Technical Report on ASQ:SE

ASQ:SE

Information relating to the development and psychometric studies completed on the ASQ:SE is contained in this appendix. In the first section, development of the ASQ:SE system, including item selection, is addressed, followed by a description of the initial field-test version. The second section describes the participants; the measures used to collect demographic, reliability, and validity data from the normative sample; and the procedures used to collect demographic and psychometric data. Third, demographic characteristics of the ASQ:SE research sample are described. Fourth, psychometric findings are reported, including internal consistency, test–retest reliability, concurrent validity, and known groups (criterion-referenced) validity. Findings on the utility of the ASQ:SE are reported in the fifth section.

DEVELOPMENT OF THE ASQ:SE

Item Selection

ASQ:SE items were developed using a variety of sources, including standardized social-emotional and developmental assessments, textbooks and other resources in developmental and abnormal psychology, education and intervention resources, and language and communication materials. Items were created using the following criteria:

1. Items need to be representative of critical adaptive and maladaptive behaviors at the targeted age intervals.
2. Items are easy for parents to understand and recognize.
3. Items are appropriate for a variety of cultural groups and families.

Each item was written using common words that did not exceed a sixth-grade reading level. When possible, quantitative descriptors (e.g., 15 minutes, within a 24-hour period) and concrete examples (e.g., kicks, bites other children) were provided to assist with interpretation of the item meanings.

Field-Test Version

Once items were written, they were assembled into a field-test version, which was titled the *Behavior-Ages & Stages Questionnaires* (B-ASQ; Squires, Bricker, Twombly, Yockelson, & Kim, 1996). The field-test version contained seven age intervals. The number of items per interval varied from 21 items at 6 months to 33 items at 48 months. The items in this field-test version were reviewed by experts in psychology, psychiatry, education, early childhood development, pediatrics, nursing, and mental health. Experts provided feedback on the appropriateness of items, ease of understanding items, scoring format, and content validity.

Concurrently, practitioners in approximately 50 programs across the United States used the B-ASQ with a diverse population of young children and parents, including the following:

- Families served by Healthy Start in Hawaii and Oregon
- Inner-city families in Cincinnati, Ohio; Portland, Oregon; and San Francisco, California
- Families served by Head Start or Migrant Head Start in California, Texas, and Washington State
- Families with young children identified with social and emotional problems in Arizona, California, Oregon, Utah, and Washington State

Utility questionnaires completed by service providers and parents provided feedback on the clarity of item meaning, appropriateness of items, missing content, and suggestions for revisions or additions of items.

Final Version

Based on the input gathered from experts, parents, and practitioners, and on preliminary data analyses, the B-ASQ was revised and renamed the *Ages & Stages Questionnaires: Social-Emotional (ASQ:SE).* Five types of changes were made to the field-test version. First, items with overlapping and similar content were combined. Second, items were added to fill content gaps (e.g., items were added to target so-called “red flags” for autism [Filipek, Accardo, Ashwall, et al. 2000; Filipek, Accardo, Baranek, et al.,

1999]). Third, some items were reworded to be more understandable for parents. Fourth, format changes were made to improve readability and utility for parents. Finally, a questionnaire was added at 60 months so that the questionnaire system could cover the entire infant through preschool age span.

The final English version of the ASQ:SE was translated into Spanish by Spanish-speaking personnel from a Migrant Head Start program in Oregon. The Spanish translation was used with 153 children whose families were non-English speakers. These translated questionnaires were not included in ASQ:SE reliability or validity analyses.

DATA COLLECTION PROCEDURES

Participant Recruitment

Children between the ages of 3 and 66 months and their parents were recruited for a national normative study of the ASQ:SE. Several recruiting methods were used, including gathering information from birth announcements and advertisements in Pacific Northwest newspapers; sending recruitment letters to child care providers in California and Oregon; making personal contacts with personnel in agencies serving high-risk families and young children with disabilities in California, Connecticut, Florida, Hawaii, Michigan, North Carolina, Ohio, and Oregon; and setting up information booths at children's fairs and shopping malls in Oregon and Washington State. An attempt was made to stratify the normative sample so that children and families would be representative of the U.S. population in terms of ethnicity, geographic region, parent education, income, and sex of children (Bureau of the Census, 2000a, 2000b, 2001). Recruitment letters and research protocols were approved by the University of Oregon Human Subjects Compliance Committee prior to the beginning of the study.

Measures

Three types of measures were used to collect data on the normative sample: a demographic form, the ASQ:SE questionnaires for each age interval, and two social-emotional measures with established psychometric properties (i.e., Child Behavior Checklist [CBCL; Achenbach, 1991, 1992] and Vineland Social-Emotional Early Childhood Scale [SEEC; Sparrow, Balla, & Cicchetti, 1998). The demographic form asked parents to provide information on the child's age, gender, and ethnicity, as well as information on the mother's education level and family income.

The ASQ:SE covers eight age intervals from 6 to 60 months and is described previously in Chapter 1 of this *User's Guide*. The questionnaires are designed to be completed by parents or caregivers who can provide information on a child's social-emotional competence.

To assess the concurrent validity of the ASQ:SE, a sample of children was given the CBCL or the SEEC. Additional children with a formal di-

agnosis of "social-emotional disability" were also recruited. These procedures are described next.

Procedures

Parents who indicated a willingness to participate in the study were given a packet of materials containing a consent form, ASQ:SE questionnaire, and demographic form. Packets were distributed in four ways: by mail (e.g., parents contacted through birth announcements and newspaper advertisements), by preschool teachers directly to parents, by personnel in agencies serving young children and families who distributed them to interested parents, and by research personnel directly to parents (e.g., at shopping malls and children's fairs). When parents received a packet, they were asked to return the completed forms within 1 week. If parents did not return the packets, telephone calls were made or reminder notes were sent.

After the packets were returned, a random sample of parents were contacted by telephone and asked if they would be willing to complete a second set of questionnaires and/or have their child participate in a direct assessment. Parents who agreed were given two options according to the age of their child: 1) complete a second ASQ:SE and/or the CBCL at home, to be returned in a pre-stamped envelope within 1 week (for children 24–66 months of age), or 2) permit an assessment of their child with one of two trained examiners using the SEEC in the family's home or another convenient location (for children 3–24 months of age). Prior to these assessments, the two trained examiners had established interrater reliability exceeding 95%.

Data from parental completion of the second ASQ:SE were used to examine test–retest reliability, while data gathered from parents completing the CBCL and from trained examiners' completion of the SEEC were used to examine the validity of the ASQ:SE. As the packets were returned, information from the demographic form and questionnaires as well as the CBCL and SEEC results were entered into computer files for analyses.

