Excerpted from:

ASQ-3 User's Guide

By Jane Squires, Ph.D., Elizabeth Twombly, M.S., Diane Bricker, Ph.D., & LaWanda Potter, M.S.

PSYCHOMETRIC STUDIES OF ASQ, *THIRD EDITION*

This section of the report presents a range of empirical information collected on the Ages & Stages Questionnaires® (ASQ) since 2004. These data were used to guide the revisions entailed in the third edition (ASQ-3). Data include 18,572 completed questionnaires on children between 1 and 66 months of age.

Reliability Studies

Reliability studies completed on the ASQ-3 include test–retest reliability, and interobserver reliability. In addition, internal consistency of ASQ-3 items was examined using correlational analyses and Cronbach's coefficient alpha (Cronbach, 1951). Each of these analyses is presented next.

Test–Retest Reliability

Test–retest reliability is designed to help determine the stability of test outcomes over time. Test–retest reliability of the ASQ-3 was examined by comparing two questionnaires completed by the same parent at a 2-week time interval. That is, parents were asked to complete the same interval questionnaire on their child twice within a 2-week time period between completions. Questionnaires completed by 145 parents were included in this analysis. Forty-two parents completed two questionnaires online, and 103 parents completed two paper questionnaires. Parents were blind to the results of the first questionnaire when they completed the second one. The two questionnaires completed by parents were then compared for agreement on classifications (i.e., typical, identified). The percent agreement for the 145 parents was 92%. Intraclass correlations ranged from .75 to .82, suggesting the ASQ-3 has strong test–retest reliability.

Interobserver Reliability

Interobserver reliability is the agreement of test outcomes that have been completed by at least two independent test administrators. The interobserver reliability of the ASQ-3 was examined by comparing questionnaires completed by parents with questionnaires completed by trained test examiners on the same children. Trained test examiners filled out a questionnaire on a child immediately after completing a standardized assessment (e.g., Battelle Developmental Inventory [BDI]). Interobserver reliability was derived by comparing the agreement between the classifications (i.e., screened or not screened) of 107 children based on the parental and trainer examiners' completion of ASQ-3. The percent agreement between ASQ-3 classifications between parents and trainer examiners was 93%. Intraclass correlations by area ranged from .43 to .69, suggesting robust agreement between parents and trainer examiners when completing the ASQ-3 on this group of 107 children.

Internal Consistency

The internal consistency of the questionnaires was addressed by examining the relationship between developmental area and overall scores. Correlational analyses and Cronbach's coefficient alpha (Cronbach, 1951) were calculated.

Pearson product–moment correlation coefficients were calculated for developmental area scores with overall score for 20 questionnaire age intervals. As shown in Table 12, the correlations by developmental area and overall score are consistent and generally range from .60 to .85. Only in the Gross Motor area are two correlations below .60. All correlations are significant at $p < .01$. These findings suggest moderate to strong internal consistency between developmental areas and total test score.

Table 13 contains the correlations between developmental area scores that have been collapsed across the 20 questionnaire age intervals. Again, all correlations are significant, suggesting congruence between developmental areas as well as developmental areas with overall scores.

Cronbach's coefficient alpha was calculated for developmental area scores for 20 age intervals. Alphas are presented in Table 14 and range from .51 to .87. These alphas indicate that ASQ-3 items have good to acceptable internal consistency.

The reliability of the questionnaires has been studied by examining the internal consistency, test–retest reliability, and interobserver reliability of the questionnaires. Internal consistency analyses indicated strong relationships across items and within areas on the questionnaires. The questionnaires also achieved substantial test–retest and interobserver reliability. Parents' evaluations of their children using the questionnaires were consistent over time. In addition, professional examiners' agreement with parental evaluations of children using the questionnaires was consistently high.

TABLE 12. CORRELATIONS BETWEEN AREA AND OVERALL SCORE

Age interval	n	Communication	Gross Motor	Fine Motor	Problem Solving	Personal-Social
2	352	.81	.51	.70	.83	.81
4	1,824	.67	.71	.85	.83	.78
6	633	.64	.74	.81	.80	.80
8	1,362	.73	.69	.74	.72	.74
10	899	.79	.72	.67	.74	.79
12	2,088	.78	.66	.68	.74	.80
14	811	.78	.66	.81	.78	.79
16	1,191	.73	.57	.74	.76	.78
18	616	.75	.60	.71	.74	.74
20	1,278	.75	.64	.73	.80	.76
22	404	.79	.67	.78	.79	.79
24	1,443	.77	.67	.69	.77	.81
27	559	.84	.66	.75	.83	.78
30	953	.79	.64	.78	.82	.76
33	546	.84	.66	.80	.83	.84
36	1,006	.80	.66	.81	.81	.78
42	956	.82	.68	.82	.84	.80
48	672	.79	.71	.82	.80	.81
54	590	.81	.68	.81	.75	.77
60	389	.77	.75	.84	.72	.71

TOTAL NUMBER OF QUESTIONNAIRES COMPLETED = 18,572 ALL CORRELATIONS ARE SIGNIFICANT AT P < .01.

