Journal of Pediatric Psychology, Vol. 22, No. 3, 1997, pp. 313–328

Revision of a Parent-Completed Developmental Screening Tool: Ages and Stages Questionnaires[1]

Jane Squires,[2] Diane Bricker, and LaWanda Potter
University of Oregon

Received October 30, 1995; accepted April 1, 1996

Examined the Ages and Stages Questionnaires (ASQ), a series of 11 developmental questionnaires designed to be completed by parents and caregivers of young children from 4 to 48 months of age. The ASQ were recently revised and additional psychometric data were gathered. Analyses on over 7,000 questionnaires indicated high test–retest reliability, interobserver reliability, and internal consistency. Concurrent validity using standardized measures yielded an overall agreement of 85%, with a range of 76–91%. Specificity was high across questionnaire intervals while sensitivity was lower and varied across intervals. Use of parent-completed screening tools such as the ASQ is attractive in terms of cost-effectiveness, parental involvement, and flexibility in administration procedures.

KEY WORDS: developmental screening; infants at risk; early identification; early intervention.

The Ages and Stages of Questionnaires: A parent-completed child monitoring system (Bricker, Squires, & Mounts, 1995) (previously called the Infant/Child Monitoring Questionnaires) were developed in 1980 as an alternative screening assessment for infants and young children. Recently, the Ages and Stages Questionnaires (ASQ) have been revised in terms of content and format, and

[1]Support for this project was provided by the National Institute on Disabilities and Rehabilitation Research, U.S. Department of Education Grant H1336600122 and the March of Dimes Research Foundation Grant 12-206. Appreciation is extended to Amy Gambrell and Josie Holmsey of the Youngstown, Ohio, Easter Seal Society for contributing Watch-Me-Grow data; to Craig Leve for data analysis; and to Debi Eisert, Jane Farrell, Younghee Kim, and Susan Stewart for their helpful comments on earlier drafts of this paper.
[2]All correspondence should be sent to Jane Squires, Center on Human Development, 5253 University of Oregon, Eugene, Oregon 97403-5253.

0146-8693/97/0600-0313$12 50/0

additional psychometric data have been collected. The purpose of this manuscript is to describe these revisions and to report additional data and analyses relating to the validity and reliability of the questionnaire system.

The ASQ are a set of 11 developmental questionnaires each composed of three sections: a brief set of demographic items; 30 simply worded questions focusing on an infant's or child's developmental repertoire; and a brief section asking seven open-ended questions (e.g., Does anything about your child worry you?). The reading level ranges from fourth to sixth grade. The 30 items are equally divided into five domains: communication, gross motor, fine motor, problem solving, and personal-social.

The questionnaires were developed to be completed by the infant's or child's parent or primary caregiver. The questionnaires are designed in a bifold mailback format and can be completed in the home, clinic, or other community setting. Once an infant or child is identified for screening, a questionnaire appropriate for the child's age (or corrected age) can be mailed to the parents at the following intervals: 4, 6, 8, 12, 16, 18, 20, 24, 30, 36, and 48 months. For each item on the questionnaires, parents or caregivers may answer "yes" indicating the child performs the behavior; "sometimes" indicating the child performs the item on occasion (an emerging behavior); or "not yet" indicating the child does not yet perform the behavior. The scoring system for the questionnaires is as follows: 10 points for yes; 5 for sometimes; and 0 for not yet. Questionnaires are scored by totaling the domain scores for the questionnaire and comparing each domain score with the screening cutoff score for that domain. The cutoff scores have been established empirically and vary by domain and age interval. If the infant or child's score falls at or below the established cutoff score in one or more domains, the child's performance for that questionnaire is considered suspect and it is recommended that the child be referred for further evaluation assessment.

Three important premises underlie the development of the ASQ. First, screening systems should be dynamic and designed to follow children's development over time rather than relying on one or two test intervals. Second, those individuals who have the greatest familiarity with the child such as parents and caregivers should be included in the screening. Further, there are data suggesting that most parents can accurately judge whether their children can or cannot perform observable behaviors (Byrne, Backman, & Smith, 1986; Diamond & Squires, 1993; Hagekull, Bohlin, & Lindhagen, 1984; Squires, Nickel, & Bricker, 1990).