DEMOGRAPHIC CHARACTERISTICS OF NORMATIVE SAMPLE

Children between the ages of 3 and 66 months were recruited to examine the psychometric properties of the ASQ:SE. Approximately 10% were recruited through birth announcements in newspapers; 10% through newspaper advertisements; 30% through agency personnel who attended national conferences and agreed to field test the ASQ:SE; 35% through early intervention/early childhood special education centers and parent education programs; and 15% through children's fairs or booths at shopping malls. Data for demographic variables such as ethnicity, family income, and mother's education level were not always provided by parents for a variety of reasons (e.g., privacy). The number of children with missing data and the type of missing data are noted for each analysis.

Table A1. Number of questionnaires and gender distribution by ASQ:SE age interval

ASQ:SE age interval	Number of questionnaires		
	Total	Males	Females
6 month	355	176	175
12 month	375	189	180
18 month	323	146	172
24 month	471	249	219
30 month	298	169	126
36 month	425	199	207
48 month	457	215	221
60 month	310	153	154
Overall	3,014[a]	1,496[a]	1,454[a]

[a]Gender data missing for 64 children.

The total number of ASQ:SE assessments completed on children was 3,014. The distribution of these questionnaires by age interval and gender is shown in Table A1. The ASQ:SE total sample included 2,633 children (87%) whose families contributed at least one completed questionnaire and 381 (13%) whose families contributed two or more questionnaires at different age intervals (e.g., questionnaires at 6 and 12 months). Of the 381 families that completed two or more questionnaires, 59 contributed four or more questionnaires.

Table A2 contains a comparison of Bureau of the Census (2001) counts of the ethnic distribution with those of the ASQ:SE normative sample. There appears to be an apparent underrepresentation of Caucasians and an overrepresentation of individuals with mixed ethnicity. This is not a straightforward comparison, given the large numbers of individuals who identified themselves as mixed ethnicity.

According to data provided by the Bureau of the Census (2001), the ASQ:SE normative sample has a higher percentage of well-educated mothers than found generally in the United States, as shown in Table A3, although again these comparisons are not straightforward given differing categories of analysis. A comparison between the U.S. Census data and the ASQ:SE sample on income level indicates the ASQ:SE sample was

Table A2. Ethnicity comparison of ASQ:SE normative sample (*N* = 2,952) with 2000 Census estimates

Ethnic category	Percentage		
	ASQ:SE normative sample	2000 U.S. Census[a]	Percentage point difference
Caucasian	58.9	69.1	–10.2
African American	8.9	12.1	–3.2
Hispanic	8.6	12.5	–3.9
Asian/Pacific Islander	6.3	3.7	+2.6
Native American	1.1	0.7	+0.4
Mixed ethnicity	16.0	1.6	+14.4

Note: Ethnicity data missing for 62 children.

[a]Bureau of the Census (2001).

Table A3. Mother's education level comparison of ASQ:SE normative sample (*N* = 2,863) with 2000 Census figures

Level of education	Percentage		Percentage point difference
	ASQ:SE normative sample	2000 U.S. Census[a]	
Less than high school graduation	13.0	20.9	–7.9
High school graduation or equivalent	47.4	51.0	–3.6
Associate degree	11.9	7.5	+4.4
4-year college degree or above	25.3	20.6	+4.7
Don't know	2.4	—[b]	—

Note: Mother's level of education data missing for 151 children.

[a]Bureau of the Census (2000b).

[b]U.S. Census does not include a "Don't know" category.

composed of a higher percentage of families with lower incomes than is found in the general population, as shown in Table A4.

Data taken from the demographic form permitted dividing the ASQ:SE normative sample into four groups according to the children's developmental status: 1) no risk (i.e., children with one or no identified environmental/medical risk factors), 2) at risk (i.e., children with two or more risk factors), 3) developmental disability (i.e., children with established developmental disabilities who were receiving early intervention/early childhood special education services through IDEA), and 4) social-emotional disability (i.e., children with identified social-emotional disabilities, according to IDEA Part B eligibility guidelines and the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* [DSM-IV; American Psychiatric Association, 1994], diagnostic classifications). Variables used to determine level of risk for the first two groups included the following:

1. Family income less than $12,000
2. Mother less than 18 years old when child was born
3. Mother's level of education less than high school graduation
4. Involvement of child protective services with family
5. Child in foster care
6. Birth weight less than 3 pounds, 5 ounces

Table A5 presents the number of children by developmental status in the normative sample.

PSYCHOMETRIC FINDINGS

The following sections discuss how the cutoff scores for the ASQ:SE were developed. In addition, data collected from subgroups of the normative sample used to examine the internal consistency, test–retest, concurrent validity, known groups validity, and utility of the ASQ:SE are presented.

Table A4. Family income level comparison of ASQ:SE normative sample (*N* = 1,992) with 1999 Bureau of the Census estimates

ASQ:SE		1999 Bureau of the Census estimates		
Income category	Percentage of normative sample	Income category[a]	Percentage of population	Percentage point difference
$0–12,000	20.6	Less than $9,999	9.2	+11.4
$12,001–24,000	19.9	$10,000–24,999	21.3	–1.4
$24,001–40,000	22.8	$25,000–39,999	18.4	+4.4
More than $40,000	29.9	More than $40,000	51.1	–21.2
Don't know	6.8	—[b]	—	—

Note: Family income level data missing for 1,022 children.

[a]Bureau of the Census (2000a).

[b]U.S. Census does not include a "Don't Know" category.

Establishing Reliability

Internal Consistency Internal consistency measures the extent to which items on the assessment tool measure the same underlying construct (Salvia & Ysseldyke, 1998). High internal consistency reflects items that assess the same characteristic or behavioral area. To measure internal consistency, coefficient alpha was calculated for each ASQ:SE age interval using the variances of individual items and the variance of the total test scores (*N* = 1,994). Cronbach's coefficient alphas for the ASQ:SE age intervals are shown in Table A6. Alphas ranged from .67 to .91, with an overall alpha of .82. An alpha of .70 is considered to be an adequate measure of internal consistency (Nunnally, 1978).

Test–Retest Reliability Test–retest reliability measures the stability of child performance across time. Test–retest reliability for the ASQ:SE was determined by comparing the results of two questionnaires completed by parents at 1- to 3-week intervals. A random sample of parents (*N* = 367) was asked to complete a second, identical ASQ:SE after returning the first completed questionnaire; parents were "blind"

Table A5. Number of children by developmental status for ASQ:SE normative sample (*N* = 2,861)

		Number of subjects by developmental status							
		No risk[a]		At risk[b]		Developmental disability[c]		Social-emotional disability[d]	
ASQ:SE age interval	*N*	*n*	Mean	*n*	Mean	*n*	Mean	*n*	Mean
6 month	286	84	19.2	166	20.6	27	29.8	9	77.4
12 month	293	103	22.1	145	26.2	38	40.4	7	67.9
18 month	264	115	22.2	100	32.2	41	68.0	8	97.5
24 month	389	172	25.8	141	37.5	56	46.4	20	86.6
30 month	245	114	33.5	78	46.2	40	86.8	13	107.2
36 month	347	191	33.3	81	47.5	48	81.8	27	119.1
48 month	378	176	31.6	123	52.2	51	76.8	28	130.5
60 month	277	134	30.1	85	47.7	29	69.5	29	132.9
Overall	2,479	1,089	27.2	919	38.8	330	62.4	141	102.4

Note: Developmental status data missing for 382 children.

