimates. To examine the possibility that these small deviations affected scores resulting from the scales, analyses compared the group scale score means resulting from a fully invariant model (i.e., a model that assumes measurement bias does not exist) to those resulting from a

model allowing partial measurement invariance (i.e., a model that includes DIF). If the pattern, size, and significance of mean differences did not differ from the fully invariant and partially invariant models, this would suggest that observed DIF would not impact estimated scores. Under the model incorporating bias, older children demonstrated significantly lower mean levels of the factor assessing behavior problems than did younger children ($M_{\text{dif}}=-0.12, p<0.01$), while no significant age-related difference was observed between mean levels of the factor assessing social skills ($M_{\text{dif}}=0.00$). Boys demonstrated significantly greater mean levels on the behavior problems factor than did girls ($M_{\text{dif}}=0.30, p<0.01$), while no significant sex-related difference was observed for the social skills factor ($M_{\text{dif}}=0.02$). The same pattern, size, and significance of differences were observed under the model ignoring bias, and the hypothesis that even the minimally observed DIF impacted scale scores was rejected.

3.4 Scale Score Properties

The Rasch modified GRM, with the four positively worded items identifying social skills loading exclusively on one factor and the four negatively worded items identifying related behavior problems loading exclusively on another factor, was accepted as the final model. Table 1 provides the item discrimination parameters and thresholds for this model.

In polytomous IRT models, the interpretation of an item's ability to discriminate among persons with different levels of the underlying trait depends on both the discrimination parameter and the spread of the thresholds. For ease of interpretation, we present probability curves for the items identifying social skills in Fig. 1 and probability curves for the items identifying behavior problems in Fig. 2. These curves show the probability of endorsing a given answer category, given the child's level of the underlying trait (as derived from the scale items and the measurement model) using standard scores ("person location"). For example, in Fig. 2, if a child's level of behavior problems is 1.5 standard deviation lower than average, then there is

Table 1 Item discrimination parameters and thresholds

Scale and item	Discrimination parameter	Thresholds		
		Between never and sometimes	Between sometimes and usually	Between usually and always
Social skills scale				
Shows respect	0.93	-3.00	-1.92	-0.92
Gets along well with other children	0.93	-3.60	-1.76	-0.37
Tries to understand other's feelings	0.93	-2.73	-0.89	0.18
Tries to resolve conflicts	0.93	-2.23	-0.61	0.48
Behavior problems scale				
Argues too much	0.98	-1.14	1.22	1.74
Bullies	0.98	0.99	2.92	3.29
Disobedient	0.98	-0.36	2.35	2.81
Stubborn, sullen, or irritable	0.98	-0.98	1.87	2.39

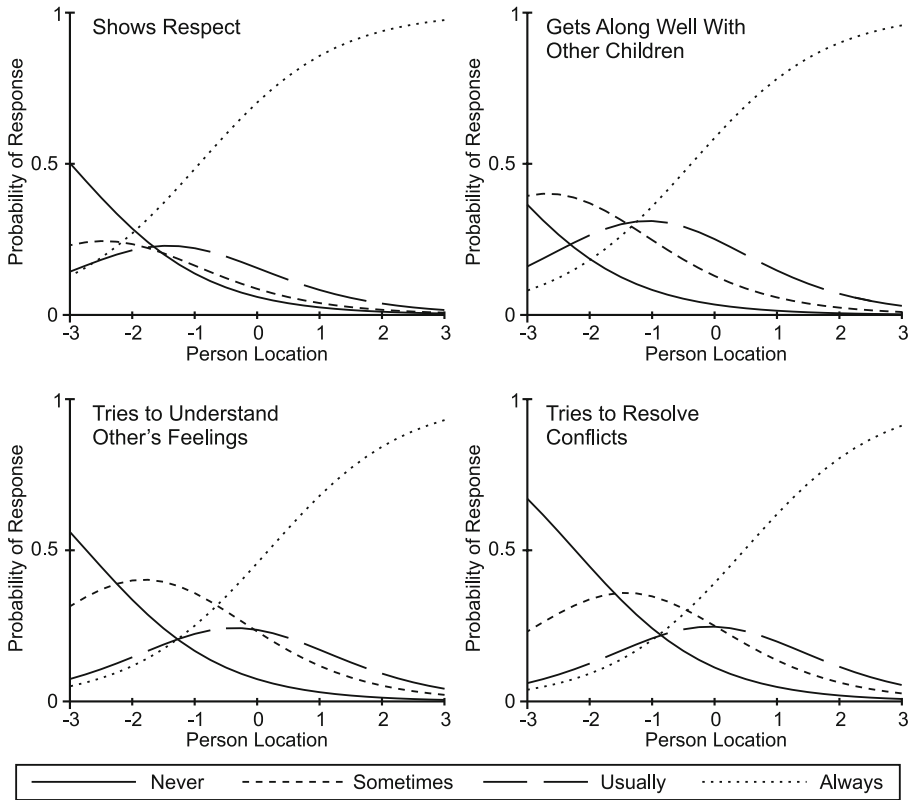


Fig. 1 Answer category probability curves for individual items on the social skills scale

a 58.7% probability that the parent will report that the child “never” argues too much, a 34.8% probability that the parent will report that the child “sometimes” argues too much, a 2.5% probability that the parent will report that the child “usually” argues too much, and a 4.0% probability that the parent will report that the child “always” argues too much. “Always” becomes the most probable response when the child’s level of behavior problems exceeds 1.3 standard deviations above the mean.