Third, for screening systems to be widely and routinely used, they must be economical. Clinical assessments of children by highly trained specialists on a routine basis do not meet this criterion; however, asking parents to complete a simple questionnaire on their child at periodic intervals does. Although some parents will be unable to complete questionnaires independently (e.g., non-

readers), this type of screening leaves more agency resources available for families who need personal contact.

Investigations funded by the National Institute on Disability and Rehabilitation Research and the March of Dimes Research Foundation were undertaken to address the questionnaires' reliability and validity (see Bricker & Squires, 1989a, 1989b; Bricker, Squires, Kaminski, & Mounts, 1988). In general, results indicated that parents were able to complete the questionnaires accurately. Second, interrater agreement and test–retest agreement on the questionnaires were high, generally in excess of 90%. Third, parents' and trained testers' agreement on the classification of the child's performance as "typically developing" or "needing further assessment" was found to be generally high. Agreements ranged from 76–92% depending on the test interval (e.g., 4 months, 24 months).

REVISION OF THE AGES AND STAGES QUESTIONNAIRES

In response to empirical analyses and utility data (e.g., Brinker, Frazier, Lancelot, & Norman, 1989) gathered on the ASQ since 1980, the questionnaires were revised. Modifications were made based on feedback from project staff, interventionists, parents, nurses, and pediatricians working with the questionnaires in clinic and research settings. Eight types of changes were made. First, 25 items across the first 10 questionnaires were reworded to clarify meaning. In most cases, the modifications entailed minimal word changes. For example, "reach for a toy" was changed to "try to get a toy"; "couch or adult chair" was changed to "furniture." In other cases, illustrations were added or modified, to clarify the meaning of the item. For example, one illustration of an infant was modified to clarify that the infant was leaning on her hands for balance rather than sitting unassisted.

Second, 8 items that parents found difficult to interpret were eliminated and replaced with items parents found easy to understand. In all cases, the substituted items appeared on a questionnaire at the preceding or succeeding interval and approximated the developmental quotient of the eliminated item. For example, at 20 months, the fine motor item "Does your child hold a spoon between her fingers and thumb when feeding herself?" was eliminated and replaced with a fine motor item from the 24-month questionnaire, "Does your child use a turning motion with her hand while trying to turn doorknobs, wind-up toys, twist tops, or screw lids on or off jars?" Both the eliminated item and the substituted item have a developmental quotient of 90–100 at 20 months.

The third change made on the revised questionnaires was the elimination of items with a developmental quotient of 125–150. These items were originally included to test parents' tendency to overestimate their child's performance (cf Gradel, Thompson, & Sheehan, 1981; Hunt & Paraskevopoulos, 1980). An

analysis of these items (Squires, Potter, & Bricker, 1995) revealed that a small percentage of children at each questionnaire interval was able to perform these developmentally advanced items. Thus, these items could not be reliably used to test parent overestimation. In addition, concurrent validity data on the ASQ reflecting strong parent–professional agreement minimized concerns about parental overestimation. That is, our questionnaire data suggest parents accurately report skills their children can currently perform and rarely misrepresent their children's developmental acumen.

A fourth change was the reordering of items within each domain according to their developmental level. The 6 items in each domain (e.g., communication, fine motor) were rearranged beginning with the lowest-age item moving to the highest-age item. For example, on the 12-month questionnaire, items with a developmental age of 8 months appear first, followed by items with a developmental age of 12 months which, in turn, are followed by the items with a developmental age of 16 months. On the original version of the questionnaires, items appeared in random developmental order within each domain. By reordering items from simple to more complex, it was felt that the educational value of the questionnaires for parents and caregivers would be enhanced.

Fifth, 6-, 18-, and 48-month questionnaires were added to the series, making a total of 11 different questionnaires. The 6- and 18-month questionnaires were added to assist medical personnel who routinely see babies at these intervals for well-baby checks. The 48-month questionnaire was developed at the request of professionals whose clientele included 4-year-olds.

