

# C

## ASQ:SE-2 Technical Report



Jane Squires, Diane Bricker, Elizabeth Twombly,  
Kimberly Murphy, and Rob Hoselton,  
with contributions from Jill Dolata and Chieh-Yu Chen

This technical report focuses on the development and study of the psychometric properties of the *Ages & Stages Questionnaires®: Social-Emotional, Second Edition (ASQ:SE-2™)*, including revisions and additions to the first edition of ASQ:SE. Psychometric studies completed on ASQ:SE-2 are also described in detail, including results across the nine questionnaire intervals.

The first section addresses the development of the first edition, including item selection and revision, and is followed by a description of the iterative process used for developing, revising, and studying ASQ:SE-2. The next section describes ASQ:SE-2 data collection procedures, including participant recruitment, as well as procedures for the validity and reliability studies. Next, this technical report offers an overview of the demographic characteristics of the ASQ:SE-2 normative sample. The following section reports psychometric findings, including internal consistency, test–retest reliability, and convergent and known group validity. Finally, findings on the utility of ASQ:SE-2 are reported.

### DEVELOPMENT OF ASQ:SE

This section briefly describes the development of the first edition of ASQ:SE.

#### Item Selection

A developmental and behavioral lens was used to select items for ASQ:SE. The tool includes items that focus on acceptable, prosocial behaviors as well as problem behaviors, with attention paid to developmentally appropriate expectations for children. The knowledge that cultural and family factors as well as situational circumstances affect appropriateness of behaviors also guided our item selection.

ASQ:SE items initially were developed using a variety of sources, including standardized social-emotional and developmental assessments, research studies, textbooks and other resources in developmental and abnormal psychology, education and intervention resources, and language and communication materials. We ensured that items are as follows:

1. Representative of critical adaptive and maladaptive behaviors at targeted age intervals
2. Easy for parents/caregivers to understand and recognize
3. 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 24 hours) and concrete examples (e.g., smiles, kicks, bites other children) were provided to assist with interpretation of the item meanings.

## Iterative Item Development and Testing

Items were assembled into a field test version, which was titled the Behavior-Ages & Stages Questionnaires (B-ASQ; Bricker, Squires, 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 items regarding their appropriateness, ease of understanding, 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 families served by Healthy Start in Hawaii and Oregon; inner-city families in Cincinnati, Ohio, Portland, Oregon, and San Francisco, California; Head Start and Migrant Head Start families in California, Texas, and Washington; Child Development and Rehabilitation Center professors in Eugene, Oregon; and families with young children identified with social-emotional delays in Arizona, California, Oregon, Utah, and Washington. Utility surveys completed by service providers and parents offered feedback on the clarity of item meaning, appropriateness of items, missing content, and suggestions for revisions or additional items.

The B-ASQ was revised and renamed the Ages & Stages Questionnaires: Social-Emotional (ASQ:SE) based on the input gathered from experts, parents, and practitioners and on preliminary data analyses. Revisions included eliminating or combining items with overlapping and similar content; rewording items that were difficult to understand; adding items to fill content gaps (e.g., items were added to target so-called “red flags” for autism and adding a questionnaire for 60 months).

## DEVELOPMENT OF ASQ:SE-2

Beginning in late 2009, we began the process of revising and updating ASQ:SE based on data and comments from professionals and parents who had used the tool for more than 6 years. Early intervention program faculty, staff, and doctoral students again delved into an updated literature base on social-emotional development, including studies documenting the rise in the awareness and prevalence of autism spectrum and related disorders in young children, as well as the continuing and persistent problem with late identification of social-emotional

delays and disabilities in the preschool population. We began to develop and revise items for the second edition of ASQ:SE with a focus on capturing symptoms related to autism and communication disorders and developing additional items to identify competence and problem behaviors in infants and toddlers. Our main goals for the second edition revisions included the following:

1. Extend the age range of ASQ:SE in both directions—to include infants from 1 month to facilitate enrolling and monitoring children from birth, and to expand items to cover children’s behavioral repertoire to age 6 (i.e., 72 months) when children generally enter school settings
2. Develop new questionnaire items that directly target social-emotional, early communication, regulatory, and autism spectrum disorder (ASD) behaviors, and do so for children at a younger age
3. Update the format, including response options, to enhance the user friendliness of ASQ:SE
4. Update the normative sample to reflect the current, more diverse population of young children in the United States
5. Incorporate the addition of a monitoring zone similar to that used in ASQ-3 to aid in score interpretation and follow-up decision making

## **Extended Age Range**

Extending the ASQ:SE-2 age windows included testing and modifying the 6 month interval to include infants as young as 1 month of age, as well as expanding the upper age range of the 60 month interval to include children to age 6 years.

### ***Development of the 2 Month Age Interval***

The 2 month questionnaire was developed in response to requests from program personnel monitoring the development of newborns and older infants, who wanted an initial ASQ:SE to give parents soon after the birth of their children. With a 2 month questionnaire, professionals would be able to screen those infants whose parents had great concerns shortly after birth.

Analyses indicated that infants 1–2 months of age were inconsistently performing several items on the ASQ:SE-2 6 month questionnaire. A 2 month questionnaire for children from 1 to 3 months was developed and was composed of 16 scored items from the 6 month questionnaire, which were revised to be more developmentally appropriate. For example, “Does your baby let you know when she is hungry or sick?” was revised to “Does your baby let you know when she is hungry, tired, or uncomfortable? For example, does she fuss or cry?”

### ***Extension of the Upper Age Range to 72 Months***

The ASQ:SE-2 60 month questionnaire (with no additional items) was given to children 60–72 months of age to assist with screening children during kindergarten and entry to school. Analyses suggested that mean scores did not significantly differ between children 60–65 and 66–72 months of age. By extending the upper age range to 72 months, programs will be better able to continuously monitor children until their entry into first grade in public or private school settings, thus improving efforts to identify social-emotional and behavioral issues during children’s early school years.

## Item Additions

Additional ASQ:SE-2 items were formulated and field-tested in a series of eight iterative versions between 2010 and 2013. Items were added primarily to target early communication and behaviors associated with ASD, as well as internalizing behaviors in young children that might go undetected, such as excessive shyness and lack of reciprocity in social situations. We identified internalizing and communication behaviors and other more subtle behaviors associated with ASD in children 14 months of age and older, as these behaviors become easier to identify at this time. For infants, we looked at what behaviors parents would notice that are indicative of the communication, social, cognitive, and behavioral impairments at young ages (e.g., no back-and-forth sharing of sounds, no or few social smiles, no babbling or gesturing by 12 months) that might lead to early identification and improved outcomes.

Between three and seven additional items were added to each of the previously existing age intervals, and minor wording changes were made to existing ASQ:SE items. We made minor edits to existing ASQ:SE items based on item response modeling analyses and feedback from users. Examples of new items include the following:

- “Does your baby make sounds and look at you while playing with you?” (6 months)
- “When you copy sounds your baby makes, does your baby repeat the same sounds back to you?” (12 months)
- “Does your child play with objects by pretending? For example, does your child pretend to talk on the phone, feed a doll, or fly a toy airplane?” (18–30 months)
- “Is your child too worried or fearful?” (24–60 months)

New items are listed in Chapter 1, Table 1.1.

## Format Changes

Three major format changes were made to ASQ:SE-2—response options, scoring indicators, and the Information Summary sheet. First, we modified response options to make items easier for parents to interpret and respond. The previous response options of *most of the time*, *sometimes*, and *rarely or never* were changed to *often or always*, *sometimes*, and *rarely or never*. Both practitioners and parents felt that using *often or always* would assist in responding to items such as “Does your child cry, scream, or have tantrums for long periods of time?” The previous option (*most of the time*) was confusing for some parents, and focus groups of parents and practitioners felt that *often or always* was easier to understand.

## Scoring Indicators

Changes to streamline the scoring procedure were made to assist professionals in adding individual items’ scores (i.e., 15, 10, 5, 0) page by page before transferring the scores to the Information Summary sheet. Scoring lines were added in the right margin for each ASQ:SE-2 item as well as at the bottom of each page for total points per page, which can then be transferred to the Information Summary for calculating the child’s total score. In addition, a “V” was added next to each item concern circle to prompt professionals to add 5 points to any item score marked by the parent as a concern.

## Information Summary Sheet

The third format change was to revise the Information Summary sheet of each ASQ:SE-2 questionnaire so that it contains more information to summarize a child’s score, interpret results,

and indicate next steps for referral and evaluation. The Information Summary revisions were based on feedback from five focus groups with students, four focus groups with ASQ:SE users, as well as a final design review with professionals representing six programs with extensive experience using ASQ:SE. The revised Information Summary includes a scoring graphic illustrating the range of normative scores from 0% to 90% (the range of children's scores is too wide to accurately illustrate in the space on each questionnaire—often up to 350 points—so the 90th percentile score is indicated). The child's total score can be reviewed on the scoring graphic to get a visual representation of where the score falls on the distribution.

## Monitoring Zone

We developed a monitoring zone, or “questionable” area, just under the empirically derived cutoff score for each of the nine age intervals. The monitoring zone cutoff for each interval was set above the median at the 65th percentile, and each monitoring zone includes a range of 10–30 points below the at-risk cutoff score, depending on the specific age interval. The monitoring zone will identify an additional 14% of children across the ASQ:SE-2 intervals. It is depicted on the Information Summary sheet as a light gray area on the scoring graphic, as it is on ASQ-3.