Figures 1 and 2 also demonstrate the levels of the underlying trait at which the items function best. When the most probable response category varies across a narrow range of underlying trait scores, the item does a good job of distinguishing among children at various levels within that range. When the most probable response category remains the same across a wide range of underlying trait scores, the item does not discriminate among children at various levels within that range. An examination of Figs. 1 and 2 reveals that the items do a good job identifying poorly performing children (i.e., those above the mean on the behavior problems scale and those below the mean on the social skills scale). The items do not do as good a job discriminating among children who are doing well.

IRT-based scale scores were estimated for the full sample and are available from the authors (SJB, KSO). One distinctive property of Rasch models, however, is that raw scale scores—derived by summing the integer responses to the scale items—are

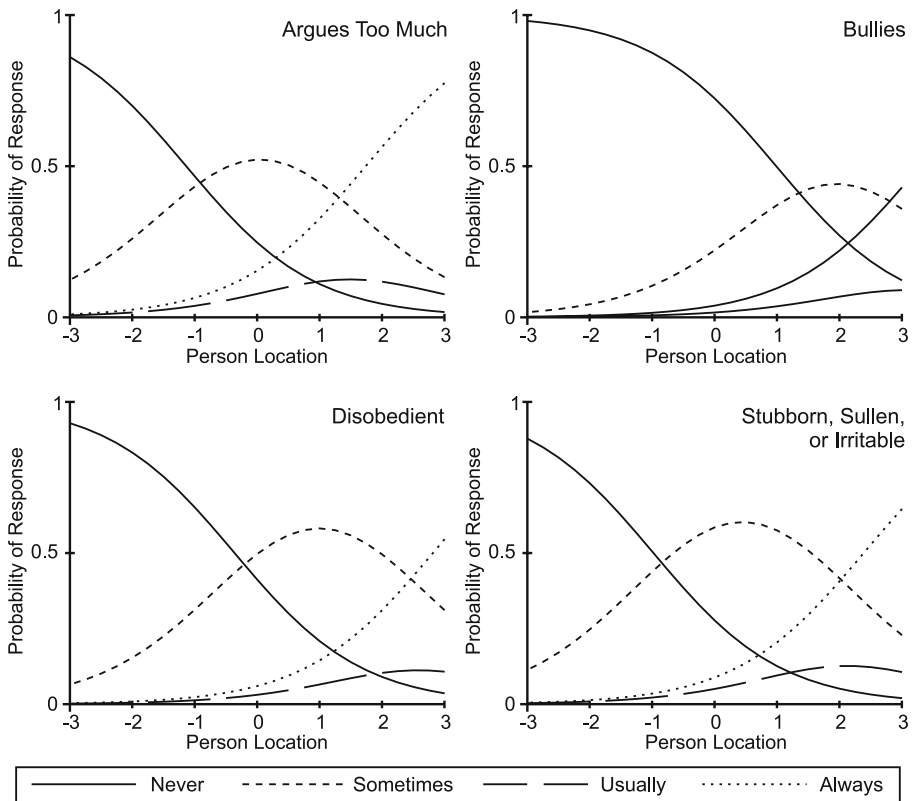


Fig. 2 Answer category probability curves for individual items on the behavior problems scale

Table 2 Descriptive statistics for raw scale scores

Scale and statistic	Score	Standard error
Social skills scale		
Mean	13.32	0.016
10th percentile	9.82	0.036
25th percentile	11.40	0.018
50th percentile (median)	13.06	0.017
75th percentile	14.80	0.037
90th percentile	^a —	—
Geometric mean	13.11	0.018
Behavior problems scale		
Mean	6.92	0.014
10th percentile	4.08	0.016
25th percentile	5.14	0.015
50th percentile (median)	6.29	0.012
75th percentile	7.38	0.020
90th percentile	8.87	0.032
Geometric mean	6.67	0.013

^a Not meaningful. The maximum score of 16 was achieved for 23.39% of the children.

sufficient for capturing all of the information provided by the IRT-based scale scores (Embretson and Reise 2000). Indeed, raw scale scores were nearly perfectly correlated with the IRT-based scale scores (Pearson's $r=0.955$ for social skills, 0.952 for behavior problems).

Descriptive statistics for the raw scale scores are provided in Table 2. Based on raw scores, Cronbach's alpha was 0.65 for the social skills scale and 0.67 for the behavior problems scale.