Sixth, the overall section at the end of each questionnaire (which is not quantitatively scored) was modified. On the first 3 questionnaires (4, 8, 12 months), questions about hearing ("Do you think your child hears well?") and motor development ("Does your baby use both hands equally well?" and "Are his feet flat on the surface most of the time?") were substituted for general questions about feeding and sleeping. On the toddler and preschool questionnaires (16, 20, 24, 30, 36, 48 months), a language question ("Do you think your child talks like other toddlers/children her age?") and a motor question ("Do you think your child walks, runs, and climbs like other toddlers/children her age?") were substituted for general questions about feeding and sleeping. Feedback from a broad range of professionals and caregivers who had used the questionnaires indicated these general questions did not consistently trigger close observation of a child's behavior while the more specific questions appeared to assist in directing the adult's observation to aspects of children's behavior that may be indicative of potential problems. Also, empirical findings reported in the literature (Hack, Klein, & Taylor, 1995; Korner et al., 1993; Nickel, 1995) indicate the addition of two questions ("Does either parent have any family history of childhood deafness or hearing impairment?"; "Has your child had any medical problems in the last several months?") could potentially be useful in the early identification of developmental delays.

Seventh, a last page was added to each questionnaire containing a summary scoring section, a simplified scoring grid displaying the cutoff scores for that interval, and computer scanning boxes. This optional page allows either parents or service providers to score questionnaires and provides a summary page of child data for record keeping.

Finally, the name of the questionnaires was changed from the Infant/Child Monitoring Questionnaires (ICMQ) to the Ages and Stages Questionnaires (ASQ), to more accurately reflect the content and structure of the questionnaires and to present a more appealing and inviting identifier to parents and professionals.

Given the minimal revisions made to the questionnaire, the data collected on the original and revised questionnaires were merged to create a larger and, thus likely, more representative sample of child performance at each developmental interval. These data were used to reexamine the validity and reliability of the questionnaires. The analyses that follow refer to this combined data set as ASQ data. The primary purpose of this paper is to report on the findings from these studies.

METHOD

The sample was composed of children with medical risk factors, environmental risk factors, and a normative sample. The children with medical risk factors were recruited from infants who remained at least 3 days in a Level IV Neonatal Intensive Care Unit (NICU) needing medical care for a specific problem (e.g., respiratory distress, prematurity, low birth weight). The environmental risk sample was recruited using one or more of the following criteria: (a) extreme poverty (according to family income level, as defined by federal guidelines, 100% poverty level); (b) maternal age was 19 years or younger at the time of the infant's birth; (c) maternal education was less than 12th grade; and/or (d) parents who had experienced involvement with Children's Protective Services for abuse of their children. The normative sample, recruited through newspaper advertisements and by contacting child care centers, comprised children who met the following criteria: (a) no previous history of developmental or serious health problems as reported by parents; (b) full-term (>37 weeks) at birth; (c) never assigned to a neonatal intensive care unit. Those who met criteria and gave informed consent were selected. Because of the overlap among risk factors, the two risk categories—medical and environmental—were combined into one "risk" category. Detailed analyses specifically related to risk status and completion of ASQ are described elsewhere (Squires et al., 1995); however, the mean domain score of children from the combined risk groups was lower on 94% of ASQ items (n = 224) than mean domain scores of the normative sample. In addition, percentage agreement between questionnaires completed by parents

from the risk group and a standardized assessment ranged from 80–91% (M = 85) whereas percentage agreement for questionnaires completed by parents in the normative sample ranged from 84–93% (M = 89).

The total number of children in the sample was 2,008. (Not all children and parents are included in all studies of the questionnaires. Most validity and reliability studies included only a subsample of children and parents.) Of these, 53% (n = 1,068) were male; 47% (n = 940) were female. Eighty-one percent (n = 1,620) were risk status; 19% (n = 388) were nonrisk status.

While the ethnicity of the sample was representative of the U.S. population (Current Population Survey, 1991) with regard to the percentage of Caucasian and African American participants, Hispanics were underrepresented and Native Americans were overrepresented, each by about 15%. Mothers' educational level was similar to percentages of U.S. population for children aged 0–2 (Current Population Survey, 1991). Annual income of families (n = 984) ranged from less than $5,000 (13%); $5,001–10,000 (13%); $10,001–15,000 (12%); $15,001–20,000 (15%); $20,001–25,000 (14%); more than $25,000 (32%).