[a]One or no identified risk factors.

[b]Two or more identified risk factors.

[c]Children receiving early intervention or early childhood special education services.

[d]Children with diagnosed social-emotional disabilities.

to the results of their first completed ASQ:SE. The percent agreement between classifications of the child's performance on the ASQ:SE at Time 1 (first questionnaire) and Time 2 (second questionnaire) were used to measure test–retest reliability.

Children were classified as "okay" on the ASQ:SE (no further evaluation of social-emotional competence was indicated) if their scores were below the empirically derived cutoff point for that interval. Children were classified as "at risk" on the ASQ:SE (further evaluation of their social-emotional status was indicated) if their scores were on or above the cutoff point. Using the McNemar Test (Agresti, 1990) assessing dependent proportions, test-retest agreement was 94% (N = 344/367).

Establishing Validity

The primary goal of a screening measure is to accurately discriminate between individuals who are typical or okay (i.e., do not have the problem or characteristic) on a targeted variable (e.g., development, medical condition such as PKU) and individuals who appear atypical or not okay (i.e., potentially may have the problem or characteristic). Establishing the validity of a screening measure generally requires a two-step process. First, it is necessary to collect sufficient normative data to establish optimal cutoff scores for the screening test. Individuals who fall above the cutoff score are classified as at risk and in need of follow-up, while individuals who score below the cutoff score are classified as okay and do not need follow-up.

Table A6. Cronbach coefficient alpha by ASQ:SE age interval (N = 1,994)

ASQ:SE age interval	Number of questionnaires	Alpha
6 month	196	.69
12 month	196	.67
18 month	210	.81
24 month	297	.80
30 month	198	.88
36 month	281	.89
48 month	317	.91
60 month	299	.91
Overall	1,994	.82

Note: Field-test versions of the B-ASQ (N = 867) and Spanish translation (N = 153) were not included in this analysis.

For any screening test, there are no absolute scores that separate typical from nontypical individuals. Rather, data must be collected and examined to determine optimal cutoff scores, that is, scores that correctly classify children as needing or not needing follow-up evaluation. Finding optimal cutoff scores requires examining a range of alternatives to discover those scores that maximize the identification of individuals who should receive further testing (i.e., true positives) while minimizing the misidentification of individuals who do not require further testing (i.e., false positives) and minimizing the nonidentification of individuals who should receive further testing (i.e., false negatives).

Once "tentative" cutoff scores are selected, the second step is to determine if they do accurately discriminate between individuals who require follow-up and individuals who do not. Thus, first it is necessary to establish what are thought to be optimal cutoff scores for the screening measure. Once cutoff scores are selected, it is then necessary to determine their accuracy and thus the validity of the screening measure. Establishing the validity of a screening measure is done by comparing an individual's classification on the screening measure with his or her classification on a selected criterion measure(s). Using this two-step process, the validity of the ASQ:SE was examined by comparing children's classification (i.e., developmentally okay or at risk) on the ASQ:SE with their classification (i.e., developmentally okay or at risk/disabled) on selected criterion measures that included the CBCL, the SEEC, and professional diagnosis of a social-emotional disability.

To discover optimal ASQ:SE cutoff scores (i.e., those that yield high true positives, high true negatives, low false positives, and low false negatives), receiver operating characteristic (ROC) curves were used (Swets & Pickett, 1982). ROC analysis permits the systematic comparison of true positive probabilities against false positive probabilities for a range of possible cutoff scores. To create these comparison data, a sample of 1,041 children with completed ASQ:SE questionnaires were given a concurrent criterion measure—either the CBCL or the SEEC—or had a professional diagnosis of a social-emotional disability. Each child's classification (i.e., okay or at risk) on the ASQ:SE was then compared with his or her classification (i.e., okay or at risk/disabled) on one of the criterion measures. Figure 16 in Chapter 5 shows a four-cell contingency table used to assess the agreement between the screening measure (i.e., ASQ:SE) and the follow-up criterion measure (i.e., CBCL, SEEC, or diagnosis of social-emotional disability). In addition, this figure shows the formulas for calculating the percentage of children identified as needing further assessment, percent agreement, sensitivity, specificity, overreferral, underreferral, and positive predictive value.

Comparison of means, medians, interquartile ranges, and ROC cutoffs is shown in Table A7. It can be noted that ROC cutoff scores for most age intervals were similar to scores derived from adding 1.5 semi-interquartile ranges to medians. The general trend of increasingly higher scores as children develop is reflected in both mean and median scores,

Table A7. Means, medians, interquartile ranges, and ROC by ASQ:SE age interval (*N* = 2,861)

ASQ:SE age interval	*N*	Mean	Median	Interquartile range	Median + 1.5 semi-interquartile ranges	ROC cutoff score[a]
6 month	331	22.5	16.7	22.5	34	45
12 month	339	27.7	25.0	22.0	42	48
18 month	307	34.6	26.0	26.6	46	50
24 month	441	35.4	28.4	33.8	54	50
30 month	289	48.6	35.2	41.5	66	57
36 month	408	49.9	35.0	48.9	72	59
48 month	447	55.7	36.0	52.6	75	70
60 month	299	47.5	35.0	45.0	69	70

Note: Data from the B-ASQ at 6, 12, 24, 30, 36, and 48 months (*N* = 867) were combined with data from ASQ:SE at the same intervals after *t* tests revealed no significant differences between the field-test version and the ASQ:SE at these age intervals.

[a]ROC cutoff based on "best fit," maximizing true positives and true negatives.

except at 60 months. The leveling or decrease in scores at 60 months may be the artifice of a smaller sample at that age interval.