Table 3 Unstandardized beta coefficients and effect sizes for logistic regression analyses using the social skills scale and the behavior problems scale to predict selected outcomes related to behavior problems and social skills

Selected outcome	Social skills scale				Behavior problems scale			
	Unadjusted		Adjusted ^a		Unadjusted		Adjusted	
	β	PV ^b	β	ΔPV^c	β	PV	β	ΔPV
Outcomes related to behavior problems								
Doctor or other health professional told parent that child has behavioral or conduct problems	-0.52	0.08	-0.35	0.03	0.56	0.08	0.40	0.03
Child's mental or emotional health puts a burden on the family	-0.47	0.06	-0.31	0.02	0.53	0.07	0.39	0.03
Child has a chronic emotional, developmental, or behavioral problem for which treatment or counseling is needed	-0.42	0.06	-0.29	0.02	0.45	0.06	0.31	0.02
Child has moderate or severe difficulties with emotions, concentration, behavior, or getting along with others	-0.41	0.08	-0.26	0.02	0.50	0.09	0.38	0.04
Parent is highly aggravated	-0.34	0.06	-0.18	0.01	0.50	0.10	0.41	0.05
School has contacted parent 2 or more times in past year because of problems in school	-0.33	0.08	-0.24	0.03	0.35	0.07	0.24	0.02
Doctor or other health professional told parent that child had ADD/ADHD	-0.29	0.04	-0.19	0.01	0.33	0.04	0.24	0.02
Parent is not coping well with the day-to-day demands of parenthood	-0.29	0.01	-0.11	0.00	0.38	0.01	0.32	0.01
Outcomes related to social skills								
Child was on a sports team or took sports lessons	0.08	0.01	0.06	0.00	-0.07	0.00	-0.03	0.00
Child participated in extracurricular clubs and organizations	0.09	0.01	0.09	0.01	-0.05	0.00	*-0.01	0.00
Child 12-17 years was involved in community service or volunteer work	0.15	0.03	0.15	0.02	-0.10	0.01	*-0.02	0.00
Child has friends and parent has met all or most of them	0.19	0.03	0.17	0.02	-0.13	0.01	-0.04	0.00
Parent described relationship with child as "very close"	0.33	0.07	0.26	0.03	-0.31	0.05	-0.18	0.01

All unstandardized beta coefficients in this table are significantly different from zero, $p<.01$, with two exceptions marked by an asterisk.

^a Adjusted models included both raw scale scores in the model.

^b PV Percentage of variance in selected indicator explained by the raw scale score, reported as a decimal (Cox and Snell 1989). PV reported as 0.00 indicates a quantity greater than zero but less than 0.005.

^c ΔPV Improvement in percentage of variance explained by the addition of the raw scale score to a logistic regression model that already included the other raw scale score

3.5 Construct Validity

To evaluate construct validity, the raw scores were compared with several other measures that were also assessed in the NSCH interview. We used logistic regression analyses to examine the predictive power of the individual scores and to control for the shared influence of the two scales. All of the measures suggesting greater likelihood of behavior problems were modestly related to both the social skills scale score and the behavior problems scale score when each was included independently in the regression models (see the unadjusted models in Table 3). As expected, the beta coefficients for the behavior problems scale scores were positive, and the beta coefficients for the social skills scale scores were negative. The relationships were stronger with direct indicators of behavior problems (i.e., diagnoses of behavioral or conduct problems and parents' ratings of the severity of the child's mental health problems) than with more indirect indicators (i.e., school contacts resulting from problems and parents' reports of feeling aggravated and not coping well with the demands of parenthood).

When both scale scores were included in the regression models, it was clear that the behavior problems scale score was generally a stronger predictor of alternative measures suggesting behavior problems than was the social skills scale score (see the adjusted models in Table 3). With the behavior problems scale score already in the model, the effect size of the social skills scale score was never greater than 0.03, an effect considered small (Cohen 1988).

In contrast, the social skills scale score was generally a stronger predictor of alternative measures suggesting greater likelihood of positive outcomes than was the behavior problems scale score. With the social skills scale score already in the model predicting positive outcomes, the effect size of the behavior problems scale score was quite low (0.01 or less). However, for both scale scores, the strength of the relationship with measures suggesting positive outcomes was generally smaller than the strength of the relationship with measures suggesting behavior problems.

4 Discussion

The current study established several findings. First, results demonstrated that nationally representative data from the eight items selected for inclusion on the 2003 NSCH were best modeled by a congeneric two factor solution: One construct was represented by four behaviors related to social skills, and the other construct was represented by four behaviors related to behavioral problems. Second, treating each four item set as a separate scale, a modified GRM that constrained the discrimination to equality in a Rasch fashion best fit the data. Third, DIF analyses noted some differential item functioning across age and sex; however, sensitivity analyses suggested that these differences were small and did not meaningfully impact scores generated from these items. Fourth, given the establishment of the general Rasch-based IRT model across age and sex, findings support the use of raw scale scores to represent the level of each child on the two underlying traits (conceptualized as social skills and behavior problems).