This monitoring zone was included to assist programs in talking to parents about results and the relationship between the child's score and normative data, and to assist with decisions regarding referral options. Significant parent concerns paired with a child's total score in the monitoring zone may indicate a need for immediate referral to a mental health agency. See Chapter 6, Table 6.3, for additional information to guide professionals on aspects to consider regarding referral, including parent concerns.

## ASQ:SE-2 Pilot Version

Once initial content revisions were finalized in 2010, we began piloting an expanded version of ASQ:SE-2 to investigate 1) potential cutoff scores and psychometric properties with a new and larger normative population; 2) the accuracy of items extending the age range of ASQ:SE down to 1 month (from the original 3 months) and up to 72 months, or 6 years (from the original 66 months); and 3) how this expanded version identified children with and without ASD diagnoses. Parents of children from 1 month to 72 months of age were invited to complete the ASQ:SE-2 expanded pilot version, in both paper and electronic formats.

## DATA COLLECTION PROCEDURES

This section describes data collection procedures and iterative analyses and revisions.

### Participant Recruitment

Children between the ages of 1 and 72 months and their parents were recruited for the normative study. Approximately 10% were recruited through newspaper and magazine advertisements; 25% through agency personnel who attended national conferences and agreed to field-test ASQ:SE-2; and 65% through recruiting efforts via electronic bulletin boards and parenting web sites. Additional recruiting methods included posting advertisements on other web sites (e.g., Craigslist), social media, and in parenting resource guides; sending recruitment letters to child care providers in Oregon and California; and contacting personnel in agencies serving

high-risk families and young children with disabilities in several states (including California, Connecticut, Florida, Hawaii, Michigan, North Carolina, Ohio, Oregon, and Washington).

An attempt was made to stratify the normative sample so that children/families would be representative of the U.S. population in terms of race/ethnicity, geographic region, parent education and income, and gender of children. Recruitment letters and research protocols were approved by the University of Oregon Human Subjects Compliance Committee prior to beginning data collection. Data from the normative sample were first collected, followed by recruitment of samples for assessing convergent validity—the agreement of the classification outcomes of ASQ:SE-2 (i.e., risk, okay) with other screening and evaluation measures—and ASQ:SE-2's internal consistency, test–retest reliability, and interrater reliability.

## Measures

ASQ:SE-2 and a demographic form were the primary measures that were completed by parents. ASQ:SE-2 is a series of questionnaires at nine age intervals from 1 month to 72 months, as described previously. The questionnaires are designed to be completed by parents or other caregivers who can provide information on a child's social-emotional competence across situations.

The demographic form asked parents to provide information on the child's age, date of birth, gender, and race/ethnicity. Information about mother's age at child's birth, mother's education level, and family income also were requested.

## Convergent Validity

Several measures were used as comparators for ASQ:SE-2 outcomes to measure convergent validity. Due to the lack of “gold standard” assessments for the evaluation of general social-emotional difficulties in children from birth to 6 years (DelCarmen-Wiggins & Carter, 2004), we chose to compare ASQ:SE-2 outcomes with several evidence-based measures that focus on social-emotional outcomes in one or more age groups within the 1- to 72-month age range. We chose the term *convergent validity* to suggest that we were investigating the correspondence between ASQ:SE-2 and these measures rather than comparing ASQ:SE-2 with one gold standard of diagnosis. Convergent measures included the Devereux Early Childhood Assessment for Infants and Toddlers (DECA-IT; Mackrain, LeBuffe, & Powell, 2007), the Infant Toddler Social Emotional Assessment (ITSEA; Carter & Briggs-Gowan, 2006), and the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2000).

DECA-IT (Mackrain et al., 2007) is an assessment designed for parent or teacher completion and measures protective and risk factors for children's social-emotional development. Reliability studies by the DECA-IT authors indicated internal consistency ranged from .90 to .94, test–retest reliability from .83 to .94, and interrater reliability from .68 to .72. The Infant Interval, designed for children from birth to 18 months, was used with a subset of infants and toddlers in the ASQ:SE-2 normative sample between 1 and 13 months of age.

ITSEA (Carter & Briggs-Gowan, 2006), a standardized norm-referenced assessment with solid psychometric properties that evaluates social-emotional competence in young children from 1 year to 3 years of age (Printz, Borg, & Demaree, 2003), was used with a subset of children between 12 and 36 months in the ASQ:SE-2 normative sample. National standardization data suggested high internal consistency (the majority of Cronbach's alphas are above .70), acceptable test–retest reliability (intraclass correlations = .61–.91), evidence for concurrent validity (problem scores correlated significantly with Child Behavior Checklist 2/3;  $r = .28-.78$ ), and acceptable factor loading on the designated subscales (Carter & Briggs-Gowan, 2006).

Three subscales (i.e., Compliance, Negative Emotionality, and Prosocial Peer Relations) were used as comparators for ASQ:SE-2, based on recommendations from the ITSEA first author (Carter) as measures critical for social-emotional assessment (Beeber et al., 2007).

The CBCL (Achenbach & Rescola, 2000) was used as a convergent measure for a subset of children whose ages were in the 18 months to 5 years range in the ASQ:SE-2 normative sample. CBCL is a well-standardized and validated checklist that has been used widely with diverse populations. CBCL is completed by parents and/or teachers, has solid psychometric evidence, and contains approximately 100 items targeting problem behaviors in young children (Whitcomb & Merrell, 2013).

## Diagnostic Status

Additional children with a formal diagnosis of social-emotional disability, developmental disability, or ASD were recruited. Multidisciplinary teams in both educational and medical settings evaluated these children who received a diagnosis for eligibility for specialized early intervention/early childhood special education services. The disability status of these children was then compared with their overall classification on ASQ:SE-2 (at risk/okay) as a measure of convergent validity.

## Procedures with Parents

Parents willing to participate in the study were presented with a packet of materials containing a consent form, ASQ:SE-2 questionnaire, demographic form, and one or more convergent validity measures, depending on the child's age and the research phase. Packets were distributed in one of three ways: by mailing packets to parents (e.g., those who were contacted through advertisements, social network sites, online parenting sites) who agreed to participate; by preschool teachers directly to parents; and via the Internet through a research web portal. The forms were automatically scored for those using the portal, and the parents immediately received results online after completion. Parents were asked to complete paper research forms within 1 week and return them by mail to the researchers.

## Iterative Questionnaire Development

In order to develop ASQ:SE-2, an initial version was developed and distributed to parents during a 2-year period, between 2009 and 2011. Between 1 and 7 items were initially added to each ASQ:SE interval to test the ability of these new items to differentiate between children with social-emotional difficulties and those without. Data were collected in an ongoing fashion and analyzed every 2–3 months to test item functioning, utility, and agreement with convergent measures. Item response modeling and correlational and chi-square analyses were performed to test the functioning of the new items and their agreement with other measures and clusters of items. Every 5–6 months, focus groups of students, researchers, practitioners, and parents were assembled to review the ASQ:SE-2 questionnaires, make comments, suggest revisions, and evaluate utility. Eight different field test versions with differing numbers of new items were tested and analyzed; items that did not differentiate children with social-emotional difficulties were deleted from subsequent data analyses.

Out of the 20 new items tested, a total of 16 were added to ASQ:SE-2 across the 9 intervals. The final item set was developed in late 2012, at which time investigation of convergent validity and test–retest reliability outcomes were instigated.

## DEMOGRAPHIC CHARACTERISTICS OF NORMATIVE SAMPLE

Children between the ages of 1 month and 72 months were recruited to examine the psychometric properties of ASQ:SE-2. Data for demographic variables such as race/ethnicity, family income, and mother’s education level were not always provided by parents or primary caregivers for a variety of reasons (e.g., privacy, information unknown). The number of questionnaires completed without corresponding demographic information is noted for each analysis.

The ASQ:SE-2 total sample included 14,074 children. The total number of ASQ:SE-2 assessments completed on these children was 16,424. The distribution of these questionnaires by age interval and gender is shown in Table C.1. The majority of parents completed one ASQ:SE-2 interval; however, some parents completed more than one ASQ:SE-2 as their child aged (e.g., contributing questionnaires for the same child at 12, 24, and 36 months).

For those children born 3 or more weeks premature, an adjusted age was used to correct for prematurity up to 24 months of age. This adjusted age corresponded to the expected due date. Note that ASQ-3 uses the same procedure for determining which questionnaire a caregiver should complete for the child.

Table C.2 contains a comparison of U.S. Census Bureau estimates of the race/ethnicity distribution with those of the ASQ:SE-2 normative sample. This comparison is not straightforward, given that the U.S. Census did not use the category of “mixed” race/ethnicities on the 2010 Census and used a two-part question that separated categories of race from ethnicity (e.g., Hispanic or Latino/White not Hispanic or Latino).

According to data provided by the U.S. Census, the ASQ:SE-2 normative sample had a higher percentage of well-educated mothers than found generally in the United States (see Table C.3). Comparisons are not straightforward, however, given differing categories of analysis. A comparison between the U.S. Census data and the ASQ:SE-2 sample on annual family income level indicates the ASQ:SE-2 sample was composed of a similar percentage of families across the income levels (see Table C.4).