The total numbers of questionnaires completed by the study sample at each questionnaire interval ranged from a high of 1,500 at 4 months to a low of 535 at 36 months. The maximum number of ASQs that could be completed for each subject was eight (i.e., at 4, 8, 12, 16, 20, 24, 30, and 36 months). (This analysis was completed prior to the development of the 6-, 18-, and 48-month questionnaires.) Only one questionnaire was completed on the majority of the study subjects (n = 950). For 198 subjects, six consecutive questionnaires were completed between 4 and 24 months. For 62, or 2% of subjects, all eight questionnaires between 4 and 36 months were completed. The decreasing numbers of questionnaires completed at each interval were likely due to several reasons. The main reasons were that families moved and left no forwarding address, and that research studies terminated after 1 or 2 years. In two studies where we were able to examine attrition, only 19 of 450 parents who participated refused to complete consecutive ASQs.

RESULTS

Reliability

Internal Consistency

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

Pearson product-moment correlation coefficients were calculated to examine the relationship between domain totals and the overall score for each ques-

Table I. Correlations Between Domain and Overall Score on the Eight Questionnaires[a]

Questionnaire's interval (months)	*n*	Area				
		Communication	Gross motor	Fine motor	Problem solving	Personal-social
4	869	.71	.70	.81	.81	.79
8	768	.72	.76	.79	.79	.79
12	617	.75	.70	.77	.78	.83
16	502	.75	.54	.76	.75	.73
20	443	.75	.70	.66	.77	.71
24	393	.69	.63	.74	.76	.76
30	305	.76	.69	.73	.83	.69
36	248	.77	.77	.78	.83	.73

[a]All correlations are significant at $p < .0001$.

tionnaire as shown in Table I. The domain total was obtained by summing across all items in each domain. The overall score was obtained by summing across the five domain totals. All correlations were significant at $p < .0001$.

Pearson product-moment correlation coefficients were also calculated for the five separate domains across questionnaires. To conduct this analysis domain totals were collapsed across the 4-, 8-, 12-, 16-, 20-, 24-, 30-, and 36-month questionnaires. As shown in Table II, all correlations were significant at $p <$ 0001. Cronbach's coefficient alpha was calculated for domain totals on individual questionnaires. For the communication domain, alphas ranged from .63 at 4 months to .75 at 24 months. For gross motor domain, alphas ranged from .53 at 4 months to .87 at 12 and 16 months. For the fine motor domain alphas ranged from .49 at 20 months to .79 at 8 months. For the problem solving domain, alphas ranged from .52 at 20 months to .75 at 8 months. Finally, for the personal-social domain, alphas ranged from .52 at 16 months to .68 at 12 months.

Table II. Correlations Between Domain Scores Collapsing Across Questionnaires[a]

Area	Area				
	Communication	Gross motor	Fine motor	Problem solving	Person-social
Communication	—				
Gross motor	.46	—			
Fine motor	.46	.49	—		
Problem solving	.64	.52	.51	—	
Personal-social	.48	.51	.39	.59	—
Overall	.77	.77	.78	.83	.73

[a]N = 4,145 ASQ. All correlations are significant at $p < .0001$.

Test–Retest Reliability

Test–retest reliability was determined by comparing the scores of two questionnaires completed by parents at a 2-week interval. To assess the test–retest reliability of the questionnaires, parents who brought their infants to the center for a standardized assessment completed a second identical questionnaire immediately before the standardized assessment was administered. The two questionnaires completed by parents were then compared for agreement. Test–retest reliability, measured as percentage agreement between classifications based upon the questionnaires completed by 175 parents at 2-week intervals, was 94%. The standard error of measurement was .10.

Interobserver Reliability

Interobserver reliability was examined by comparing children's classifications based on questionnaires completed by parents with the classifications based on questionnaires completed by examiners immediately after the standardized assessments. Interobserver reliability, measured as percentage agreement between classifications based on the questionnaires completed by 112 parents and those completed by two professional examiners, was 94%. The standard error of measurement was .12. To be conservative in calculating interobserver reliability (Shavelson, Webb, & Rowley, 1992), 74 protocols were eliminated from this analysis because one or more test areas had two or more uncompleted items. This outcome occurred because the professional examiner had little opportunity to observe children engaged in certain activities (e.g., eating, pretend play, self-help skills). For example, on the 20-month ASQ, the item, "Does your child copy activities you do, such as wipe up a spill, vacuum, shave, or comb your hair?" was difficult for examiners to observe and therefore was not scored. We have no reason to believe that any particular bias was operating in the elimination of infants from this analysis.