The total raw scale scores are sufficient for identifying the underlying level of the measured traits for a specific child because the four scale items are equally related to the specific trait they were intended to measure (Embretson and Reise 2000). This equality was confirmed by the good fit of an IRT model that constrained the discrimination parameters to equality. Children with the same raw score derived from the four scale items are estimated to have equivalent levels of the underlying trait, assuming that all four scale items are answered and regardless of specific answers provided for specific items. Moreover, the relative ordering of children will be similar regardless of whether raw scores or IRT-model-based scores are used. However, scores based on the IRT model are optimal scalings of the underlying trait levels and have stronger psychometric properties, especially for scores at the ends of the score distribution (Embretson and Reise 2000).

The finding that the items are equally related to the underlying trait should not be misconstrued to mean that a specific answer category (e.g., “always”) is equally likely to be used for each item in the subscale. In fact, one advantage of using an IRT model is that an examination of the thresholds (Table 1) can reveal which items are more difficult to answer (that is, least likely to have a “sometimes,” “usually,” or “always” response for a behavior problem or more likely to have a “never” or “sometimes” response for a social skill) given a specific level of the underlying trait. The thresholds note the point where there is more than 50% likelihood that a parent will endorse a given category or a more severe one. For example, a child would need to have an underlying level of social skills better than 1.76 standard deviations below the mean before there was more than a 50% probability that the parent would endorse “usually” or “always” for the “gets along well with other children” item. Since 96% of children would be expected to have social skills better than 1.76 standard deviations below the mean, nearly all children have a greater than 50% probability of a “usually” or “always” answer for the “gets along” item. An even greater number of children have a greater than 50% probability of a “usually” or “always” response for the “shows respect” item (threshold=-1.92), but fewer children have a greater than 50% probability of a “usually” or “always” response for the “tries to resolve conflicts” item (threshold=-0.61). If this threshold between “never/sometimes” and “usually/always” is used to assess item difficulty, one might say that “usually shows respect” is an easier threshold to meet than is “usually tries to resolve conflicts”, and the “shows respect” item is less difficult than the “resolves conflicts” item in the social skills scale.

Turning now to the behavior problems scale, it is clear that the “bullies” item is an exceptionally difficult item. A child would need to have an underlying level of behavior problems greater than 0.99 standard deviations above the mean before there is a greater than 50% probability that the parent will respond “sometimes,” “usually,” or “always.” In contrast, a child would need to have an underlying level of behavior problems greater than 1.14 standard deviations *below* the mean before there is a greater than 50% probability that the parent will respond that the child sometimes, usually, or always argues too much. As might be expected, therefore, reports of arguing too much are more likely than reports of bullying.

Data from the “bullies” item in the 2003 NSCH (independent of the other behavior problems items) have been used recently in two studies about bullying (Montes and Halterman 2007; Van Cleave and Davis 2006). Other studies have

confirmed that reports of “sometimes” or more often to this single-item measure have strong construct validity (Zimmerman et al. 2005) and may be useful for prevalence estimates about bullying (Solberg and Olweus 2003). As just noted, a child with an underlying level of behavior problems greater than 0.99 standard deviations above the mean on the behavior problems scale has a greater than 50% probability that the parent will respond that the child sometimes, usually, or always bullies. An examination of children who have similar scale scores but who are reported to never bully may reveal a group of children at risk of being bullies. Researchers studying aggressive behavior in children may wish to study the validity of using the IRT scale scores for this purpose.

Another advantage provided by the IRT model is that an examination of the thresholds can reveal ways to improve the validity of the scores. For example, the analyses here conclusively demonstrated that the answer categories are appropriately ordered. That is, a child with poor social skills is more likely to get item-level reports of “never” than is a child with better social skills, and as the level of social skills increases, the likelihood of item-level reports of “sometimes,” “usually,” or “always” increase as expected. This structure was revealed by the ordered threshold values for each item. However, the thresholds and the probability curves derived from the thresholds reveal several weaknesses of the scales. First, the answer category “usually” is not the most probable response at any level of the underlying construct for any item in the social skills or behavior problems scales. This result suggests that the scales may be improved by modifying this answer category or by combining it with an adjacent category during analysis. Second, neither “usually” nor “sometimes” is the most probable response at any level of the underlying construct for the “shows respect” item, suggesting that this skill is more likely to be reported in a dichotomous (always/never) fashion. The social skills scale may be improved by modifying this item stem to capture a more continuous response. Third, the answer category “never” is not the most probable response for the “gets along well” item except when the child’s underlying level of social skills is more than 3.6 standard deviations below the mean, suggesting that many parents will report that a child at least “sometimes” gets along well with other children when even the barest level of social skills is present. The social skills scale may be improved by modifying this item stem to eliminate this floor effect.