Data taken from the demographic form permitted dividing the ASQ:SE-2 normative sample into four groups according to the children’s developmental status in order to compare

**Table C.1.** Number of questionnaires and gender distribution by ASQ:SE-2 age interval<sup>a</sup>

ASQ:SE-2 age interval	Number of questionnaires		
	Total	Males	Females
2 month	287	148	139
6 month	2,042	1,066	976
12 month	2,274	1,278	996
18 month	2,214	1,329	885
24 month	1,808	1,101	707
30 month	1,509	892	617
36 month	2,221	1,290	931
48 month	2,523	1,456	1,067
60 month	1,516	900	616
Total	16,394	9,460	6,934

<sup>a</sup>There are gender data missing for 30 children.

**Table C.2.** Race/ethnicity comparison of ASQ:SE-2 normative sample ( $N = 14,137$ )<sup>a</sup> with 2010 U.S. Census Bureau estimates<sup>b</sup>

Race/ethnic category	Percentage		Difference
	ASQ:SE-2 normative sample	2010 U.S. Census estimate <sup>b</sup>	
White	72.9	72.4	+0.5
Black or African American	7.8	12.6	-4.8
Hispanic or Latino	7.3	N/A <sup>c</sup>	—
Asian	4.3	4.8	-0.5
Native American	0.8	0.9	-0.1
Native Hawaiian and other Pacific Islander	0.3	0.2	+0.1
Mixed race/ethnicity	6.6	N/A	—
Some other race	N/A	6.2	—
Two or more races	N/A	2.9	—

<sup>a</sup>There are race/ethnicity data missing for 2,287 cases.

<sup>b</sup>From U.S. Census Bureau (2010b).

<sup>c</sup>U.S. Census Bureau (2010b) reported Hispanic or Latino 16.3%, White Not Hispanic or Latino 83.7%; our demographic data were collected in different categories.

Note: The Pew Research Center report dated March 14, 2014, stated, "As many as 6.2% of census respondents selected only 'some other race' in the 2010 census, the vast majority of whom were Hispanic" (Krogstad & Cohn, 2014).

total scores across groups. We hypothesized that children with fewer risk factors would have lower scores on ASQ:SE-2, indicating social-emotional development in the typical range, and that as risk factors increased, ASQ:SE-2 scores would increase, indicating more potential social-emotional difficulties. The four groups are as follows: 1) No and Low Risk (i.e., children with no or one identified environmental/medical risk factor); 2) Risk (i.e., children with two or more environmental/medical 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

**Table C.3.** Education level comparison of ASQ:SE-2 normative sample ( $N = 14,290$ )<sup>a</sup> with 2010 U.S. Census Bureau estimates

Highest level of education	Percentage		Difference
	ASQ:SE-2 normative sample	2010 U.S. Census estimate <sup>b</sup>	
Less than high school diploma	3.7	12.0	-8.3
High school diploma	22.2	48.0	-25.8
Associate degree	14.0	10.0	+4.0
4-year college degree or above	56.7	30.0	+26.7
Do not know	3.3	— <sup>c</sup>	—

<sup>a</sup>There are level of education data missing for 2,134 caregivers.

<sup>b</sup>Based on U.S. Census Bureau (2010a).

<sup>c</sup>The U.S. 2010 Census Bureau does not include a "Do not know" category.

**Table C.4.** Income level comparison of ASQ:SE-2 normative sample ( $N = 14,647$ )<sup>a</sup> with 2010 U.S. Census Bureau estimates

Income category	ASQ:SE-2	2010 U.S. Census Bureau estimates		
	Percentage of ASQ:SE-2 normative sample	2010 U.S. Census income category <sup>b</sup>	Percentage of population	Difference
\$0–\$12,000	8.5	Less than \$9,999	7.8	+0.7
\$12,001–\$24,000	17.2	\$10,000–\$24,999	18.0	–0.8
\$24,001–\$40,000	14.2	\$25,000–\$39,999	15.9	–1.7
More than \$40,000	60.0	More than \$40,000	58.4	+1.6

<sup>a</sup>There are income level data missing for 1,777 children.

<sup>b</sup>Based on U.S. Census Bureau (2010a).

with identified social-emotional disabilities, according to IDEA Part B eligibility guidelines and *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition [DSM-IV]* diagnostic classifications). Variables used to determine level of risk for the Risk group included the following:

1. Family income less than \$12,000 per year
2. Mother younger than 18 years old when child was born
3. Mother’s highest level of education less than high school diploma
4. Involvement of child protective services with family; or a child in foster care
5. Child’s birth weight less than 3 pounds, 5 ounces
6. Child enrolled in Head Start or Early Head Start services

Children with developmental or social-emotional disabilities were classified in the Developmental Disability or Social-Emotional Disability group, regardless of the number of environmental risk factors. Table C.5 presents children in the normative sample by developmental status. (Developmental status information was missing for 6,039 children due to parents/primary caregivers choosing not to answer certain demographic questions; these were omitted from this analysis.)

## PSYCHOMETRIC FINDINGS

This section describes how the cutoff scores for ASQ:SE-2 were developed. In addition, this section presents data collected from subgroups of the normative sample, which were used to examine the internal consistency, test–retest reliability, convergent validity, known groups validity (including comparisons related to ASD measures), and utility of ASQ:SE-2.

### Establishing Reliability

This section explains how we established reliability, including internal consistency and test–retest reliability.

**Table C.5.** Number of questionnaires by developmental status for the ASQ:SE-2 normative sample ( $N = 10,385$ )<sup>a</sup>

ASQ:SE-2 age interval	N	Developmental status							
		No risk <sup>b</sup>		At risk <sup>c</sup>		Developmental disability <sup>d</sup>		Social-emotional disability <sup>e</sup>	
		n	Median	n	Median	n	Median	n	Median
2 month	160	125	20	34	10	1	50	—	—
6 month	1,121	1,010	20	103	10	8	60	—	—
12 month	1,279	1,148	30	103	35	25	65	3	155
18 month	1,309	1,133	36	110	35	55	95	11	115
24 month	1,127	953	35	108	35	51	60	15	135
30 month	937	766	50	88	45	69	80	14	132.5
36 month	1,514	1,043	60	311	40	104	90	56	157.5
48 month	1,876	1,079	55	627	35	98	77.5	72	145
60 month	1,062	659	50	307	35	43	95	53	165
Total/overall	10,385	7,916	40	1,791	35	454	80	224	150

<sup>a</sup>Developmental status data are missing for 6,039 children.

<sup>b</sup>No identified risk factors.

<sup>c</sup>One or more identified risk factors.

<sup>d</sup>Children receiving early intervention or early childhood special education services.

<sup>e</sup>Children with diagnosed social-emotional disabilities.

## Internal Consistency

Internal consistency measures the extent to which the different items on the assessment tool measure the same underlying construct (Saliva, Ysseldyke, & Bolt, 2012). 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-2 age interval using the variances of individual items and the variance of the total test scores ( $N = 11,489$ ). Cronbach's coefficient alphas for the ASQ:SE-2 intervals are shown in Table C.6. Alphas ranged from .71 to .90, with an overall alpha of .84. 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 over time. Test–retest reliability for ASQ:SE-2 was determined by comparing the results of two of the same questionnaires completed by parents at 1- to 3-week intervals. A random sample of parents ( $N = 281$ ) was asked to complete the same ASQ:SE-2 after returning the first completed questionnaire. The percent agreement between classifications of the child's performance on ASQ:SE at Time 1 (first questionnaire) and Time 2 (second questionnaire) was used to measure test–retest reliability.

Children were classified as Okay on ASQ:SE-2 if their scores were at or below the empirically derived cutoff point for that interval and no further evaluation of social-emotional competence was needed. Children were classified as At Risk on ASQ:SE-2 if their scores were above

**Table C.6.** Cronbach coefficient alpha by ASQ:SE-2 age interval (N = 11,489<sup>a</sup>)

ASQ:SE-2 age interval	Number of questionnaires	Alpha
2 month	270	.71
6 month	1,857	.74
12 month	1,980	.79
18 month	1,956	.87
24 month	1,581	.87
30 month	719	.88
36 month	1,095	.90
48 month	1,236	.90
60 month	795	.90
Overall	11,489	.84

<sup>a</sup>There are data missing for 4,935 children.

the cutoff point and further evaluation of their social-emotional status was needed. Test–retest agreement between classifications was 89% between Time 1 and Time 2 questionnaires. Intra-class correlations also were calculated and were .91 across intervals.