Summary

The reliability of the questionnaires was 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 ASQ. The questionnaires also achieved substantial test–retest and interobserver reliability. Parents' evaluation of their children using the questionnaires was consistent over time. In addition, professional examiners' agreement with parental evolutions of children using the questionnaires was consistently high, near 95% agreement on classifications.

Validity

Determination of Screening Cutoff Points

Originally each ASQ was scored using a cutoff point that was determined by subtracting 2 standard deviations from the mean in each domain. The rationale for employing this strategy rested on the assumption that most standardized assessments utilize a cut point 1½ to 2 standard deviations below the mean to categorize performance. With further consideration of the importance of maximizing the accuracy of a screening tool, the ASQ were analyzed to determine the most useful scoring criteria—that is to maximize sensitivity and specificity and minimize the likelihood of overscreening and underscreening. To address this need, an analytical procedure called the "relative" or receiver operating characteristic (ROC; Swets & Picket, 1982) was employed. The ROC, based in statistical decision theory (Peterson, Birdsall, & Fox, 1954), has been used in a variety of disciplines including human perception and decision making (Green & Swets, 1966) as well as developmental screening (Glascoe & Byrne, 1993; Meisels, Henderson, Liaw, Browning, & Have, 1993). This technique involves computing changes in sensitivity and specificity when cutoffs are raised or lowered. The original technique for determining scoring cutoffs and the ROC analysis are described below in conjunction with their empirical cutoff points.

ROC Analysis

Using data from the questionnaires, means, standard deviations, and cutoff points were generated. Data for male and female subjects were combined for each questionnaire interval because a *t* test by gender indicated no significant differences. A matrix showing the conditional probabilities that exist at each cutoff point based on standard deviation units was designed. Second, a ROC curve that graphically displayed the probabilities at each cutoff point was generated.

For each questionnaire interval, a matrix was designed using the cutoff points of 2, 1.5, and 1 standard deviation units below the mean. The following conditional probabilities were calculated at each cutoff point: (a) sensitivity, (b) specificity, (c) true positive proportion, (d) false positive proportion, (e) underscreening, and (f) overscreening. A sample matrix for the 12-month questionnaire can be found in Table III. As expected, when the cutoff becomes less conservative (i.e., 1.5 or 1 *SD* from the mean), the overscreening rate increases as the underscreening rate decreases.

Further demonstration of the trading relationship of true positive and false positive proportions as the cutoff point is adjusted is provided with the graphic representation of the ROC curve in Figure 1. It can be seen from Table III how

Table III. Cutoff Points by Standard Deviation Units and Accompanying Conditional Probabilities for the 12-Month Questionnaire

Standard deviation unit	Developmental area	Cutoff	Sensitivity	Specificity	True positive	False positive	Overscreening	Underscreening
2 *SD* below the mean			.85	.86	.85	.13	.12	.02
	Communication	1.54						
	Gross motor	1.80						
	Fine motor	2.87						
	Problem solving	2.51						
	Personal-social	1.98						
1.5 *SD* below the mean			.95	.71	.95	.28	.25	.006
	Communication	2.21						
	Gross motor	2.57						
	Fine motor	3.39						
	Problem solving	3.10						
	Personal-social	2.62						
1 *SD* below the mean			.97	.59	.97	.41	.36	.003
	Communication	2.88						
	Gross motor	3.33						
	Fine motor	3.90						
	Problem solving	3.68						
	Personal-social	3.26						