TABLE 13. CORRELATIONS BETWEEN AREA SCORES COLLAPSING ACROSS QUESTIONNAIRES

	AREA				
Area	Communication	Gross Motor	Fine Motor	Problem Solving	Personal-Social
Communication					
Gross Motor	.33				
Fine Motor	.36	.36			
Problem Solving	.50	.37	.52		
Personal-Social	.54	.41	.46	.53	
Overall	.76	.65	.73	.78	.79

ALL CORRELATIONS ARE SIGNIFICANT AT P < .01.

TABLE 14. STANDARDIZED ALPHAS BY AREA AND AGE INTERVAL

Age interval	n	Communication	Gross Motor	Fine Motor	Problem Solving	Personal-Social
2	352	.76	.57	.56	.78	.51
4	1,194	.60	.64	.73	.73	.60
6	602	.57	.61	.70	.70	.61
8	1,328	.69	.68	.70	.69	.54
10	446	.69	.81	.71	.69	.67
12	2,035	.68	.82	.55	.61	.63
14	481	.73	.87	.60	.70	.63
16	1,176	.70	.81	.64	.66	.59
18	592	.74	.77	.58	.54	.56
20	1,002	.77	.71	.57	.53	.58
22	399	.80	.72	.57	.56	.61
24	1,371	.80	.64	.51	.53	.58
27	546	.78	.68	.65	.61	.58
30	935	.75	.62	.75	.65	.65
33	537	.76	.62	.77	.69	.65
36	982	.71	.69	.77	.69	.61
42	950	.72	.68	.76	.72	.66
48	667	.80	.69	.76	.70	.68
54	586	.83	.73	.79	.75	.71
60	387	.66	.72	.83	.78	.67

NOTE: ANALYSES INCLUDE ONLY QUESTIONNAIRES WITH NO MISSING ITEMS.

Concurrent Validity

Concurrent validity was measured by comparing the classification of children based on their performance on a standardized test with their classification based on their performance on ASQ-3. Agreements meant that the child's ASQ-3 classification was the same as the standardized test, whereas disagreement meant that the ASQ-3 classification did not match the standardized test's classification.

Two groups of children were included in this analysis: 1) those who were tested for eligibility for Individuals with Disabilities Education Act Amendments (IDEA) of 1997 (PL 105-17) services and 2) those not receiving services and presumed to be developing without problems (hereafter called the typical group). The eligibility group (N = 257) participated in early intervention/early childhood special education (EI/ECSE) programs in Oregon, New York, and California, whereas the typical group (N = 322) was recruited from child care centers and preschool programs as well as Internet advertising for research participants. In most cases, the standardized measure used was the BDI, first and second editions (Newborg et al., 1987; Newborg, 2005). The BDI was administered and scored by trained examiners.

For the eligibility group, test data were gathered from children's files and included scores and classifications based on both the ASQ-3 and the BDI. For the typical group the ASQ-3 was completed by parents, and a trained examiner administered the BDI.

The child's performance on the standardized test was designated as "identified" if the child's scaled BDI score was equal to or less than 75 on any scale or subscale. This score was chosen because a child scoring at or below this point is likely to be functioning below developmental expectations for his or her chronological age and should be seen for further diagnostic assessment. In addition, a delay of 1.5–2 standard deviations (SD) on a standardized test meets eligibility criteria established by many states for entrance into EI/ECSE programs (www.NECTAC.org). (Although 75 was 2.5 points less than the BDI 1.5 SD cutoff score of 77.5, 75 was the mean cutoff score used for BDI eligibility decisions.)

The child's performance on ASQ-3 was considered "identified," or eligible, when his or her score was at or below the cutoff score set at 2 SD below the mean in one or more developmental areas. One of the following four outcomes was possible:

1. Both tests classified the child as typical or not identified.
2. Both tests classified the child as identified.
3. The standardized measure indicated the child as typical, and the questionnaire as identified.
4. The questionnaire classified the child as typical, and the standardized measure as identified.

Agreement between the BDI and ASQ-3 classifications for the total group of 579 children across questionnaires is contained in Figure 7. An examination of these data suggests the ASQ-3 has moderate to high agreement with BDI classifications. These findings are important because they provide objective evidence that in most cases ASQ-3 results will accurately identify children who require further evaluation and, importantly, will not identify those who are developing typically as needing further evaluation. Users of the ASQ-3 can be relatively confident that ASQ-3 results will identify those children whose development is suspect and those whose development is falling within typical developmental norms.

		Eligibility status		
		Eligible	Typical	Total
Status determined by ASQ	Eligible	217	47	264
	Typical	35	280	315
	Total	252	327	579

Sensitivity	Specificity	False positive	False negative	Percent agreement	Under-identified	Over-identified
86.1%	85.6%	14.4%	13.9%	85.8%	6.0%	8.1%

FIGURE 7. OVERALL ASQ-3 CONCURRENT VALIDITY.