## Establishing Validity

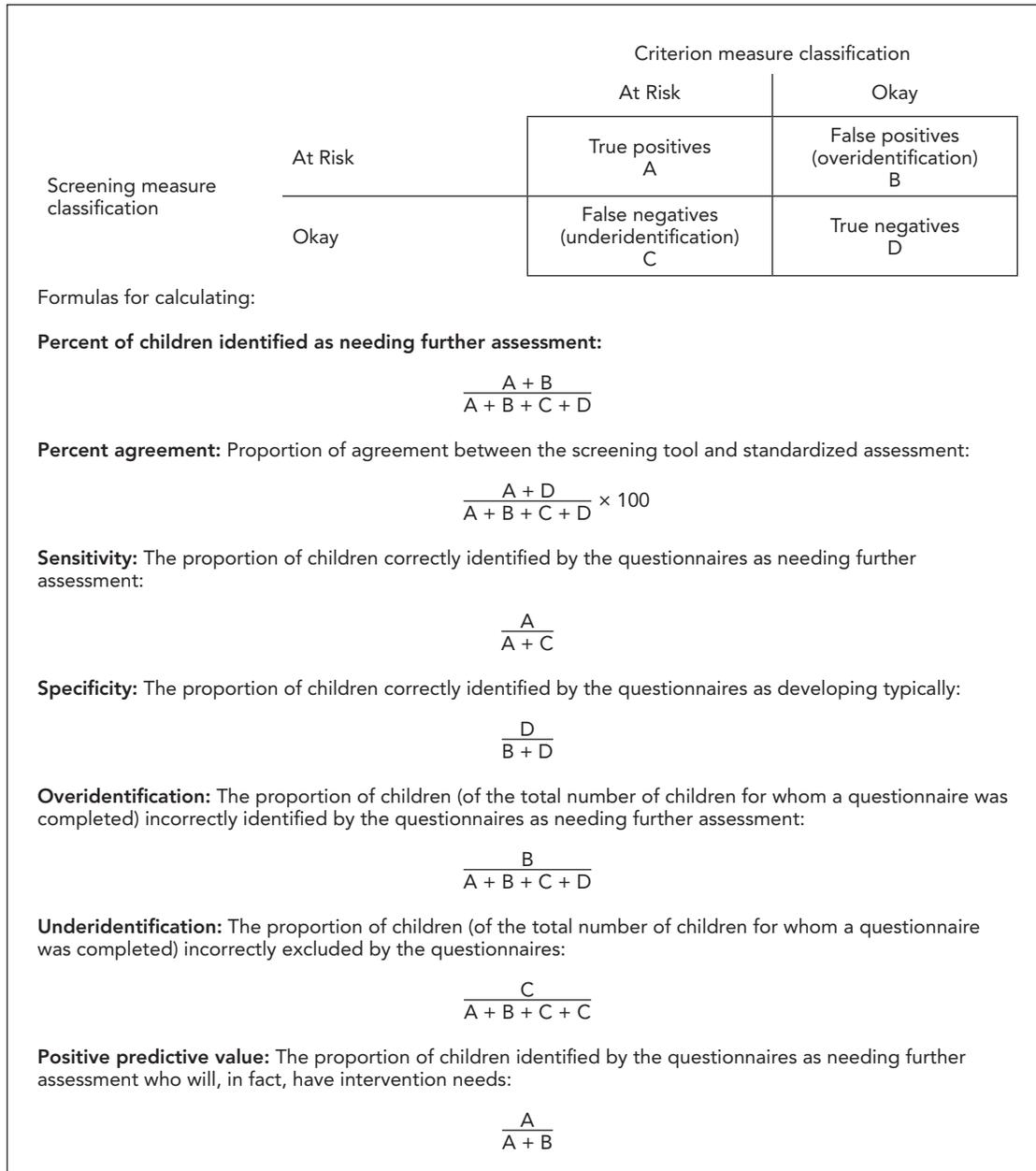
The primary goal of a screening measure is to accurately discriminate between individuals who are typical or okay (i.e., do not have the problem or characteristic) on a targeted variable (e.g., social-emotional development) and individuals who appear atypical or not okay (i.e., potentially may have the 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 whose scores are above the cutoff score are classified as At Risk and in need of follow-up, whereas individuals whose scores are below the cutoff score are classified as Okay and do not need follow-up.

There are no absolute scores that separate individuals who are typically developing from individuals who are not typically developing for any screening test. Rather, data must be collected and examined to determine the optimal cutoff 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 non-identification of individuals who should receive further testing (i.e., false negatives).

After tentative cutoff scores are selected, the second step is to determine their accuracy and, thus, the validity of the screening measure. Determining the accuracy of the cutoff scores is done by comparing children’s classification on the screening measure with their classification on a selected criterion or convergent measure(s). ASQ:SE-2 was examined by comparing children’s classification (i.e., developmentally Okay, At Risk) on ASQ:SE-2 with their classification (i.e., developmentally Okay, At Risk/Disabled) on selected convergent measures

that included the DECA-IT (Mackrain et al., 2007), ITSEA (Carter & Briggs-Gowan, 2006), CBCL (Achenbach & Rescola, 2000), and professional diagnosis of a social-emotional disability.

Receiver operating characteristic (ROC) curves were used to calculate optimal cutoff scores (i.e., those that yield high true positives, low false positives, and low false negatives for ASQ:SE-2). ROC analysis permits the systematic comparison of true positive probabilities against false positive probabilities for a range of possible cutoff scores (Swets & Pickett, 1982). A sample of 2,862 children with completed ASQ:SE-2 questionnaires were given a concurrent



**Figure C.1.** Contingency table for comparing screening measure classification with criteria measure classification and formulas for deriving comparison data.

criterion measure—DECA-IT (Mackrain et al., 2007), ITSEA (Carter & Briggs-Gowan, 2006), CBCL (Achenbach & Rescola, 2000)—or had a professional diagnosis of a social-emotional disability to create these comparison data. Each child’s classification (i.e., Okay, At Risk) on ASQ:SE-2 was then compared with the classification (i.e., Okay, At Risk/Disabled) of the child by one of the criterion measures. Figure C.1 shows a four-cell contingency table used to assess the agreement between the screening measure (i.e., ASQ:SE-2) and the criterion measure (i.e., DECA-IT, ITSEA, CBCL, diagnosis of social-emotional disability). In addition, this figure shows the formulas for calculating percentage of children identified as needing further assessment and the percent agreement, sensitivity, specificity, overidentification, underidentification, and positive predictive value of ASQ:SE-2.

Comparison of range of scored points, medians, interquartile ranges, and ROC cutoffs is shown in Table C.7. Note that ROC cutoff scores for most 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 median scores, except at 48 and 60 months.

Cutoff scores for screening tools frequently 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. Using means to calculate cutoff scores presumes a normal distribution of scores, however. Score distribution for ASQ:SE-2 questionnaires was positively skewed—that is, the majority of children obtained low scores (i.e., indicating no problem or Okay) and relatively few children obtained high scores (i.e., indicating a potential problem or risk). The majority of children had scores below 25. Figure C.2 shows the positively skewed distribution of scores for the 48 month ASQ:SE-2; 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-2 scores across intervals. Instead, ROC analyses were conducted to determine the best cutoff point for each interval.

### **Modifying Cutoff Scores**

If programs want to modify cutoff scores, then semi-interquartile ranges [i.e., median + (Quartile 1 – Quartile 3)/2] should be used as the basis for modification. See [www.agesandstages.com](http://www.agesandstages.com) for additional discussion of guidelines for altering ASQ:SE-2 cutoff points.

### **Monitoring Zone**

A monitoring zone has been designated to assist with ASQ:SE-2 score interpretation, as discussed in Chapter 6. The monitoring zone begins at the 65th percentile and ends at the referral cutoff. The monitoring zone identifies a group of children whose scores are close to the cutoff score. The range of points for each questionnaire’s monitoring zone can be found in Table C.8.

### **Gender Differences**

Scores for males and females were compared in order to examine gender differences. Mean and median scores by gender are presented in Table C.9. Box plots were then derived to examine the score distributions by gender. Box plots provide a visual picture of a distribution. Box plots for the 30 month and 36 month ASQ:SE-2 male and female score distributions are shown in Figure C.3. 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

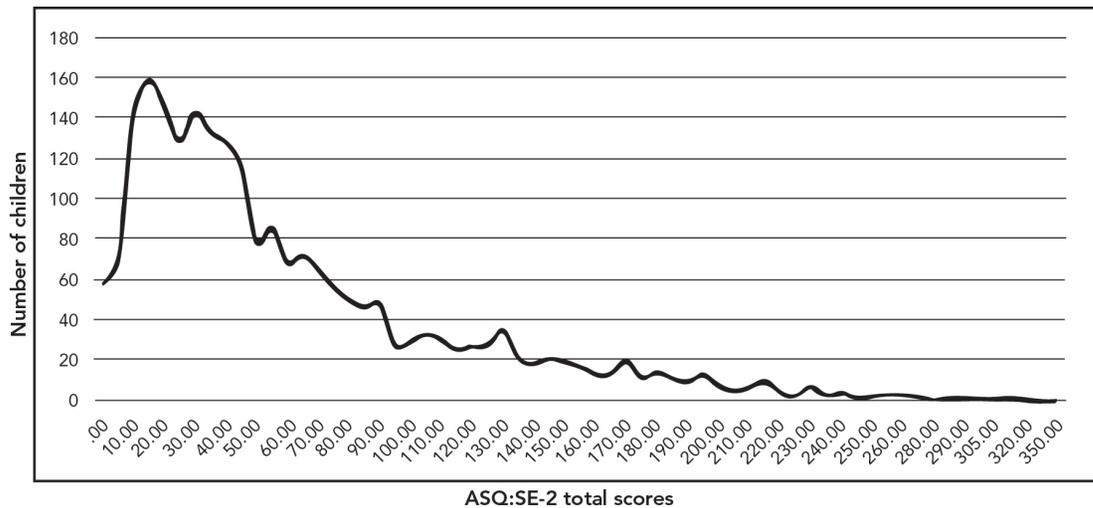
**Table C.7.** Range of points, medians, interquartile ranges, receiver operating characteristic (ROC) cutoffs, and percentages identified by ASQ:SE-2 interval (N = 16,424)<sup>a</sup>

ASQ:SE-2 age interval	N	Range of points	Median	Median + 1.5 semi-interquartile ranges	ROC cutoff score <sup>a</sup>	Percent identified
2 month	288	0–140	20	35.0	35	18.4
6 month	2,043	0–210	20	38.8	45	14.7
12 month	2,276	0–245	30	52.5	50	24.9
18 month	2,216	0–405	35	68.8	65	23.3
24 month	1,809	0–335	35	66.5	65	22.8
30 month	1,514	0–375	45	90.0	85	23.7
36 month	2,232	0–370	55	106.0	105	22.0
48 month	2,530	0–350	45	93.8	85	26.0
60 month	1,516	0–310	46	102.3	95	23.7

<sup>a</sup>ROC cutoff based on “best fit,” maximizing true positives and true negatives.

mean is indicated by the “x.” 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 “\*.”