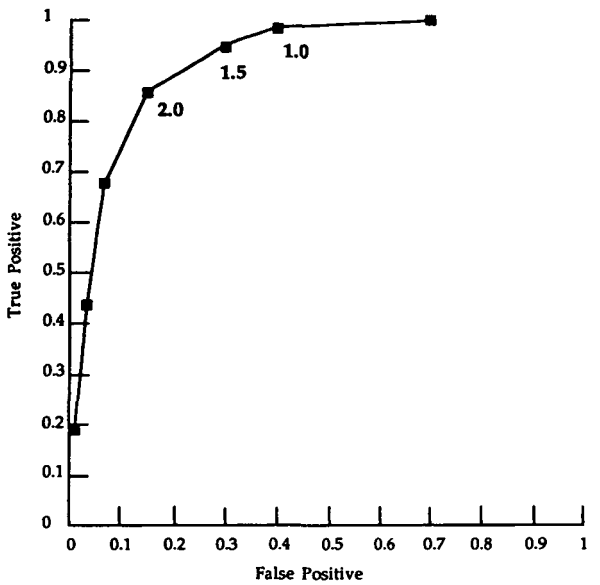

Fig. 1. ROC curve generated for 12-month questionnaire using 1, 1.5, and 2 standard deviations.

the cutoff points differ in their placement on the curve. The 2 *SD* cutoff point, although not perfect, appeared the most balanced cutoff point in terms of the true positive and false positive propositions.

For every questionnaire interval, the process of evaluation described above was undertaken and conditional probabilities across questionnaires were computed to arrive at a mean figure for each interval. These probabilities are reported in Table IV. The same trend as described for the 12-month questionnaire is evident in Table IV for other questionnaire intervals. Specifically, the sensitivity or true positive proportion could be maximized as the cutoff point was raised but at the expense of specificity, false positive proportions, overscreening, and underscreening rates.

Based on our analyses using the contingency table conditional probabilities and the ROC curve, a screening cutoff point of 2 standard deviations below the mean across questionnaire intervals was recommended. (Agencies with sufficient resources to assess additional children can raise the cutoff points to 1.5 or 1.0 *SD*.) Errors will always occur when using screening tools because of their brevity, the uneven nature of early human development, and the balancing of the conditional probabilities used to determine a test's accuracy. Because of the serial nature of the questionnaires however, a child with a verifiable delay has a high

Table IV. Mean Conditional Probabilities Across Combined Questionnaire Intervals by Cutoff Point

Cutoff point	Sensitivity	Specificity	True positive	False positive	Over-screening	Under-screening
2 *SD* below the mean	.75	.86	.75	.14	.12	.04
1.5 *SD* below the mean	.82	.74	.82	.26	.23	.03
1 *SD* below the mean	.91	.60	.91	.40	.35	.01

probability of being identified at some test interval, even if the delay is missed at one or two previous test intervals.

Concurrent Validity

Concurrent validity was measured by comparing the classifications of the child's performance based on the parent-completed ASQ with the classification of the child's performance on a professionally administered standardized test administered within 29 days. The Revised Gesell and Armatruda Developmental and Neurologic Examination (Knobloch, Stevens, & Malone, 1980) and the Bayley Scales of Infant Development (Bayley, 1969) were used for infants up to 30 months; the Stanford-Binet Intelligence Test-4th Edition (Thorndike, Hagen, & Sattler, 1985) was used for children over 3 years; and the McCarthy Scales of Children's Abilities (McCarthy, 1972) was used for 4-year-olds. The child's performance on the ASQ was classified as "refer" if any domain score fell at or below the 2 standard deviation cutoff point on any questionnaire. The child's performance on the standardized test was designated as "fail" if the child scored at or below 1.5 standard deviations below the mean on any scale or subscale. This scaled score (76 on the Bayley, McCarthy, and Stanford-Binet; and 75 on the Gesell) was chosen because it was felt that a child scoring at or below this point was suspect for developmental delay and should be seen for further diagnostic assessment. In addition, a 1.0–1.5 standard deviation delay on a standardized test meets eligibility criteria established by many states for entrance into early intervention programs (Brown & Brown, 1993). Results obtained when combining assessments in order to make overall statements about the validity of the ASQ can be found in Table V.

Concurrent Validity with Children with Disabilities

A study of the concurrent validity of the questionnaires with a subsample of children with disabilities was undertaken. Children in this study ranged from 4 to

Table V. Cross-Tabulation of Agreement Between Standardized Assessments and ASQ Collapsed Across Questionnaires

ASQ	Standardized assessment[a] Fail	Pass	Total
Refer for assessment	149	187	336
Developing typically	60	1,115	1,175
Total	209	1,302	1,511

% Overall ASQ analysis

Sensitivity	Specificity	False positive	True positive	False negative	Agreement	Under-screening	Over-screening	Screened	Positive predictive value
74.56	86.21	13.78	74.56	23.43	84.11	3.99	11.88	2[illegible]95	45.98

[a] Includes Bayley Scales of Infant Development, Gesell Developmental Schedules, Stanford-Binet Intelligence Scale, and McCarthy Scales of Children's Abilities.