Frequently, cutoff scores for screening tools are set by using means and standard deviations. That is, the mean score plus one standard deviation is a likely choice for a cutoff score. However, using means to calculate cutoff scores presumes a normal distribution of scores. Score distribution for the ASQ:SE questionnaires was positively skewed—that is, the majority of children obtained low scores (i.e., indicating they have no problem or are okay) and relatively few children obtained high scores (i.e., indicating they have a potential problem or are at risk). Figure A1

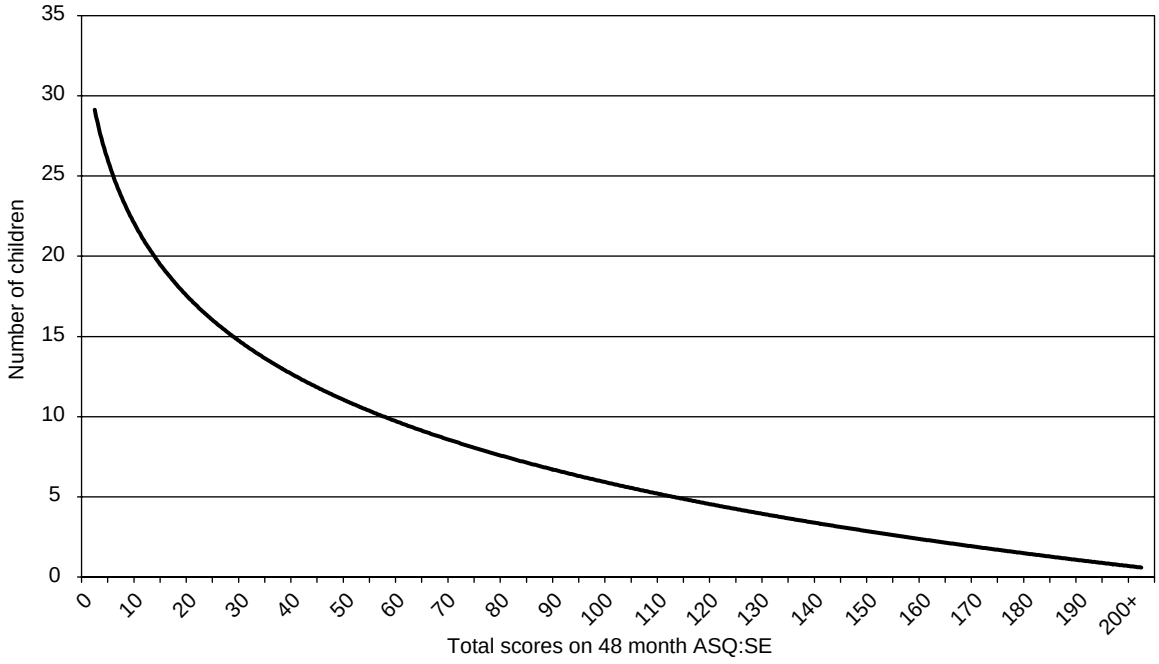

Figure A1. ASQ:SE total scores by number of children, showing a positively skewed distribution.

shows the positively skewed distribution of scores for the 48 month ASQ:SE; other age intervals showed similar score distributions. Means and standard deviations were not used for determining cutoff points because of the positive skew of ASQ:SE scores across intervals. Instead, ROC analyses were conducted to determine the best cutoff point for each interval.

To examine gender differences, scores for males and females were compared. Mean and median scores by gender are presented in Table A8. Box plots were then derived to examine the score distributions by gender. Box plots for the 30 and 36 month ASQ:SE male and female score distributions are shown in Figure A2. Box plots provide a visual "picture" of a distribution. The bottom line of the box is the 25th percentile, or Quartile 1. The top line of the box is the 75th percentile, or Quartile 3. The middle line is the median, or Quartile 2; the mean is indicated by the +. Whiskers (lines) extend to the highest and lowest observations, but not further than 1.5 interquartile ranges from the median. Outliers beyond 1.5 interquartile ranges are indicated by O; outliers beyond 3 interquartile ranges are indicated by O*.

As shown in Figure A2, the majority of scores for males at the 30 month interval range between 25 and 75, with the 1.5 interquartile range extending to 0 and to 160. Outliers extend upward to 300. For females, the range is between 20 and 50, with the 1.5 interquartile range extending to 75. Outliers extend beyond 200. A similar distribution for both males and females can be seen at the 36 month interval. Similar distribution patterns occurred at all age intervals and indicate, in general, that males tend to have greater dispersal of scores and more extreme scores.

When gender group differences are compared using the nonparametric Kruskal-Wallis Test (Heiman, 1992), significant differences are found at 30, 36, 48, and 60 months, as shown in Table A8. It is important to note that the validity sample currently does not have adequate numbers of females with social-emotional problems to indicate whether separate cutoff scores for females are needed. Consequently, girls whose scores are close to the cutoffs at the 30, 36, 48, and 60 month intervals should be considered for referral. As additional data are added to the validity sample, revised cutoffs, if necessary, will be posted on the Paul H. Brookes Publishing Co. website (http://www.brookespublishing.com).

MODIFYING CUTOFF SCORES

If programs want to modify cutoff scores, semi-interquartile ranges (i.e., median + [quartile 1 – quartile 3] / 2) should be used as the basis for modification. See the Paul H. Brookes Publishing Co. website (http://www.brookespublishing.com/asqse) for discussion of guide-lines for altering ASQ:SE cutoff points.

Table A8. ASQ:SE mean and median scores for males and females by age interval (*N* = 2,801)

ASQ:SE age interval	Male (*N* = 1,421)			Female (*N* = 1,380)		
	N	Mean	Median	*N*	Mean	Median
6 month	164	25.4	20.0	163	20.5	15.0
12 month	171	27.7	25.0	163	27.0	25.0
18 month	140	37.7	25.0	164	33.2	30.0
24 month	233	39.1	31.2	205	32.5	25.0
30 month	163	57.3	37.6	123	39.1***	33.4
36 month	190	58.3	40.0	200	40.4***	30.0
48 month	212	61.8	46.4	214	40.3***	26.6
60 month	148	57.8	40.6	148	36.4***	25.0

Note: Gender data missing for 60 children.

***Significant at $p < .001$.

Once optimal cutoff scores were established, the next step was to examine the agreement between the classification of children using these cutoffs with selected criterion measures. Both concurrent and known groups validity of the ASQ:SE have been examined, and the findings are reported in the following two sections.

Examining Concurrent Validity

To determine how accurately the ASQ:SE discriminates between children whose social-emotional development is proceeding without problem and children who have or who are at risk for developing social-emotional problems, a comparison with selected criterion measures was necessary. Criterion measures chosen to examine the concurrent validity, or discriminative power, of the ASQ:SE were the CBCL and SEEC.

The CBCL is a well-studied tool with reported adequate psychometric properties (Achenbach, 1991, 1992) and is consider the "gold standard" against which most new tools assessing social-emotional competence are measured (McConaughy, 1992). The CBCL has two forms, one for ages 2–3 years (CBCL/2-3; Achenbach, 1992) and one for ages 4–18 years (CBCL/4-18; Achenbach, 1991). Children who had scores of 61 or above on the CBCL/2-3 and of 64 or above on the CBCL/4-18 were classified as having social-emotional disabilities. (Achenbach & Rescorla's CBCL/1½–5 was not published until 2000; a decision was made to retain the CBCL/2-3 for all ASQ:SE psychometric studies.)