As shown in Figure C.3, the majority of scores for males at the 30 month interval range between 25 and 75, with the 1.5 interquartile ranges extending to 180. Outliers extend upward to 370. For females, the range is between 20 and 61, with the 1.5 interquartile ranges extending to 150. Outliers extend upward to 325. A similar distribution for both males and females can be seen at the 36 month interval. Similar distribution patterns occurred at all test intervals and indicate, in general, that males tended to have greater dispersal of scores and more extreme scores.



**Figure C.2.** ASQ:SE-2 total scores by number of children for the 48 month questionnaire (N = 2,530), showing a positively skewed distribution.

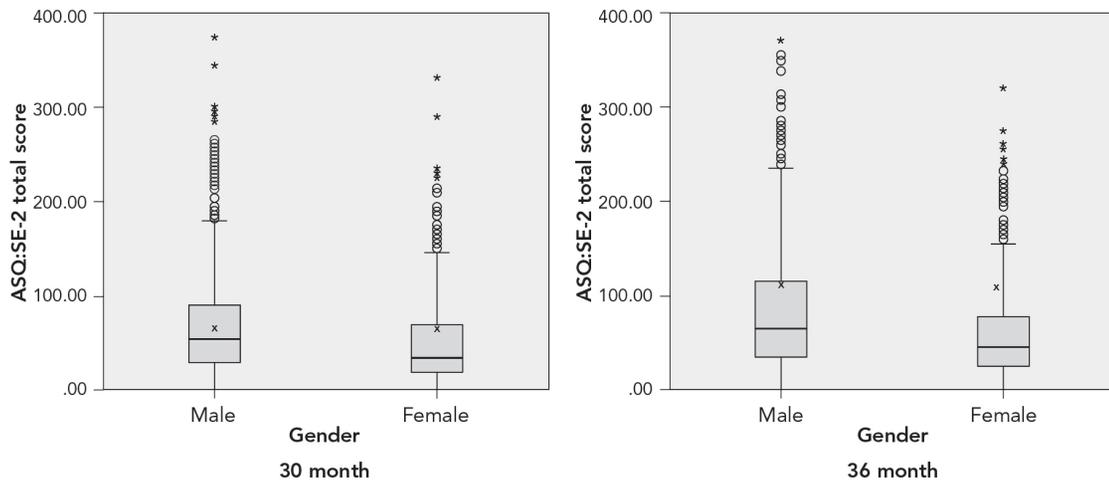
**Table C.8.** ASQ:SE-2 monitoring zone score ranges

ASQ:SE-2 interval	Monitoring zone
2 month	25–35
6 month	30–45
12 month	40–50
18 month	50–65
24 month	50–65
30 month	65–85
36 month	75–105
48 month	70–85
60 month	70–95

If gender group differences are compared using nonparametric tests (i.e., Kruskal-Wallis Test [Heiman, 1992]), significant differences are found at 12, 18, 24, 30, 36, 48, and 60 months, as shown in Table C.9. It is important to note that the validity sample did not have adequate numbers of girls identified with social-emotional problems to determine if separate cutoff scores for females are needed, especially at younger ages.

Gender differences also were compared using item response modeling (IRT), specifically differential item functioning (DIF). Differences between scores of boys and girls were analyzed across the nine ASQ:SE-2 age intervals using the Mantel (1963) approach (Dorans & Holland, 1993); 22 out of 295 items were identified with DIF or differential functioning for boys and girls. The majority of those 22 items were revised to include examples and/or modified wording.

Among all DIF items, one specific item across three age intervals was identified, potentially causing different responses based on the gender of the child being assessed. This item was “Does your child play with objects by pretending? For example by feeding a doll or talking on



**Figure C.3.** Box plots for 30 month and 36 month ASQ:SE-2 by gender. Box plots illustrate the distribution of scores. 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. Mean is indicated by an “x.” 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 “\*.” Box width varies by *n*. (Note: 30 month: male = 892, female = 617; 36 month: male = 1,290, female = 931).

**Table C.9.** ASQ:SE-2 means and medians for males and females by age interval ( $N = 16,394$ )<sup>a</sup>

ASQ:SE-2 age interval	Male			Female		
	N	Mean	Median	N	Mean	Median
2 month	148	26.0	20	139	22.8	20
6 month	1,066	28.1	21	976	26.6	20
12 month	1,278	42.8	35	996	34.6*	30
18 month	1,329	51.4	40	885	44.1*	30
24 month	1,101	53.4	40	707	41.8*	30
30 month	892	69.5	55	617	52.3*	35
36 month	1,290	81.3	65	931	59.2*	45
48 month	1,456	75.1	55	1,067	52.5*	35
60 month	900	70.7	51	616	58.7*	40
Total	9,460	—	—	6,934	—	—

<sup>a</sup>Gender data are missing for 30 children.

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

the phone?” Based on the analysis, we revised this item for subsequent versions to “Does your child play with objects by pretending? For example by talking on the phone, feeding a doll, or flying a toy airplane?” The order of items within the questionnaires was changed so that they followed the developmental order identified by the IRT analysis. Second, and perhaps most important, we included additional examples that might have more appeal to boys, with materials such as balls.

Furthermore, items with differential functioning indicated specific patterns in social-emotional competence areas across intervals by gender. For example, the scores of boys indicated a greater frequency of problem behaviors than girls on several items in social-communication and self-regulation areas; girls’ scores indicated a greater frequency of problem behaviors than boys on items assessing sleeping and autonomy. The different patterns of scoring may have resulted from potential differences between genders or the parents’ varying expectations for boys and girls.

As additional data are added to the validity sample, revised cutoffs, if necessary, will be shared via [www.agesandstages.com](http://www.agesandstages.com). Based on our current normative sample, however, we recommend that programs carefully review with parents/caregivers ASQ:SE-2 results of girls with scores in the monitoring zone. Because girls in our normative sample in general had scores lower than boys, especially at 12 through 60 months of age, we believe that girls and their families may need to be considered for referral for further assessment and/or community services when their scores are near as well as above the screening cutoff points.

## Examining Convergent Validity

A comparison with selected social-emotional measures was necessary to determine how accurately ASQ:SE-2 discriminates between children whose social-emotional development is proceeding without a problem from those children who have or are at risk for developing a social-emotional problem. DECA-IT (Mackrain et al., 2007), ITSEA (Carter & Briggs-Gowan, 2006), and CBCL (Achenbach & Rescola, 2000) were convergent measures chosen to examine the validity of ASQ:SE-2.

Parents or other primary caregivers of children in the validity sample ( $N = 2,862$ ) completed the DECA-IT (Mackrain et al., 2007), ITSEA (Carter & Briggs-Gowan, 2006), and/or CBCL (Achenbach & Rescola, 2000) (depending on the age of their child) within 2–3 weeks of also completing ASQ:SE-2. Included in the validity sample was a group of 146 children ranging in age from 6 to 60 months who had been professionally diagnosed as having a social-emotional disability and were receiving intervention services. Each diagnosis was based on a multidisciplinary battery of diagnostic assessments. Table C.10 summarizes the numbers of children by convergent measure.

Children in the validity sample were classified as either Okay or At Risk based on their ASQ:SE-2 score using the established cutoffs and were independently classified as either Okay or At Risk/Disabled using their score on the DECA-IT (Mackrain et al., 2007), ITSEA (Carter & Briggs-Gowan, 2006), and/or CBCL (Achenbach & Rescola, 2000) or professional diagnosis. These two independent classifications were then compared for all children in the validity sample. One of four outcomes was possible: 1) ASQ:SE-2 and the criterion measure both classified the child as Okay (i.e., true negatives), 2) ASQ:SE-2 and the criterion measure both classified the child as At Risk/Disabled (i.e., true positives), 3) ASQ:SE-2 classified the child as Okay while the criterion measure classified the child as At Risk/Disabled (i.e., false negatives), and 4) ASQ:SE-2 classified the child as At Risk, whereas the criterion measure classified the child as Okay (i.e., false positives).