The scale scores do not discriminate among children with good social skills or with few behavior problems. An examination of the thresholds reveals that consistent responses of “always” to the four social skills items would be expected for children with underlying levels of social skills greater than 0.48 standard deviations above the mean. Indeed, as shown in Table 2, 23% of children had the maximum score on the social skills scale. Consistent responses of “never” to the four behavior problems items were less likely due to the “argues too much” and the “stubborn” items. Parents were willing to report “sometimes” to these items even when children’s underlying trait of behavior problems was below the mean. Yet, the level of trait necessary to yield a probable response of “usually” was quite high, even for these two items. Therefore, the ability of the behavior problems scale scores to discriminate among children with few behavior problems is limited. If researchers determine that it is necessary to differentiate among children with above average social skills or with few behavior problems, the scales could be modified by adding

additional items that would reveal exceptional social skills or mild behavior problems or by adding an additional answer category that indicates “more often than sometimes” but “less often than usually.”

Other limitations are worth noting. The responses to the scale items and the data used to validate scores based on these items were obtained from a parent or another adult who lives with the child. Parent observations may be biased, especially when children become older and spend more time away from home. Validation of the reports using teacher, friend, or classmate ratings has not occurred. Ethnic, cultural, and class differences may also exist in how parents perceive their children and what they will say about them. An examination of differential item functioning by race, ethnicity, immigrant status, and social class would be valuable (cf. Spencer et al. 2005). It is also unknown whether mothers differ from fathers in their interpretation of and responses to the items. Finally, the population-based estimates presented here are derived from a sample survey that is subject to nonrandom error, including nonresponse bias, survey design flaws, data-processing mistakes, and respondent classification and reporting errors. Although all these concerns merit consideration, it is important to remember that many child indicators are derived from surveys of parents (Brown 1997), and parents can be generally reliable observers and reporters of their children’s behaviors (e.g., Glascoe and Dworkin 1995, Goodman 2001).

4.1 An Indicator of Child Well-Being

These new brief measures of social skills and behavior problems satisfy many, albeit not all, of the criteria suggested for indicators of child well-being (Moore 1997). Indicators of child well-being are intended to monitor whether the circumstances of children are growing better or worse, to motivate public policy when the levels are beneath the standards that society sets for itself, and to serve as outcome measures to evaluate the results of efforts aimed at improving children’s circumstances. To achieve these goals, at a minimum, data must be gathered on a regular basis and permit geographically detailed analyses (Brown 1997). The inclusion of the NSCH Social Competence Scale as a regular part of the NSCH, which collects representative data in a rigorous manner at the national and state level every four years, satisfies these minimum criteria.

However, the NSCH Social Competence Scale would not have been included on this survey if it did not satisfy several other criteria as well. In particular, this measure is efficient to collect. The scale contains only eight items, and these individual items are brief and quick to administer. These items and the preceding instructions require approximately 1 min and 15 s to administer over the telephone.

The items are clear, enjoy considerable face validity, and are readily understood by both policymakers and survey respondents. With declining response rates over time to all manner of surveys, it is increasingly important that survey respondents recognize the value in the questions they are asked. In the 2003 NSCH, parents of fewer than 0.03% of the children refused to answer these questions. The mix of the positively worded social skills items with the negatively worded behavior problems items likely helped to increase the acceptability of the items to parents, by permitting them to comment on their children’s desirable traits as well as their deficits.

The inclusion of positively worded items and their ability to assess a positive aspect of well-being are desirable traits for child indicators. The NSCH Social Competence Scale was developed to identify not only the absence of behavior problems, but also the presence of social skills. The balance of positive and negative aspects of social competence helps to modernize this scale and is responsive to calls for indicators that can identify if negative outcomes have been prevented *and* if worthwhile goals have been achieved. It should be noted, however, that the social skills scale did not discriminate well among children with good social skills, and the examination of the construct validity of the social skills scale was limited by the paucity of positive social outcomes available in the NSCH. Additional work is needed to determine whether children with high scores on the social skills scale have, in fact, achieved high levels of social competence.

The value of the NSCH Social Competence Scale as an indicator is enhanced by its strong internal validity. Responses to the items that comprise each scale (social skills and behavior problems) were internally consistent and were predicted by children's level on a monotonically increasing underlying trait and by the items' parameters. The strong fit of the IRT model also indicates that, controlling for trait level and item parameters, the responses to individual items were independent of responses to the other items. Reliability—defined as analogous to measurement precision—was also strong, as indicated by the discrimination parameters. The internal validity and reliability were shown to be consistent across a substantial span of child ages—6 to 17 years—and gender subgroups. However, internal validity and reliability across other demographic subgroups or over time has not been evaluated here. These are important criteria for indicators of child well-being, but the parameters of IRT models are generally not sample dependent. We expect that future researchers will find that the parameters will be largely equivalent over time and for most population subgroups.