The SEEC is a measure frequently used to assess the social-emotional competence of young children. Psychometric data on the SEEC suggest it is both reliable and valid (Sparrow et al., 1998), although new studies have not been conducted since the original study of the Vineland Adaptive Behavior Scale (Sparrow, Balla, & Cicchetti, 1984) in the 1980s. While the CBCL was completed by parents or caregivers in their homes, the SEEC was completed through an interview with the parent. Children were classified as having a social-emotional disability if their scores on the SEEC were 70 or below (Sparrow et al., 1998).

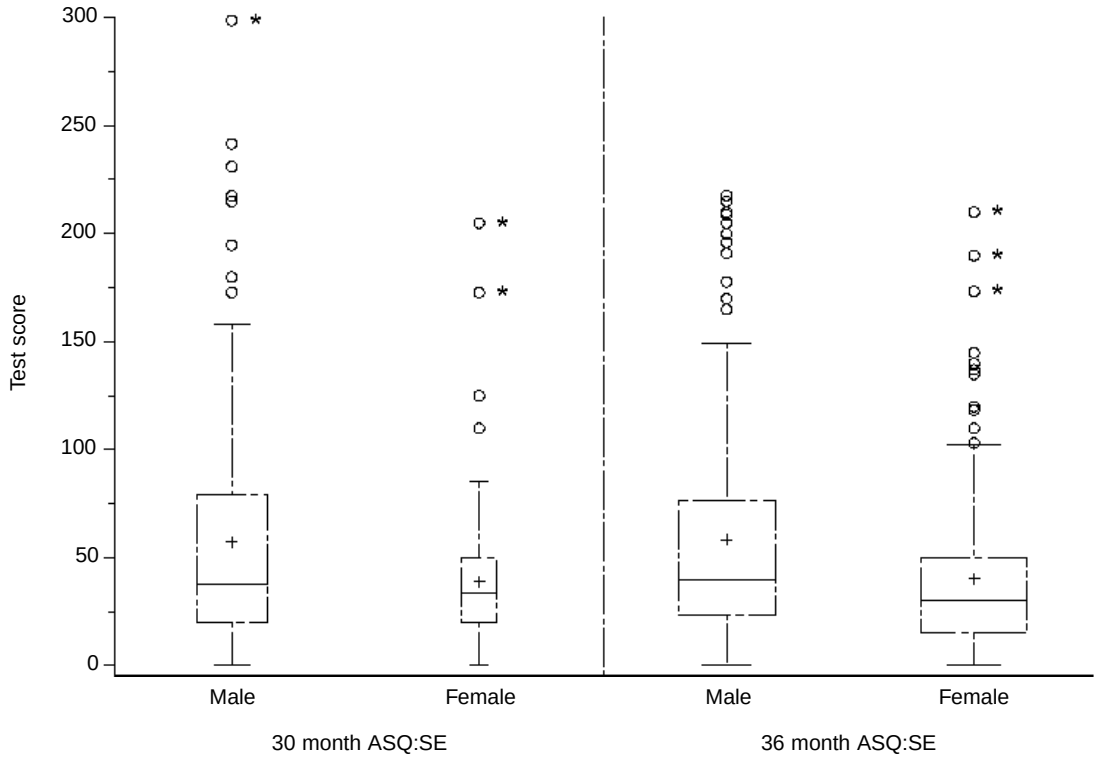

Figure A2. Box plots for 30 and 36 month ASQ:SE by gender. Box plots illustrate the distribution of scores. Bottom line of box is 25th percentile, or Quartile 1. Top line of box is 75th percentile, or Quartile 3. Middle line of box is the median, or Quartile 2. Mean is indicted by +. Whiskers (lines) extend to the highest and lowest observations, but not further than 1.5 interquartile ranges. Outliers beyond 1.5 interquartile ranges are indicated by **O**. Outliers beyond 3 interquartile ranges are indicated by **O***. Box width varies with *n.*

Parents or other primary caregivers of children in the validity sample (N = 1,041) completed either the CBLC or the SEEC within 2–3 weeks of completing the ASQ:SE on their child. In addition, the validity sample included 88 children ranging in age from 2½ to 5 years who had been professionally diagnosed as having a social-emotional disability and were receiving intervention services. Parents completed an ASQ:SE on these children as well.

Children in the validity sample were classified as either okay or at risk based on their ASQ:SE scores and the established cutoffs and were independently classified as either okay or at risk/disabled using their scores on the CBLC or SEEC or based on professional diagnosis. These two independent classifications were then compared for all children in the validity sample. One of four outcomes was possible: 1) the ASQ:SE and criterion measure both classified the child as okay (i.e., true negatives); 2) the ASQ:SE and criterion measure both classified the child as at risk/disabled (i.e., true positives); 3) the ASQ:SE classified the child as okay, while the criterion measure classified the child as at risk/disabled (i.e., false negatives); and 4) the ASQ:SE classified the child as at risk while the criterion measure classified the child as okay (i.e., false positives).

To conduct these comparisons, contingency tables containing four cells (i.e., A = true positives, B = false positives, C = false negatives, and D = true negatives, as shown in Figure 16 in Chapter 5) were developed for each of the ASQ:SE age intervals using the ROC cutoff scores listed in Table A7. Each contingency table contained in Figure A3 shows the absolute agreement for true positives, false positives, false negatives, and true negatives. From the data contained in the contingency table, the sensitivity, specificity, false positive rate, true positive rate, false negative rate, percent agreement, underreferral, overreferral, percent referral, and positive predictive value were calculated for each ASQ:SE age interval. An overall comparison across all intervals is shown in Figure A4.

Table A9 presents a comparison of the sensitivity, specificity, false positive rate, false negative rate, percent agreement, underreferral, and overreferral across ASQ:SE age intervals. Sensitivity ranged from a low of 70.8% at 24 months to a high of 84.6% at 60 months. Specificity ranged from 89.5% at 30 months to 98.2% at 6 months. Percent agreement ranged from 87.8% at 30 months to 94.0% at 60 months. Underreferral ranged from 2.4% at 60 months to 4.7% at 12 months, while overreferral ranged from 3.0% at 18 months to 8.6% at 30 months. These findings suggest the ASQ:SE is generally accurate in discriminating between children who are developing okay and those who need follow-up. In practical terms, the underreferral rate appears acceptable across intervals and never exceeds 4.6%, while the false positive and overreferral rate are consistently high. This finding suggests that parents using the ASQ:SE are consistently identifying problems in their children that the criterion measures do not. At least two possible explanations seem reasonable. First, the ASQ:SE may be consistently overscreening children, or second, the criterion measures may consistently be missing children who have social-emotional problems. Only follow-up of overreferred children (using the criterion measure classification) will determine which explanation is correct.