Contingency tables containing four cells (i.e., A = true positives, B = false positives, C = false negatives, and D = true negatives, as shown in Figure C.1) were developed for each ASQ:SE-2 age interval using the ROC cutoff scores (see Table C.7) to conduct these comparisons. Each contingency table contained in Figure C.4 shows the agreement for true positives, false positives, false negatives, and true negatives. Using the data contained in the contingency table, the sensitivity, specificity, false positive rate, false negative rate, percent agreement, underidentified rate, overidentified rate, and positive predictive value were calculated for each ASQ:SE-2 age interval. An overall comparison across all intervals is shown in Figure C.5.

**Table C.10.** Measures and  $N$  used for ASQ:SE-2 convergent validity analyses

ASQ:SE-2 interval	Autism diagnosis	CBCL	DECA-IT	ITSEA	Social-emotional diagnosis	Total
2 month	—	—	71	—	—	71
6 month	—	—	438	—	2	440
12 month	1	—	264	232	3	500
18 month	4	16	9	458	13	500
24 month	9	52	—	336	11	408
30 month	7	45	—	297	11	360
36 month	35	83	—	103	42	263
48 month	48	116	—	—	37	201
60 month	29	63	—	—	27	119
Total	133	375	782	1,426	146	2,862

Key: CBCL, Child Behavior Checklist (Achenbach & Rescola, 2000); DECA-IT, Devereux Early Childhood Assessment-Infant/Toddler (Mackrain, LeBuffe, & Powell, 2007); ITSEA, Infant Toddler Social Emotional Assessment (Carter & Briggs-Gowan, 2006).

Table C.11 presents a comparison of the sensitivity, specificity, percent agreement, false positive rate, false negative rate, underidentified rate, and overidentified rate across ASQ:SE-2 age intervals. Sensitivity ranged from a low of 77.8% at 2 months to a high of 84% at 24 months. Specificity ranged from 76.2% at 18 months to 98% at 60 months. Percent agreement ranged from 77% at 18 months to 89.1% at 60 months. The underidentified rate ranged from 1.8% at 6 months to 10.9% at 48 months, whereas the overidentified rate ranged from .8% at 60 months to 19.4% at 12 months. These findings suggest that ASQ:SE-2 is generally accurate in discriminating between children who are Okay and those who need follow-up. Because of the overidentification rate at 12 months (19.4%) and 18 months (19.0%), programs should consider rescreening in follow-up and referral actions.

		2 month		Criterion measure classification <sup>a</sup>				
				At risk	Okay	Total		
ASQ:SE-2 classification	At risk	7	12			19		
	Okay	2	50			52		
	Total	9	62			71		

Sensitivity	Specificity	False positive	False negative	Percent agreement	Under-identified	Over-identified	Positive predictive value
77.8%	80.6%	19.4%	22.2%	80.3%	2.8%	16.9%	36.8%

		6 month		Criterion measure classification <sup>b</sup>				
				At risk	Okay	Total		
ASQ:SE-2 classification	At risk	29	47			76		
	Okay	8	356			364		
	Total	37	403			440		

Sensitivity	Specificity	False positive	False negative	Percent agreement	Under-identified	Over-identified	Positive predictive value
78.4%	88.3%	11.7%	21.6%	87.5%	1.8%	10.7%	38.2%

**Figure C.4.** Contingency tables showing agreement between ASQ:SE-2 classification, criterion measure classification, and ASQ:SE-2 sensitivity, specificity, false positive rate, false negative rate, percent agreement, percent underidentified, percent overidentified, and positive predictive value by age interval (definitions and formulas are contained in Figure C.1). Criterion measure classification includes CBCL, DECA-IT, ITSEA, and professional diagnoses. (Key: <sup>a</sup>Includes DECA-IT; <sup>b</sup>Includes DECA-IT and professional diagnosis; <sup>c</sup>Includes DECA-IT, ITSEA, and professional diagnosis; <sup>d</sup>Includes DECA-IT, CBCL, ITSEA, and professional diagnosis; <sup>e</sup>Includes CBCL, ITSEA, and professional diagnosis; <sup>f</sup>Includes CBCL and professional diagnosis.)

(continued)

Figure C.4. (continued)

12 month		Criterion measure classification <sup>c</sup>			Total		
		At risk	Okay				
ASQ:SE-2 classification	At risk	40	97		137		
	Okay	11	352		363		
Total		51	449		500		
Sensitivity	Specificity	False positive	False negative	Percent agreement	Under-identified	Over-identified	Positive predictive value
78.4%	78.4%	21.6%	21.6%	78.4%	2.2%	19.4%	29.2%

18 month		Criterion measure classification <sup>d</sup>			Total		
		At risk	Okay				
ASQ:SE-2 classification	At risk	81	95		176		
	Okay	20	304		324		
Total		101	399		500		
Sensitivity	Specificity	False positive	False negative	Percent agreement	Under-identified	Over-identified	Positive predictive value
80.2%	76.2%	23.8%	19.8%	77.0%	4.0%	19.0%	46.0%

24 month		Criterion measure classification <sup>e</sup>			Total		
		At risk	Okay				
ASQ:SE-2 classification	At risk	63	39		102		
	Okay	12	294		306		
Total		75	333		408		
Sensitivity	Specificity	False positive	False negative	Percent agreement	Under-identified	Over-identified	Positive predictive value
84.0%	88.3%	11.7%	16.0%	87.5%	2.9%	9.6%	61.8%

30 month		Criterion measure classification <sup>e</sup>			Total		
		At risk	Okay				
ASQ:SE-2 classification	At risk	47	44		91		
	Okay	12	257		269		
Total		59	301		360		
Sensitivity	Specificity	False positive	False negative	Percent agreement	Under-identified	Over-identified	Positive predictive value
79.7%	85.4%	14.6%	20.3%	84.4%	3.3%	12.2%	51.6%

36 month		Criterion measure classification <sup>e</sup>		
		At risk	Okay	Total
ASQ:SE-2 classification	At risk	98	19	117
	Okay	27	119	146
Total		125	138	263

Sensitivity	Specificity	False positive	False negative	Percent agreement	Under-identified	Over-identified	Positive predictive value
78.4%	86.2%	13.8%	21.6%	82.5%	10.3%	7.2%	83.8%

48 month		Criterion measure classification <sup>f</sup>		
		At risk	Okay	Total
ASQ:SE-2 classification	At risk	105	11	116
	Okay	22	63	85
Total		127	74	201

Sensitivity	Specificity	False positive	False negative	Percent agreement	Under-identified	Over-identified	Positive predictive value
82.7%	85.1%	14.9%	17.3%	83.6%	10.9%	5.5%	90.5%

60 month		Criterion measure classification <sup>f</sup>		
		At risk	Okay	Total
ASQ:SE-2 classification	At risk	57	1	58
	Okay	12	49	61
Total		69	50	119

Sensitivity	Specificity	False positive	False negative	Percent agreement	Under-identified	Over-identified	Positive predictive value
82.6%	98.0%	2.0%	17.4%	89.1%	10.1%	0.8%	98.3%

## Known Groups Validity

This section describes the validity of known groups: risk and disability status, and ASD.

### Risk and Disability Status

Examining the differences in scores across groups is another approach to assessing the validity of a screening measure suggested by Spector (1992). For this analysis, a subsample of children in the validity sample were divided into one of three groups based on developmental status—No Risk, Developmentally Disabled, and Socially-Emotionally Disabled. Children were assigned to the

Overall	Criterion measure classification <sup>a</sup>			
		Risk	Okay	
ASQ:SE-2 classification	Risk	527	365	892
	Okay	126	1,844	1,970
Total		653	2,209	2,862

Sensitivity	Specificity	False positive	False negative	Percent agreement	Under-identified	Over-identified	Positive predictive value
80.7%	83.5%	16.5%	19.3%	82.8%	4.4%	12.8%	59.1%

<sup>a</sup>Includes CBCL, DECA-IT, ITSEA, and professional diagnosis.

**Figure C.5.** Contingency table showing overall agreement (combined across age intervals) between ASQ:SE-2 classification with criterion measure classification and ASQ:SE-2 sensitivity, specificity, false positive rate, false negative rate, percent agreement, percent underidentified, percent overidentified, and positive predictive value (definitions and formulas are contained in Figure C.1). Criterion measure classification includes CBCL, DECA-IT, ITSEA, and professional diagnosis.

No Risk group if caregivers reported no risk factors ( $N = 7,916$ ); children were assigned to the Developmentally Disabled group if they were receiving general early intervention services ( $N = 454$ ); and children were assigned to the Socially-Emotionally Disabled group if they had been diagnosed with a behavior or emotional problem and were receiving intervention services ( $N = 224$ ). Risk factors included the following: 1) family income less than \$12,000; 2) mother younger than 18 years old when child was born; 3) mother's level of education less than high school diploma; 4) involvement of child protective services with family or a child in foster care; 5) birth weight less than 3 pounds, 5 ounces; and 6) child enrolled in Early Head Start or Head Start services.

