Validity of the Ages and Stages Questionnaires in Term and Preterm Infants

WHAT’S KNOWN ON THIS SUBJECT: The Ages and Stages Questionnaires (ASQ) has been validated in many countries and translated into numerous languages. In most publications, it has been reported that the ASQ is accurate in detecting true problems in apparently healthy children and even in children with biological risk factors.

WHAT THIS STUDY ADDS: This report compares the third version of the ASQ and the Bayley Scales of Infant and Toddler Development, Third Edition, assessments. Psychometric properties showed a tendency to improve with testing age and when comparing term versus extremely preterm children.

BACKGROUND: This study assessed the concurrent validity of the parent-completed developmental screening measure Ages and Stages Questionnaires, Third Edition (ASQ-3) compared with the Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III) in children born term, late preterm, or extremely preterm at 8, 18, or 30 months of corrected gestational ages (CGA).

METHODS: Data were collected from 306 term and preterm children ages 8, 18, and 30 months’ CGA recruited from an ambulatory well-child clinic in Santiago, Chile. Parents completed the ASQ-3 in their homes, and afterward a trained professional administered the Bayley-III in a clinic setting. On the ASQ-3, the presence of any domain screened ≤2 SDs below the mean area score was considered a positive screen (indicating failure or delay). A Bayley-III score less than ≤1 SD indicated mild or severe delay.

RESULTS: ASQ-3 showed adequate psychometric properties (75% sensitivity and 81% specificity) and modest agreement with the Bayley-III (r = 0.58). Sensitivity, specificity, and correlations between measures improved with testing age and in children who were born extremely preterm.

CONCLUSIONS: Considering its psychometric properties, the ASQ-3 can be recommended for routine use in screening low-risk children at 8, 18, and 30 months’ CGA and is advisable to be included in follow-up programs for children with biological risk factors such as those born preterm. Pediatrics 2013;131:e1468–e1474

AUTHORS: Luisa Schonhaut, MD, MPH; Iván Armijo, BSc; Marianne Schönstedt, BSc; Jorge Álvarez, MD; and Miguel Cordero, MSc

AFFILIATIONS: Clínica Alemana, Facultad de Medicina, Universidad del Desarrollo, Santiago, Chile; Universidad Gabriela Mistral, Santiago, Chile; and Ministry of Health, Santiago, Chile

KEY WORDS: Ages and Stages Questionnaires, developmental delay, developmental screening, infant, low birth weight, premature


FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: Supported by the Clínica Alemana research grants program. Grant SA720043 provided further funding from the National Fund for Health Research, and grant S0CHIP 208013 was provided by the Chilean Society of Pediatrics.
The American Academy of Pediatrics (AAP) recommends a routine standardized screening assessment at well-child visits at 9, 18, and 30 (or 24) months and when there is clinical suspicion of a child being at risk for developmental delay. This guideline was adapted in 2008 for the Chilean Ministry of Health National Clinical Guidelines.

Standardized screening measures have been shown to promote early detection of developmental delay and to provide an opportunity for early intervention. A large variety of screening measures are available, as recommended by the AAP. Moreover, the AAP suggests that the practice setting, population served, and health care provider’s preferences should determine test choice. One of the main differences lies in application methodology; some tools are designed to be administered by a very high-trained professional, whereas others are based on parent or primary caregiver reports.

The Bayley Scales of Infant Development have traditionally been considered a comprehensive developmental assessment for monitoring and follow-up in children with increased biological risk factors; in fact, many investigators consider the Bayley scales to be the reference or gold standard for infant developmental assessment. However, its universal clinical applicability is limited due to its high cost, timing, and required administration by trained professionals. Given that the Bayley scales are not widely available in the Chilean National Health Services and elsewhere due to these requirements, it is important to have a tool to identify those children at risk for developmental delay who need a complete developmental assessment. Parent-completed developmental screening measures, such as the Ages and Stages Questionnaires (ASQ), balance these limitations.

The ASQ has gained popularity in busy pediatric clinics for screening children during well-child examinations. Many studies support its easy administration, short completion time, easy interpretation, and its capacity to dramatically enhance a clinician’s ability to detect children who have suspected developmental delays. In addition, the ASQ performs well with children with biological risk factors as well as those with environmental risk factors such as foster care placement. As an additional benefit, these measures provide an excellent opportunity for discussing parental concerns regarding developmental issues in children, with the goal of improving communication and physician-parent relationships. Moreover, these measures have a positive impact on promoting parental involvement in early child development.

During the last decade, the ASQ has been validated in many countries, in different languages and settings. There are reports of concurrent validation of the ASQ, Second Edition, and the Bayley Scales of Infant Development, Second Edition, in low-risk term children. In most of those publications, it has been reported that the ASQ is accurate in detecting true problems in apparently healthy children and even in children with biological risk factors. Recently, Simard et al published a study that contrasted these previous findings by using the second edition of the ASQ compared with the subscales of the Bayley Scales of Infant Development, Second Edition. To the best of our knowledge, there are no reports comparing the third editions of these questionnaires. Moreover, there are no reports regarding the psychometric properties of ASQ across different testing ages and prematurity conditions.

We conducted the current study to explore the concurrent validity of the ASQ, Third Edition (ASQ-3) compared with the Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III), in children born term, late preterm, or extremely preterm, assessed at 8, 18, and 30 months’ corrected gestational ages (CGA).