Examining Known Groups Validity

Another approach to assessing validity of a screening measure suggested by Spector (1992) requires examining the differences in scores across groups. For this analysis, children in the validity sample were divided into four groups based on their developmental risk status: no risk, at risk, developmental disability, and social-emotional disability. Children were assigned to the no-risk group if parents reported one or no risk factors (N = 812), were assigned to the at-risk group if parents reported two or more risk factors (N = 790), were assigned to the developmental disability group if they were receiving general early intervention services (N = 297), or were assigned to the social-emotional disability group if they had been diagnosed with a behavior or emotional problem and were receiving intervention services (N = 88). Risk factors included 1) annual family income less than $12,000; 2) mother less than 18 years old when child was born; 3) mother's level of education less than high school graduation; 4) in-

6 month ASQ:SE

		Criterion measure classification		
		At risk	*Okay*	
ASQ:SE classification	*At risk*	11	1	12
	Okay	3	56	59
	Total	14	57	71

Sensitivity	Specificity	False positive	True positive	False negative	Percent agreement	Under-referral	Over-referral	Percent referral	Positive predictive value
78.6%	98.2%	8.3%	78.6%	5.1%	94.0%	4.2%	1.4%	17.0%	91%

12 month ASQ:SE

		Criterion measure classification		
		At risk	*Okay*	
ASQ:SE classification	*At risk*	10	2	12
	Okay	4	69	73
	Total	14	71	85

Sensitivity	Specificity	False positive	True positive	False negative	Percent agreement	Under-referral	Over-referral	Percent referral	Positive predictive value
71.4%	97.2%	16.7%	71.4%	5.5%	93.0%	4.7%	2.4%	15.0%	83%

18 month ASQ:SE

		Criterion measure classification		
		At risk	*Okay*	
ASQ:SE classification	*At risk*	9	3	12
	Okay	3	84	87
	Total	12	87	99

Sensitivity	Specificity	False positive	True positive	False negative	Percent agreement	Under-referral	Over-referral	Percent referral	Positive predictive value
75.0%	96.6%	25.0%	75.0%	3.4%	93.9%	3.0%	3.0%	12.0%	75%

24 month ASQ:SE

		Criterion measure classification		
		At risk	*Okay*	
ASQ:SE classification	*At risk*	17	9	26
	Okay	7	119	126
	Total	24	128	152

Sensitivity	Specificity	False positive	True positive	False negative	Percent agreement	Under-referral	Over-referral	Percent referral	Positive predictive value
70.8%	93.0%	34.6%	70.8%	5.6%	89.5%	4.6%	5.9%	17.0%	65%

(continued)

Figure A3. Contingency tables showing agreement between ASQ:SE classification and criterion measure classification and ASQ:SE sensitivity, specificity, false positive rate, true positive rate, false negative rate, percent agreement, underreferral, overreferral, percent referral, and positive predictive value by age interval (definitions and formulas are contained in Figure 16 in Chapter 5). Criterion measure classification includes CBCL, SEEC, and professional diagnosis.

Figure A3. *(continued)*

30 month ASQ:SE

		Criterion measure classification		
		At risk	*Okay*	
ASQ:SE classification	*At risk*	10	19	29
	Okay	3	84	87
	Total	13	103	116

Sensitivity	Specificity	False positive	True positive	False negative	Percent agreement	Under-referral	Over-referral	Percent referral	Positive predictive value
80.0%	89.5%	38.5%	80.0%	4.5%	87.8%	3.4%	8.6%	23.0%	61%

36 month ASQ:SE

		Criterion measure classification		
		At risk	*Okay*	
ASQ:SE classification	*At risk*	28	10	38
	Okay	8	133	141
	Total	36	143	179

Sensitivity	Specificity	False positive	True positive	False negative	Percent agreement	Under-referral	Over-referral	Percent referral	Positive predictive value
77.8%	93.0%	26.3%	77.8%	5.7%	89.9%	4.5%	5.7%	21.0%	73%

48 month ASQ:SE

		Criterion measure classification		
		At risk	*Okay*	
ASQ:SE classification	*At risk*	20	8	28
	Okay	6	140	146
	Total	26	148	174

Sensitivity	Specificity	False positive	True positive	False negative	Percent agreement	Under-referral	Over-referral	Percent referral	Positive predictive value
76.9%	94.6%	28.6%	76.9%	4.1%	92.0%	3.4%	4.6%	16.0%	71%

60 month ASQ:SE

		Criterion measure classification		
		At risk	*Okay*	
ASQ:SE classification	*At risk*	22	6	28
	Okay	4	136	140
	Total	26	142	168

Sensitivity	Specificity	False positive	True positive	False negative	Percent agreement	Under-referral	Over-referral	Percent referral	Positive predictive value
84.6%	95.8%	21.4%	84.6%	2.9%	94.0%	2.4%	3.6%	18.1%	71%

Overall

		Criterion measure classification		
		At risk	*Okay*	
ASQ:SE	*At risk*	131	48	179
classification	*Okay*	37	825	862
	Total	168	873	1,041

Sensitivity	Specificity	False positive	True positive	False negative	Percent agreement	Under-referral	Over-referral	Percent referral	Positive predictive value
78.0%	94.5%	26.8%	78.0%	4.3%	91.8%	3.6%	4.6%	17.2%	26.8%

Figure A4. Contingency table showing overall agreement (combined across age intervals) between ASQ:SE classification and criterion measure classification and ASQ:SE sensitivity, specificity, false positive rate, true positive rate, false negative rate, percent agreement, underreferral, overreferral, percent referral, and positive predictive value by age interval (definitions and formulas are contained in Figure 16 in Chapter 5). Criterion measure classification includes CBCL, SEEC, and professional diagnosis.

volvement of child protective services with family; 5) child in foster care; and 6) birth weight less than 3 pounds, 5 ounces.

Figure A5 presents the mean scores for the four groups across the 6, 12, 18, 24, 30, 36, 48, and 60 month ASQ:SE intervals. Differences between risk groups were examined using the nonparametric Kruskal-Wallis Test (Heiman, 1992). Significant differences ($p < .0001$) were found between groups at all ASQ:SE age intervals. These findings suggest the ASQ:SE can accurately discriminate between children whose social-emotional development is typical and those who have disabilities. An example of box plots showing the distribution of risk groups for the 48 month ASQ:SE is presented in Figure A6. The box plots clearly show that mean (marked with +) and median (middle horizontal line in each box) scores increase as risk factors increase. In addition, there is almost no overlap in the distribution of scores between the no risk and social-emotional disability groups. Children with diagnosed social-emotional disabilities had the highest scores, while children in the no risk group had the lowest scores.