36 months and were enrolled in state-funded early intervention programs for children with moderate to severe disabilities. These children had received a multidisciplinary assessment and were found by evaluators to meet state eligibility guidelines for receiving early intervention services. Of the 52 children whose parents completed questionnaires, 50 (96%) were categorized as "refer" for further assessment (i.e., scored below the established cutoff points). Those who were not identified by the questionnaires were eligible for early intervention services due to chronic health impairments, and might be expected to perform above the screening cutoffs in some developmental domains.

The validity of the ASQ has been evaluated extensively. The concurrent validity of the questionnaires as reported in percentage agreement between questionnaires and standardized assessments ranged from 76% for the 4-month ASQ to 91% for the 36-month ASQ, with 88% agreement overall. Sensitivity ranged from 51% for the 4-month ASQ to 90% for the 36-month ASQ with 75% overall. Specificity of the questionnaires ranged from 81% for the 16-month ASQ to 92% for the 36-month ASQ, with 86% overall. Although sensitivity, or the ability to detect delayed development, was lower, averaging 75%, specificity, or the ability of the ASQ to correctly identify typically developing children, remained high across questionnaire intervals and standardized assessments. A separate analysis of the ability of the questionnaires to identify children with established developmental delays yielded high agreement (96%).

DISCUSSION

The ASQ appear to be a reliable screening system that uses parents to complete simple developmental questionnaires on their children. The central goal in revising the ASQ was to "fine tune" the questionnaires, thus improving their accuracy and efficiency. By clarifying some items and eliminating items difficult to interpret, parents are likely to find the questionnaires easier to understand and use with their child. The potential teaching effect of the questionnaires is also likely improved with the revision. By eliminating items with developmental quotients above 125, there is less confusion about what is developmentally appropriate at each questionnaire interval. As well, arranging the items into a hierarchy beginning with the easiest items may increase parental understanding of the developmental growth sequence in each domain. Another improvement useful to professionals is the addition of three questionnaire intervals (i.e., 6, 18, and 48 months). This addition allows professionals more flexibility in using the questionnaires across the age continuum from 4 to 48 months.

The revision of the overall section on the ASQ likely benefits both parents and professionals. The items in this section were modified to address the more specific concerns of professionals (e.g., hearing, equal use of hands and feet).

The addition of these items may prompt parents to evaluate their child's development and performance on the ASQ with these items in mind.

For the most part, revisions made to the original questionnaires were minor and therefore data gathered using the original and revised questionnaires have been combined to conduct analyses of internal consistency as well as to establish screening cutoff points. The combining of these data sets significantly increases the number of questionnaires at each interval, which in turn, likely provides better information on the instrument's reliability and validity.

In studying the validity and reliability, analyses were hampered by a small number of children whose performance on the standardized assessment suggested a delay or problem. Overall, only 14% of children (n = 209 across 8 test intervals) scored at or below the 1.5 standard deviation cutoff point on the standardized assessment. Excluding the 4-month questionnaire, this figure drops to 11%. At several test intervals, decisions relating to sensitivity and specificity were made on less than 5 children. Although a separate study of children with disabilities was undertaken, this sample was not combined with the overall sample because a different procedure was used to measure concurrent validity; however 50 out of 52 questionnaires completed on children with disabilities suggested a potential delay.

Parent-completed questionnaires may not be appropriate for all parents and children. Some parents are unable to complete questionnaires satisfactorily due to reading, organizational, mental health, and cognitive disabilities; however, substantial numbers of parents can be relied upon to accurately monitor the development of their children over time using the questionnaires. Limited agency resources can then be allocated to families who need personal contact.

Previous investigations (Bricker et al., 1989) have found the ASQ to be a reliable, accurate, and economical tool. The increased demand for such a tool, not only by agencies interested in using the ASQ in state tracking programs but also by culturally diverse programs, requires continued efforts to examine and improve this dynamic screening system for infants and young children.