Figure C.6 presents the median scores for the four groups by interval. Score profiles by risk group indicate that ASQ:SE-2 can discriminate between children whose social-emotional development is typical and those who have disabilities. An example of box plots showing the

**Table C.11.** ASQ:SE-2 cutoff scores and classification statistics<sup>a</sup> by age interval based on receiver operating characteristic (ROC) cutoff score ( $N = 2,862$ )

ASQ:SE-2 age interval	N	Cutoff score	Sensitivity	Specificity	Percent agreement	False positive rate	False negative rate	Under-identified	Over-identified
2 month	71	35	77.8	80.6	80.3	19.4	22.2	2.8	16.9
6 month	440	45	78.4	88.3	87.5	11.7	21.6	1.8	10.7
12 month	500	50	78.4	78.4	78.4	21.6	21.6	2.2	19.4
18 month	500	65	80.2	76.2	77.0	23.8	19.8	4.0	19.0
24 month	408	65	84.0	88.3	87.5	11.7	16.0	2.9	9.6
30 month	360	85	79.7	85.4	84.4	14.6	20.3	3.3	12.2
36 month	263	105	78.4	86.2	82.5	13.8	21.6	10.3	7.2
48 month	201	85	82.7	85.1	83.6	14.9	17.3	10.9	5.5
60 month	119	95	82.6	98.0	89.1	2.0	17.4	10.1	0.8
Overall	2,862	—	80.7	83.5	82.8	16.5	19.3	4.4	12.8

<sup>a</sup>See Figure C.1 for formula used in calculating classification statistics.

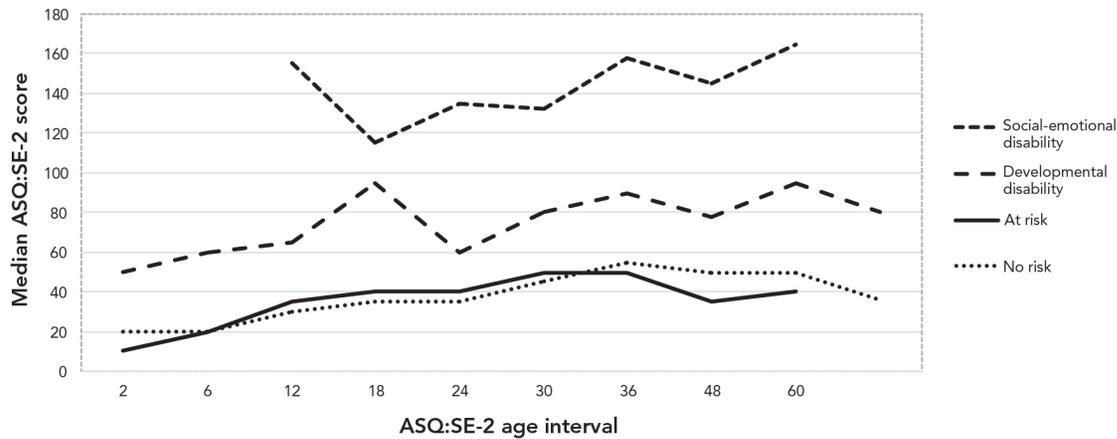


Figure C.6. Median ASQ:SE-2 scores by group risk status.

distribution of risk groups for the 48 month ASQ:SE-2 is presented in Figure C.7. The box plots show mean (marked with a “x”) and median (middle horizontal line in each box) scores are elevated for children with developmental and social-emotional disabilities. 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. There is an overlap in scores for the Risk and No Risk groups, possibly due to a selection factor involving a large number of parents with concerns for their child’s social-emotional development who completed the ASQ:SE-2 online. The majority of these families experienced low or no risk. The No Risk and Low Risk groups could be collapsed for future analyses.

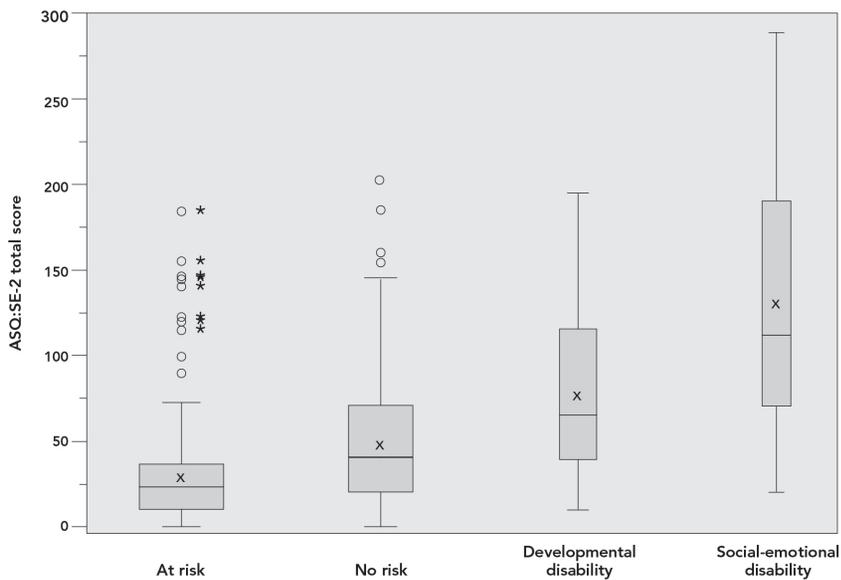
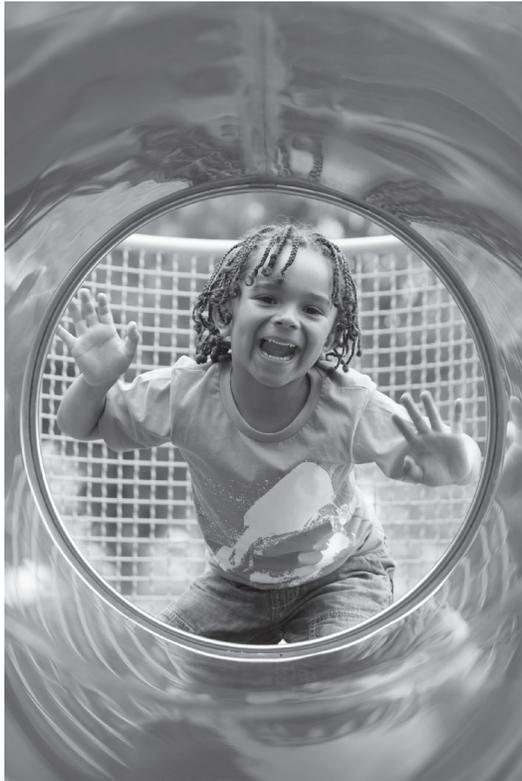


Figure C.7. Box plot distribution of total scores on the 48 month ASQ:SE-2 by developmental status. Box plots illustrate the spread of distribution. The bottom line of the box is the 25th percentile, or Quartile 1. The top line of the box is the 75th percentile, or Quartile 3. The middle line is the median, or Quartile 2. Mean is indicated by a “x.” Whiskers (lines) extend to the highest and lowest observations but not further than 1.5 interquartile ranges. Outliers beyond 1.5 interquartile range are indicated by “O.” Outliers beyond 3 interquartile ranges are indicated by “\*.” Box width varies with *n*.

## Autism Spectrum Disorder

For comparisons related to ASD, the classification of children on ASQ:SE-2 was compared for 133 children between 18 and 60 months of age with a diagnosis of ASD. Overall, the ASQ:SE-2 classification (Okay, Risk) agreed 83.5% with the child's classification of ASD (Risk). These children had received a diagnosis of ASD through educational or medical diagnostic procedures, and their parents had completed an ASQ:SE-2 questionnaire.

These preliminary analyses suggest that ASQ:SE-2 will identify the majority of children with ASD. Ongoing research is being conducted in which in-depth ASD measures such as the Autism Diagnostic Observation Schedule, Second Edition (Lord, Rutter, DiLavore, & Risi, 2012) are used concurrently with ASQ:SE-2. These results may assist programs in making decisions regarding referral for ASD testing based on ASQ:SE-2 results. In addition, a summary of ASQ:SE-2 and ASQ-3 items related to behaviors associated with ASD appears in Table C.12. ASQ-3 items pertaining to autism are included to provide a better understanding of how to use the tools together.



## UTILITY SURVEY

Utility of a screening tool measures the usefulness or practicality of the test or procedure (Bricker & Squires, 1989). A random sample of parents ( $N = 332$ ) who completed a paper version of ASQ:SE-2 was asked to complete a utility survey that requested their opinion about the length, appropriateness, and ease of completion of ASQ:SE-2. Summary results of the utility survey can be found in Table C.13.

Thirty-one percent of the respondents indicated that it took less than 10 minutes to complete; thirty-nine percent indicated that it took 10–20 minutes to complete. Ninety-three percent of the respondents indicated that ASQ:SE-2 was easy to understand, and eighty percent felt that question content was appropriate for their child. Seventy-three percent reported that ASQ:SE-2 helped them think about their child's behavior, and ninety-four percent said that they would like to fill out another questionnaire when their child is older. Thus, parents reported that ASQ:SE-2 was easy to understand, it took little time to complete, questions were appropriate, and they enjoyed the process. In addition, parents felt that completing ASQ:SE-2 was interesting and helped them think about the social-emotional development of their children.