Table A9. ASQ:SE cutoff scores and classification statistics by age interval based on ROC cutoff score (*N* = 1,041)

ASQ:SE age interval	*N*	Cutoff score	Sensitivity	Specificity	False positive rate	False negative rate	Percent agreement	Under-referral	Over-referral
6 month	71	45	78.6	98.2	8.3	5.1	94.0	4.2	1.4
12 month	85	48	71.4	97.2	16.7	5.5	93.0	4.7	2.4
18 month	99	50	75.0	96.6	25.0	3.4	93.9	3.0	3.0
24 month	152	50	70.8	93.0	34.6	5.6	89.5	4.6	5.9
30 month	115	57	80.0	89.5	38.5	4.5	87.8	3.4	8.6
36 month	179	59	77.8	93.0	26.3	5.7	89.9	4.5	5.7
48 month	174	70	76.9	94.6	28.6	4.1	92.0	3.4	4.6
60 month	168	70	84.6	95.8	21.4	2.9	94.0	2.4	3.6
Overall	1,041		78.0	94.5	26.8	4.3	91.8	3.6	4.6

Note: See Figure 16 in Chapter 5 for formulas used in calculating classification statistics.

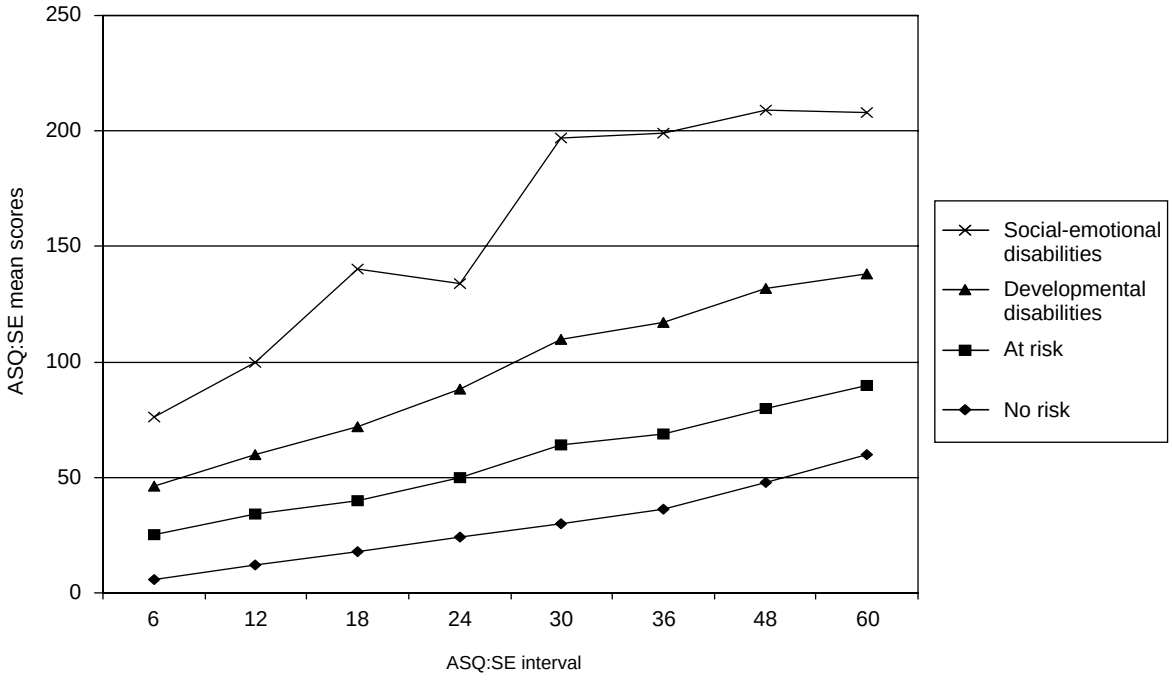

Figure A5. Mean ASQ:SE scores by group risk status.

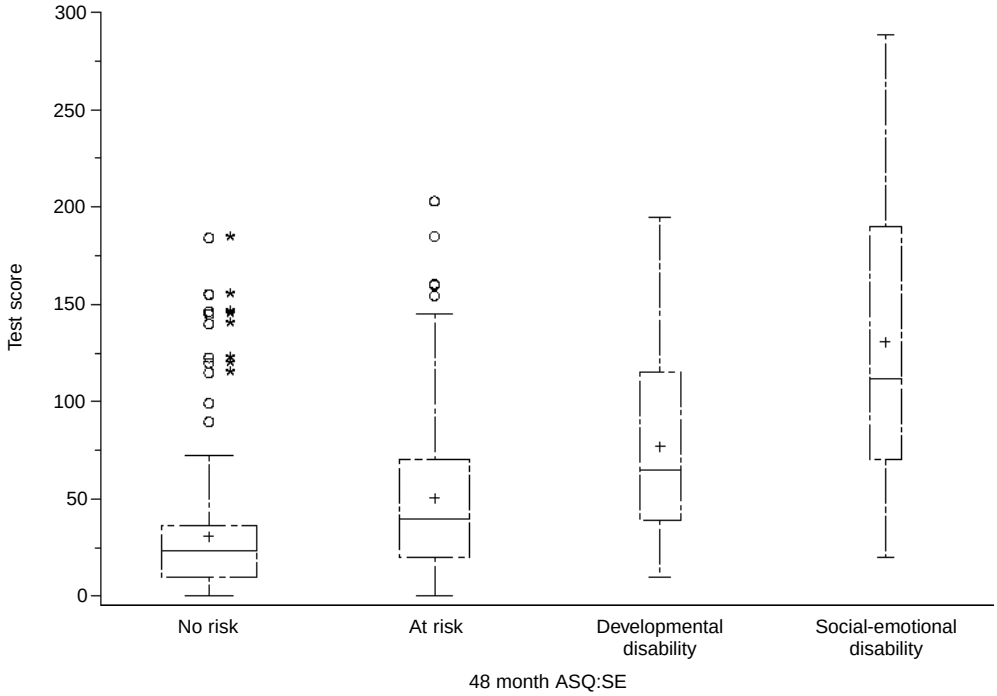

Figure A6. Box plots of distribution of total scores on 48 month ASQ:SE by developmental status. Box plots illustrate the spread of distribution. Bottom line of box is 25th percentile, or Quartile 1. Top line of box is 75th percentile, or Quartile 3. Middle line of box is the median, or Quartile 2. Mean is indicted by +. Whiskers (lines) extend to the highest and lowest observations, but not further than 1.5 interquartile ranges. Outliers beyond 1.5 interquartile ranges are indicated by **O**. Outliers beyond 3 interquartile ranges are indicated by **O***. Box width varies with *n.*

UTILITY STUDIES

Utility of a screening tool measures the usefulness or practicality of the test or procedure (Bricker & Squires, 1989). A random sample of parents (N = 731) who completed the ASQ:SE were asked to complete a utility questionnaire that requested their opinion about the length, appropriateness, and ease of completion of the ASQ:SE. Summary results of the utility questionnaire can be found in Table A10. Sixty percent of the respondents indicated that it took less than 10 minutes to complete. Ninety-six percent of the respondents indicated the ASQ:SE was easy to understand; ninety percent noted that question content was appropriate. Thus, parents reported that the ASQ:SE was easy to understand, that it took little time to complete, and that the questions were appropriate.