4.2 Percentage of Children who are Socially Competent

We recognize that indicators are most useful when they can be easily and meaningfully presented to the press and to the public (Moore et al. 2004). Scale

Table 4 Percentage of children aged 6–17 years with sufficient social skills and minimal behavior problems, based on hypothetical criteria^a U.S. National Survey of Children's Health, 2003

Population	Sufficient social skills		Minimal behavior problems		Sufficient social skills and minimal behavior problems	
	Percent	Standard error	Percent	Standard error	Percent	Standard error
Sex						
Male	78.1	0.40	83.9	0.36	69.7	0.44
Female	85.0	0.37	84.1	0.38	75.3	0.44
Age (in years)						
6–11	82.9	0.39	84.3	0.37	73.4	0.45
12–17	80.1	0.39	83.7	0.36	71.5	0.43
All children	84.0	0.26	81.5	0.28	72.5	0.31

^a See text.

scores based on raw scores or IRT models cannot be meaningfully presented to lay audiences. In contrast, rates or percentages based on dichotomous measures are easily understandable, and we expect that the public and the press will want to know, for example, what percent of adolescents are socially competent. Achieving such dichotomous measures would require us to identify the score X at which a child can be considered to have sufficient social skills and the score $X-1$ at which a child can be considered to not have sufficient social skills. Similarly, we would need to identify the score Y at which a child can be considered to have an acceptably small level of behavior problems and the score $Y+1$ at which the child can be considered to have too many behavior problems.

The analyses presented here do not follow such an approach. The fact that the patterns of responses fit a Rasch-based IRT model is a strong indication that the underlying traits are best considered to be on continua. There were no natural gaps in the distributions of scores or clear substantive differences between children with one score and children with an adjacent score. There are no public policies or program eligibility criteria that might provide logical breaks in the continua, and there are no *a priori* reasons to establish cutoff scores at particular population prevalence estimates.

Hypothetically, however, let us suppose that cutoff scores were established that met the following face-valid criteria. Perhaps a child could be considered to have sufficient social skills if the child's parent responded "usually" or "always" to all 4 social skills items. A response of "usually" to all four items would result in a score of 12, so scores equal to or greater than 12 may be considered to represent sufficient social skills. Perhaps a child could be considered to have minimal behavior problems if a child's parent did not respond "usually" or "always" to any of the four behavior problems items. A score of 8 is the highest score attainable without a response of "usually" or "always," so scores less than or equal to 8 may be considered to represent minimal behavior problems. (See U.S. Department of Health and Human Services 2005, and www.childhealthdata.org for an alternative dichotomous assessment of behavior problems.) Based on these hypothetical cutoff scores and 2003 NSCH data, the percentages of children with sufficient social skills, with minimal behavior problems, and with both indicators of social competence are presented in Table 4. Additional criterion-based assessment data and further analyses will be necessary to validate these (or any other) cutoff scores.

Appendix

National Survey of Children's Health Social Competence Scale

I am going to read a list of items that sometimes describe children. For each item, please tell me how often this was true for your child during the past month. Would you say never, sometimes, usually, or always?

1. [He/She] argues too much.
2. [He/She] bullies or is cruel or mean to others.
3. [He/She] shows respect for teachers and neighbors.
4. [He/She] gets along well with other children.

5. [He/She] is disobedient.
6. [He/She] is stubborn, sullen, or irritable.
7. [He/She] tries to understand other people's feelings.
8. [He/She] tries to resolve conflicts with classmates, families, or friends.