REFERENCES

Bayley, N. (1969). *The Bayley Scales of Infant Development*. New York: Psychological Corp.

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

Bricker, D., & Squires, J. (1989b). Low cost system using parents to monitor the development of at-risk infants. *Journal of Early Intervention, 13,* 50-60.

Bricker, D., Squires, J., Kaminski, R., & Mounts, L. (1988). The validity, reliability and cost of a parent completed questionnaire system to evaluate at-risk infants. *Journal of Pediatric Psychology, 13,* 56-68.

Bricker, D., Squires, J., & Mounts, L. (1995). *Ages and Stages Questionnaire: A parent-completed, child-monitoring system.* Baltimore, MD: Paul H. Brookes.

Brinker, R., Frazier, W., Lancelot, B., & Norman, J. (1989). Identifying infants from the inner city for early intervention. *Infants and Young Children, 2,* 49-58.

Brown, W., & Brown, C. (1993). Defining eligibility for early intervention. In W. Brown, S. Thurman, & L. Pearl (Eds.), *Family-centered early intervention with infants and toddlers* (pp. 21-42). Baltimore, MD: Paul H. Brookes.

Byrne, J., Backman, J., & Smith, I. (1986). Developmental assessment: The clinical use and validity of parent report. *Journal of Pediatric Psychology, 11,* 549-559.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334.

Current Population Survey. (1991). Washington, DC: U.S. Bureau of the Census.

Diamond, K., & Squires, J. (1993). The role of parental report in the screening and assessment of young children. *Journal of Early Intervention, 17,* 107-115.

Glascoe, F., & Byrne, K. (1993). The accuracy of three developmental screening tests. *Journal of Early Intervention, 17,* 368-379.

Gradel, K., Thompson, M., & Sheehan, R. (1981). Parental and professional agreement in early childhood assessment. *Topics in Early Childhood Special Education, 1,* 31-40.

Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Hack, M., Klein, N., & Taylor, G. (1995). Long-term developmental outcomes of low birth weight infants. *Future of Children, 5,* 176-196.

Hagekull, B., Bohlin, G., & Lindhagen, K. (1984). Validity of parental reports. *Infant Behavior and Development, 7,* 77-92.

Hunt, J., & Paraskevopoulos, J. (1980). Children's psychological development as a function of the accuracy of their mothers' knowledge of their abilities. *Journal of Genetic Psychology, 136,* 258-298.

Knobloch, H., Stevens, F., & Malone, A. (1980). *Manual of developmental diagnosis: The administration and interpretation of the revised Gesell and Armatruda developmental and neurological examination.* Hagerstown, MD: Harper & Row.

Korner, A., Stevenson, D., Kraemer, H., Spiker, D., Scott, D., Constantinou, J., & Dimiceli, S. (1993). Prediction of the development of low birthweight preterm infants by a new neonatal medical index. *Developmental and Behavioral Pediatrics, 14*(2), 106-111.

McCarthy, D. (1972). *McCarthy scales of children's abilities.* New York: Psychological Corp.

Meisels, S., Henderson, L., Liaw, F., Browning, K., & Have, T. (1993). New evidence for the effectiveness of the early screening inventory. *Early Childhood Research Quarterly, 8,* 327-346.

Nickel, R. (1995). Community consultants in the care of children with special health care needs. Portland, OR: Child Development and Rehabilitation Center.

Peterson, W., Birdsall, T., & Fox, W. (1954). *Transactions of IRE Professional Group Information Theory, 4,* 171.

Shavelson, R., Webb, N., & Rowley, G. (1992). Generalizability theory. In A. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (pp 233-256). Washington, DC: American Psychological Association.

Squires, J., Potter, L., & Bricker, D. (1995). *Ages and stages user's guide.* Baltimore, MD: Paul H. Brookes.

Squires, J., Nickel, R., & Bricker, D. (1990). Use of parent-completed developmental questionnaires for child-find and screening. *Infants and Young Children, 3*(2), 46-57.

Swets, J., & Picket, R. (1982). *Evaluation of diagnostic systems.* New York: Academic Press.

Thorndike, R., Hagen, E., & Sattler, J. (1985) *Stanford-Binet intelligence scale* (4th ed.). Chicago: Riverside.