## SUMMARY

Psychometric studies on ASQ:SE-2 are summarized in this technical report. Normative data are based on 16,424 completed questionnaires. Validity studies were conducted using 2,862

**Table C.12.** ASQ:SE-2 and ASQ-3 items related to autism spectrum disorder (ASD)

Item content (not exact wording)	ASD related skill/indicator	Intervals for 12–24 months	Intervals for 25–60 months
When you hold out your hand and ask for her toy, does your baby let go of it into your hand?*	Social reciprocity, social-communication, restricted/repetitive patterns	12	—
Does your baby like to play games such as Peekaboo?	Social reciprocity	12*	—
Does your baby look for you when a stranger approaches?	Social reciprocity	12, 18	—
When you talk to your baby, does he turn his head, look, or smile?	Social reciprocity	12, 18	—
Does your child respond to her name when you call her?	Social-communication	12, 18, 24	—
When you point at something, does your child look in the direction you are pointing?	Social reciprocity	12, 18, 24	30
Does your child try to show you things? (with point and check-in at later intervals)	Social reciprocity	12, 18, 24	30, 36
Does your child play with objects by pretending? (symbolic in later intervals)*	Social reciprocity	12**, 18, 24	30, 36
Does your baby roll or throw a ball back to you so you can return it to him?*	Social reciprocity	12, 14	—
Does your baby shake his head when he means “no” or “yes?”**	Social-communication	14	—
After you have shown your baby how, does he try to get a toy using a tool?*	Social reciprocity	14, 16, 18	—
Does your child come to you when she needs help?*	Social-communication	16, 18	—
Does your child look at you when you talk to him?	Social-communication	18, 24	30, 36, 48, 60
Does your child do things over and over and get upset when you try to stop her?	Restricted/repetitive patterns	18, 24	30, 36, 48, 60
Does your child let you know how she is feeling with gestures or words?	Social-communication	18, 24	30, 36, 48, 60
Does your child check to make sure you are near when exploring new places?	Social reciprocity	18, 24	30, 36
Does your child like to be around other children?	Social reciprocity	18, 24	—
Does your child copy specific gestures?*	Social reciprocity	20, 22	27, 30
Does your child copy or imitate you and line up blocks? (creating a bridge at later interval)**	Social reciprocity	20, 22, 24	27, 30, 33, 36, 42
Does he put a box on his head, pretending it is a hat?*	Social reciprocity	22, 24	27
Does your child correctly use at least two words such as <i>me</i> and <i>you</i> ?*	Social-communication	22, 24	27

(continued)

**Table C.12.** (continued)

Item content (not exact wording)	ASD related skill/indicator	Intervals for 12–24 months	Intervals for 25–60 months
Does your child greet or say hello to familiar adults?	Social-communication	24	30
After your child watches you draw a line/circle, does your child copy your shape?***	Social reciprocity	—	27, 30, 33, 36, 42
When you point to the figure and ask your child, “What is this?” does your child say a word that means a person?***	Social reciprocity	—	27, 30, 33, 36, 42
Does your child do what you ask him to do?	Social reciprocity	—	30, 36, 48, 60
Does your child move from one activity to the next with little difficulty?	Restricted/repetitive patterns	—	30, 36, 48, 60
Does your child take turns and share when playing with an adult (other children at later intervals)?*	Social reciprocity	—	36**, 42**, 60
Can your child name a friend?*	Social reciprocity	—	36, 48, 54*
Do other children like to play with your child?	Social reciprocity	—	36, 48, 60
Does your child like to play with other children?	Social reciprocity	—	36, 48, 60
Does your child follow rules?	Restricted/repetitive patterns	—	48, 60
Does your child show concern for other people’s feelings?	Social reciprocity	—	48, 60
Does your child have simple conversations with you?	Social reciprocity	—	48, 60
Can your child draw a picture of a person/girl/boy?***	Social reciprocity	—	54, 60

Note: No asterisk denotes ASQ:SE-2 item only. \*Denotes ASQ-3 and ASQ:SE-2 item. \*\*Denotes ASQ-3 only item. Wording may differ slightly between intervals and across measures. See questionnaires for exact wording.

children. Internal consistency measured by coefficient alpha was found to be high across intervals, ranging from .71 to .90, with an overall alpha of .84. Test–retest reliability, measured as the agreement between two ASQ:SE-2 questionnaires completed by parents at 1 day to 60 days apart was .89. Sensitivity ranged from .78 at 2 months to .84 at 24 months, with .81 overall sensitivity. Specificity of the questionnaires ranged from .76 at 18 months to .98 at 60 months, with .84 overall. Percent agreement between questionnaires and standardized assessments/disability status ranged from .77 at 18 months to .89 at 60 months with overall agreement at .83. The ability of ASQ:SE-2 to detect atypical social-emotional development (sensitivity) was generally less than specificity (i.e., the ability of ASQ:SE-2 to correctly identify typically developing children).

Research is continuing on ASQ:SE-2. Results of ASQ:SE-2 questionnaires completed by parents are being compared with in-depth ASD diagnostic evaluations to study the sensitivity and specificity of ASQ:SE-2 related to identifying young children with ASD. Additional research findings will be posted on the ASQ web site as they become available ([www.agesandstages.com](http://www.agesandstages.com)).

**Table C.13.** Parent responses to utility survey items (N = 332)

Question	Percent of parents reporting
1. How long to complete ASQ:SE-2?	
a. Less than 10 minutes	31.3
b. 10–20 minutes	39.5
c. 20–30 minutes	17.2
d. 30 minutes–1 hour	0.3
e. More than 1 hour	0.9
Missing	10.8
2. Was ASQ:SE-2 easy to understand?	
a. Yes	93.1
b. Sometimes	5.4
c. No	0.3
Missing	1.2
3. Were ASQ:SE-2 questions appropriate?	
a. Yes	80.4
b. Sometimes	8.7
c. No	1.2
Missing	9.7
4. The ASQ:SE-2 questionnaires was . . . (check all that apply)	
a. Fun to do	44.9
b. Interesting	62.0
c. Took too long	1.2
d. Helped me think about my child	73.2
e. Waste of time	1.2
f. Didn't tell me much	6.6
5. Would you fill out another ASQ:SE-2?	
a. Yes	94.3
b. No	1.5
Missing	4.2

## REFERENCES

- Achenbach, T., & Rescorla, L. (2000). *Manual for the ASEBA preschool forms and profiles*. Burlington: University of Vermont, Research Center for Children, Youth, and Families.
- Beeber, L., Chazan Cohen, R., Squires, J., Jones Harden, B., Boris, N., Heller, S., & Malik, N. (2007). The early promotion and intervention research consortium (E-PIRC): Five approaches to improving infant/toddler mental health in Early Head Start. *Infant Mental Health Journal, 28*(2), 151–170.
- 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.
- Bricker, D., Squires, J., Twombly, E., Yockelson, S., & Kim, Y. (1996). *Behavior–Ages and Stages Questionnaires*. Unpublished manuscript.
- Carter, A., & Briggs-Gowan, M. (2006). *Manual for the Infant-Toddler Social & Emotional Assessment (ITSEA) and Brief-ITSEA (BITSEA)*. San Antonio, TX: Psychological Corporation.
- DelCarmen-Wiggins, R., & Carter, A. (Eds.). (2004). *Handbook of infant, toddler, and preschool mental health assessment*. New York, NY: Oxford University Press.
- Dorans, N.J., & Holland, P.W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P.W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35–66). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Heiman, G. (1992). *Basic statistics for the behavioral sciences*. Boston, MA: Houghton Mifflin.

- Krogstad, J., & Cohn, D. (2014). U.S. Census looking at big changes in how it asks about race and ethnicity. Retrieved from Pew Research Center web site: <http://www.pewresearch.org/fact-tank/2014/03/14/u-s-census-looking-at-big-changes-in-how-it-asks-about-race-and-ethnicity/>
- Lord, C., Rutter, M., DiLavore, P., & Risi, S. (2012). *Autism Diagnostic Observation Schedule*. Los Angeles, CA: Western Psychological Services.
- Mackrain, M., LeBuffe, P., & Powell, G. (2007). *Devereux Early Childhood Assessment for Infants and Toddlers (DECA-IT)*. Lewisville, NC: Kaplan Early Learning.
- Printz, P.H., Borg, A., & Demarree, M.A. (2003). *A look at social, emotional, and behavioral screening tools for Head Start and Early Head Start*. Newton, MA: Education Development Center, Center for Children & Families.
- Saliva, J., Ysseldyke, J., & Bolt, S. (2012). Assessment. In *Special and inclusive education* (12th ed.). Belmont, CA: Wadsworth.
- Spector, P.E. (1992). A consideration of the validity and meaning of self-report measures of job conditions (Paper 567). *Psychology Faculty Publications*. Retrieved from [http://scholarcommons.usf.edu/psy\\_facpub/567](http://scholarcommons.usf.edu/psy_facpub/567)
- Swets, J.A., & Pickett, R.M. (1982). *Evaluation of diagnostic systems: Methods from signal detection theory*. San Diego, CA: Academic Press.
- U.S. Census Bureau. (2010a). *Educational attainment in the United States: 2010—Detailed tables*. Retrieved from <http://www.census.gov/hhes/socdemo/education/data/cps/2010/tables.html>
- U.S. Census Bureau. (2010b). *Overview of race and Hispanic origin: 2010*. Retrieved from <http://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf>
- Whitcomb, S., & Merrell, K. (2013). *Behavioral, social, and emotional assessment of children and adolescents* (4th ed.). New York, NY: Routledge.