References

- Achenbach, T. M. (1991). *Child behavior checklist/4–18*. Burlington: University of Vermont, Department of Psychiatry.
- Achenbach, T. M. (2001). *Teacher's report form for ages 5–18*. Burlington: University of Vermont, Department of Psychiatry.
- Achenbach, T. M., & Edelbrock, C. S. (1981). Behavior problems and competencies reported by parents of normal and disturbed children aged four through sixteen. *Monographs of the Society for Research in Child Development*, 46, Serial No. 188, 1–82.
- Ahorn, M. (1985). *Statistical legacies of the social indicators movement*. Paper presented at the annual meeting of the American Statistical Association, Las Vegas, Nevada.
- Anderson, S., & Messick, S. (1974). Social competency in young children. *Developmental Psychology*, 10, 282–293.
- Andrews, A. B., & Ben-Arieh, A. (1999). Measuring and monitoring children's well-being across the world. *Social Work*, 44(2), 105–115.
- Ben-Arieh, A., & Wintersberger, H. (Eds.) (1997). *Monitoring and measuring the state of children – beyond survival, Eurosocial report no. 62*. Vienna, Austria: European Centre for Social Welfare Policy and Research.
- Blumberg, S. J., Olson, L., Frankel, M., Osborn, L., Srinath, K. P., & Giambo, P. (2005). The design and operation of the National Survey of Children's Health, 2003. National Center for Health Statistics. *Vital and Health Statistics*, 1(43).
- Brown, B. V. (1997). Indicators of children's well-being: A review of current indicators based on data from the federal statistical system. In R. M. Hauser, B. V. Brown, & W. R. Prosser (Eds.) *Indicators of child well-being* (pp. 3–35). New York: Russell Sage Foundation.
- Brown, L., & Hammill, D. D. (1990). *Behavior rating profile- Second edition*. Austin, TX: PRO-ED.
- Burks, H. F. (1977). *Burks' behavior rating scales*. Los Angeles: Western Psychological Services.
- Cavell, T. (1990). Social adjustment, social performance, and social skills: A tri-component model of social competence. *Journal of Clinical Child Psychology*, 19, 111–122.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233–255.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cox, D. R., & Snell, E. J. (1989) *Analysis of binary data (2nd ed.)*. London: Chapman and Hall.
- Dickey, W. C., & Blumberg, S. J. (2004). Revisiting the factor structure of the Strengths and Difficulties Questionnaire: United States, 2001. *Journal of the American Academy of Child and Adolescent Psychiatry*, 43, 1159–1167.
- Dodge, K. A. (1985). Facets of social interaction and the assessment of social competence in children. In B. Schneider, K. H. Rubin, & J. Ledingham (Eds.) *Children's peer relations: Issues in assessment and intervention* (pp. 3–22). New York: Springer-Verlag.
- Eisenberg, N., & Fabes, R. A. (1998). Prosocial development. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology, vol. 3: Social, emotional, and personality development* (5th ed.) (pp. 701–778). New York: John Wiley.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Epps, S. R., Park, S. E., Huston, A. C., & Ripke, M. (2005). A scale of positive social behaviors. In K. Moore, & L. Lippman (Eds.) *What do children need to flourish? Conceptualizing and measuring indicators of positive development* (pp. 163–179). New York: Springer.
- Epstein, M., & Sharma, J. (1998). *Behavioral and emotional rating scale: A strength-based approach to assessment*. Austin, TX: PRO-ED.

- Federal Interagency Forum on Child and Family Statistics (2003). *America's Children: Key National Indicators of Well-Being, 2003*. Washington, DC: U.S. Government Printing Office.
- Glascoe, F. P., & Dworkin, P. H. (1995). The role of parents in the detection of developmental and behavioral problems. *Pediatrics*, *95*, 829–836.
- Goodman, R. (1994). A modified version of the Rutter parent questionnaire including extra items on children's strengths: A research note. *Journal of Child Psychology and Psychiatry*, *35*, 1483–1494.
- Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy for Child and Adolescent Psychiatry*, *40*, 1337–1345.
- Green, S. B., & Babyak, M. A. (1997). Control of type I errors with multiple tests of constraints in structural equation modeling. *Multivariate Behavioral Research*, *32*, 39–51.
- Gresham, F. (1986). Conceptual and definitional issues in the assessment of children's social skills: Implications for classification and training. *Journal of Clinical Child Psychology*, *15*, 3–15.
- Gresham, F., & Elliott, S. (1990). *Social skills rating system*. Circle Pines, MN: American Guidance Service.
- Hosp, J. L., Howell, K. W., & Hosp, M. K. (2003). Characteristics of behavior rating scales: Implications for practice in assessment and behavioral support. *Journal of Positive Behavior Interventions*, *5*(4), 201–208.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.
- Hutton, J. B., & Roberts, T. G. (1986). *Social-emotional dimension scale*. Austin, TX: PRO-ED.
- Huston, A. C., & Ripke, M. N. (2006). Experiences in middle childhood and children's development: A summary and integration of research. In A. C. Huston & M. N. Ripke (Eds.), *Developmental contexts in middle childhood: Bridges to adolescence and adulthood*. (pp. 409–434): Cambridge University Press.
- Kamphaus, R. W., & Reynolds, C. R. (1998). *BASC (behavior assessment system for children) monitor*. Circle Pines, MN: American Guidance Service.
- Levy, F., & Murnane, R. J. (2001). Key competencies critical to economic success. In D. S. Rychen, & L. H. Salganik (Eds.) *Defining and selecting key competencies* (pp. 151–173). Gottingen, Germany: Hogrefe & Huber.
- Malecki, C. K., & Elliot, S. N. (2002). Children's social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly*, *17*(1), 1–23.
- Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Journal of Multivariate Behavioral Research*, *39*, 479–515.
- Montes, G., & Halterman, J. S. (2007). Bullying among children with autism and the influence of comorbidity with ADHD: A population-based study. *Ambulatory Pediatrics*, *7*, 253–257.
- Moore, K. A. (1997). Criteria for indicators of child well-being. In R. M. Hauser, B. V. Brown, & W. R. Prosser (Eds.) *Indicators of child well-being* (pp. 36–44). New York: Russell Sage Foundation.
- Moore, K. A., Evans, V. J., Brooks-Gunn, J., & Roth, J. (2001). What are good child outcomes? In A. Thornton (Ed.), *The well-being of children and families: Research and data needs*. Ann Arbor, MI: University of Michigan Press.
- Moore, K. A., & Halle, T. G. (2001). Preventing problems vs. promoting the positive: What do we want for our children? In T. J. Owens & S. L. Hofferth (Eds.), *Children at the millennium: Where have we come from? Where are we going? (Advances in life course research series)*. New York: Elsevier Science.
- Moore, K. A., & Lippman, L. H. (Eds.). (2005). *What do children need to flourish? Conceptualizing and measuring indicators of positive development*. New York: Springer.
- Moore, K. A., Lippman, L., & Brown, B. (2004). Indicators of child well-being: The promise for positive youth development. *Annals, AAPSS*, *591*, 125–145.
- Muthén, L. K., & Muthén, B. O. (2006). *Mplus user's guide*. Los Angeles: Muthén & Muthén.
- NICHD Early Child Care Research Network (2006). The relations of classroom contexts in the early elementary years to children's classroom and social behavior. In A. C. Huston, & M. N. Ripke (Eds.) *Development contexts in middle childhood* (pp. 217–236). New York, NY: Cambridge University Press.
- Quay, H. C., & Peterson, D. R. (1987). *Revised behavior problem checklist*. Coral Gables, FL: University of Miami, Department of Psychology.
- Quint, J. C., Bos, H., & Polit, D. F. (1997). *New chance: Final report on a comprehensive program for young mothers in poverty and their children*. New York: MDRC.
- Raver, C. C., & Zigler, E. F. (1997). Social competence: An untapped dimension in evaluating Head Start's success. *Early Childhood Research Quarterly*, *12*, 363–385.