In addition, parents indicated that completing the ASQ:SE was interesting and helped them to think about the social and emotional development of their young children. Six parents thought that questions related to sexual interest on the 36, 48, and 60 month ASQ:SE were inappropri-

Table A10. Parent responses (N = 731) to ASQ:SE utility questionnaire items

Question	Percentage of parents reporting
1. How long to complete ASQ:SE?	(Missing 4 responses)
a. Less than 10 minutes	60
b. 10–20 minutes	32
c. 20–30 minutes	4
d. 30 minutes–1 hour	0
e. More than 1 hour	0
2. Was ASQ:SE easy to understand?	(Missing 7 responses)
a. Yes	96
b. Sometimes	3
c. No	0
3. Were ASQ:SE questions appropriate?	(Missing 13 responses)
a. Yes	90
b. Sometimes	7
c. No	1
4. The ASQ:SE questionnaire was . . . (check all that apply)	
a. Fun to do	38
b. Interesting	57
c. Took too long	1
d. Helped me think about my child	71
e. Waste of time	2
f. Didn't tell me much	10
5. Would you fill out another ASQ:SE?	(Missing 33 responses)
a. Yes	91
b. No	5
6. Would you change the ASQ:SE?	
a. Yes	16
b. No	84

ate; however, a decision was made to retain these items because of their importance to identification of disturbances related to sexual abuse and early exposure to domestic violence.

SUMMARY

Psychometric studies on the ASQ:SE are summarized in this Technical Report. Normative data were based on 3,014 completed questionnaires; validity studies were conducted using 1,041 children. Internal consistency measured by coefficient alpha was found to be high across intervals, ranging from .67 to .91 with an overall alpha of .82. Test–retest reliability, measured as the agreement between two ASQ:SE questionnaires completed by parents at 1- to 3-week intervals was 94%. Sensitivity ranged from 71% at 24 months to 85% at 60 months, with 78% overall sensitivity. Specificity of the questionnaires ranged from 90% at 30 months to 98% at 6 months, with 94% overall. Percent agreement between questionnaires and standardized assessments/disability status ranged from 88% at 30 months to 94% at 60 months, with overall agreement of 92%. Underreferral ranged from 2.4% at 60 months to 4.7% at 12 months, while overreferral ranged from 3.0% at 18 months to 8.6% at 30 months. The ability of the ASQ:SE to detect atypical social-emotional development (sensitivity) was generally lower across intervals, while specificity, or the ability of the ASQ:SE to correctly identify typically developing children, was high. Specificity may have been elevated in the 6, 12, and 18 month intervals because of the large number of "identified" children in these samples and the small number of low-moderate risk children.

Research is continuing on the ASQ:SE. Specifically, additional young children with atypical social-emotional development—particularly girls—are being recruited for validity studies. In addition, results of ASQ:SE completed by parents and teachers are being compared to study its interrater reliability. Research findings will be posted at the Paul H. Brookes Publishing Co. website (http://www.brookespublishing.com) as they become available.

REFERENCES

Achenbach, T. (1991). *Manual for the Child Behavior Checklist/4–18 and 1991 profile.* Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T. (1992). *Manual for the Child Behavior Checklist/2–3 and 1992 profile.* Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T., & Rescorla, L. (2000). *Manual for the ASEBA preschool forms and profiles.* Burlington, VT: ASEBA.

Agresti, A. (1990). *Categorical data analysis.* New York: John Wiley & Sons.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Bricker, D., & Squires, J. (1989). The effectiveness of screening at-risk infants: Infant Monitoring Questionnaire. *Topics in Early Special Childhood Education, 3*(9), 67–85.

Bureau of the Census. (2000a, September). *Table 13. Money income in the United States: Current population reports: Consumer income, 1999, P60-209* [On-line]. Available: http://www.census.gov.

Bureau of the Census. (2000b, December). *Table 1: Educational attainment of the population 15 years and over, by sex (female, all races)* [On-line]. Available: http://www.census.gov.

Bureau of the Census. (2001, April). *PHC-T-1: Population by race and Hispanic or Latino origin* [On-line]. Available: http://www.census.gov.

Filipek, P.A., Accardo, P.J., Ashwall, S., et al. (2000). Practice parameter. Screening and diagnosis of autism: Report of the Quality Standards Subcommittee of the American Academy of Neurology and the Child Neurology Society. *Neurology, 55*(4), 468–479.

Filipek, P.A., Accardo, P.J., Baranek, G.T., et al. (1999). The screening and diagnosis of autistic spectrum disorders. *Journal of Autism and Developmental Disorders, 29*(6), 439–484.

Heiman, G. (1992). *Basic statistics for the behavioral sciences.* Boston: Houghton Mifflin.

McConaughy, S.H. (1992). Objective assessment of children's behavioral and emotional problems. In C.E. Walker & M.C. Roberts (Eds.), *Handbook of clinical child psychology* (pp. 163–180). New York: John Wiley & Sons.

Nunnally, J.C. (1978). *Psychometric theory* (2nd ed.). New York, McGraw-Hill.

Salvia, J., & Ysseldyke, J. (1998). *Assessment* (7th ed.). Boston: Houghton Mifflin.

Sparrow, S., Balla, D., & Cicchetti, D. (1984). *Vineland Adaptive Behavior Scales (VABS).* Circle Pines, MN: American Guidance Service.

Sparrow, S., Balla, D., & Cicchetti, D. (1998). *Vineland Social-Emotional Early Childhood Scale (SEEC).* Circle Pines, MN: American Guidance Service.

Spector, P. (1992). *Summated rating scale construction.* Newbury Park, NJ: Sage University Press.

Squires, J., Bricker, D., Twombly, E., Yockelson, S., & Kim, Y. (1996). *Behavior-Ages & Stages Questionnaires.* Eugene: University of Oregon, Center on Human Development.

Swets, J.A., & Pickett, R.M. (1982). *Evaluation of diagnostic systems: Methods from signal detection theory.* San Diego: Academic Press.