- Research Triangle Institute (2004) *SUDAAN language manual, Release 9.0*. Research Triangle Park, NC: Research Triangle Institute.
- Reynolds, C. R., & Kamphaus, R. W. (1992). *Behavior assessment system for children*. Circle Pines, MN: American Guidance Service.
- Rose-Krasnor, L. (1997). The nature of social competence: A theoretical review. *Social Development, 6*, 111–135.
- Rubin, K. H., & Rose-Krasnor, L. (1992). Interpersonal problem solving. In V. B. Van Hassett, & M. Hersen (Eds.) *Handbook of social development* (pp. 283–323). New York: Plenum.
- Rutter, M. (1967). A children's behavior questionnaire for completion by teachers: Preliminary findings. *Journal of Child Psychology and Psychiatry, 8*, 1–11.
- Rychen, D. S. (2003). Key competencies: Meeting important challenges in life. In D. S. Rychen, & L. Salganik (Eds.) *Key competencies for a successful life and a well-functioning society* (pp. 63–107). Göttingen, Germany: Hogrefe & Huber.
- Rydell, A.-M., Hagekull, B., & Bohlin, G. (1997). Measurement of two social competence aspects in middle childhood. *Developmental Psychology, 33*, 824–833.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement, 34*(4), 100–114.
- Solberg, M. E., & Olweus, D. (2003). Prevalence estimation of school bullying with the Olweus Bully/Victim Questionnaire. *Aggressive Behavior, 29*, 239–268.
- Spencer, M. S., Fitch, D., Grogan-Kaylor, A., & Mcbeath, B. (2005). The equivalence of the Behavior Problem Index across U.S. ethnic groups. *Journal of Cross-Cultural Psychology, 36*, 573–589.
- Steiger, J. H. (1998). A note on multiple sample extensions of the RMSEA fit index. *Structural Equation Modeling, 5*, 411–419.
- Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika, 52*, 393–408.
- Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland, & H. Wainer (Eds.) *Differential item functioning* (pp. 67–113). Hillsdale, NJ: Lawrence Erlbaum Associates.
- U.S. Department of Health and Human Services, Health Resources and Services Administration, Maternal and Child Health Bureau (2005). *The health and well-being of children: A portrait of states and the nation 2005*. Rockville, MD: U.S. Department of Health and Human Services.
- Van Cleave, J., & Davis, M. M. (2006). Bullying and peer victimization among children with special health care needs. *Pediatrics, 118*, e1212–e1219.
- Walker, H. M., & McConnell, S. R. (1995). *The Walker-McConnell scale of social competence and school adjustment*. San Diego: Singular.
- Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology, 85*(2), 357–364.
- Zill, N., & Brim, O. G. (1975). Childhood social indicators. *Newsletter*. Society for Research in Child Development. Fall.
- Zimmerman, F. J., Glew, G. M., Christakis, D. A., & Katon, W. (2005). Early cognitive stimulation, emotional support, and television watching as predictors of subsequent bullying among grade-school children. *Archives of Pediatric and Adolescent Medicine, 159*, 384–388.