**METHODS**

**Study Population**

The current research was part of a larger national validation study of the ASQ-3 in a representative sample of the Chilean population in which we reported agreement between the ASQ-3 and Bayley-III in term children with low biological and social risk factors. The current sample was composed of children who attended a well-child clinic in Santiago, Chile, from April 2008 to April 2011. A total of 306 children were recruited; 119 born term (37–41 weeks’ gestational age), 124 late preterm (32–36 weeks’ gestational age), and 63 extremely preterm (<32 weeks’ gestational age or weight <1500 g).

The term sample was recruited from children and families who presented at their primary care provider for a routine well-child visit. Ten families refused to participate due to scheduling problems. For late and extremely preterm samples, 2 recruiting methods were used: (1) direct recruitment by a pediatrician; and (2) direct telephone contact by a pediatrician, with the use of the neonatal service databases as a source of contact information. Twenty cases refused to participate due to scheduling problems. All parents who agreed to participate signed a written informed consent form. All parents were Spanish speakers.

Development was assessed at 8, 18, or 30 months’ GA. We corrected gestational age for late and extremely preterm children by following universal recommended standards for GA, which was determined by using the mother’s last menstrual date and/or early obstetric ultrasound.

**ARTICLE**

PEDIATRICS Volume 131, Number 5, May 2013

Downloaded from by guest on July 25, 2017
e1469
Data are presented as n (%), mean ± SD, or %. NS, not significant.

** All of the comparisons are significant (P < .05).

** The difference in deficit frequency by age was not significant (P = .072).

** The difference in deficit frequency by age was significant (P = .001).

* Detected difference in deficit proportions.

b $\chi^2$ for associations between the deficit and administered test, corrected by a Fisher's exact test.

d Significant indicators (P < .05).

### Table 1: Demographic Characteristics of the Study Sample

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Age Group</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8 Months</td>
<td>18 Months</td>
</tr>
<tr>
<td></td>
<td>(n = 110)</td>
<td>(n = 100)</td>
</tr>
<tr>
<td>Gestational age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37–41 wk</td>
<td>43 (39)</td>
<td>39 (39)</td>
</tr>
<tr>
<td>32–36 wk</td>
<td>44 (40)</td>
<td>41 (41)</td>
</tr>
<tr>
<td>&lt;32 wk or &lt;1500 g</td>
<td>25 (21)</td>
<td>20 (20)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>45 (41)</td>
<td>47 (47)</td>
</tr>
<tr>
<td>Male</td>
<td>65 (59)</td>
<td>53 (53)</td>
</tr>
<tr>
<td>Multiple pregnancies</td>
<td>34 (31)</td>
<td>31 (31)</td>
</tr>
<tr>
<td>Maternal age, mean ± SD, y</td>
<td>32.8 ± 3.6</td>
<td>34.2 ± 3.3</td>
</tr>
<tr>
<td>Maternal years of education</td>
<td>17.7 ± 2.6</td>
<td>17.4 ± 2.5</td>
</tr>
<tr>
<td>Maternal occupation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paid work</td>
<td>71</td>
<td>77</td>
</tr>
<tr>
<td>Homemaker</td>
<td>29</td>
<td>23</td>
</tr>
<tr>
<td>Family income eighth and ninth deciles of income</td>
<td>95</td>
<td>95</td>
</tr>
</tbody>
</table>

### Table 2: Deficit Frequency of Psychomotor Development By Age Group According to the 2 Tests

<table>
<thead>
<tr>
<th>Testing Age</th>
<th>No. of Children</th>
<th>ASQ*, n (%)</th>
<th>Bayley-III, % (n)</th>
<th>$Z_{91-120}$</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 mo</td>
<td>110</td>
<td>41 (37.3%)</td>
<td>31 (28.2%)</td>
<td>1.44</td>
<td>20.07**</td>
</tr>
<tr>
<td>18 mo</td>
<td>100</td>
<td>24 (24.0%)</td>
<td>14 (14.0%)</td>
<td>1.89</td>
<td>22.83**</td>
</tr>
<tr>
<td>30 mo</td>
<td>96</td>
<td>23 (24.0%)</td>
<td>11 (11.5%)</td>
<td>2.43</td>
<td>20.98**</td>
</tr>
<tr>
<td>Total</td>
<td>306</td>
<td>88 (28.8%)</td>
<td>56 (18.3%)</td>
<td>3.06</td>
<td>66.12**</td>
</tr>
</tbody>
</table>

Sample size was calculated by using G*power software,27 over the basis of an estimated difference in developmental risk proportions of 0.15, a confidence level of 0.95, and an expected power of 0.9. A total sample of 112 cases per group was determined for the term and late preterm sample. All of the extremely preterm sample cases in the neonatal database available were included. Recruiting of cases was stopped when targets were met for sample size for each group. Each child participated only once in the study over a 3-year period.

Parents were asked to complete the ASQ-3 in their homes before Bayley-III testing. The time interval between both measures was no more than 2 weeks. Parents completed a form with personal, socioeconomic, demographic, and medical information. The research ethics board of Clínica Alemana–Universidad del Desarrollo approved the study.

### Outcome Measures

**ASQ-3**

The ASQ28 is a brief measure in which parents rate their child’s current skills and development, from 1 to 66 months of age. Twenty-one questionnaires are available within this age range. Parents answer 30 questions covering 5 domains of development, including communication, gross motor, fine motor, problem-solving, and adaptive skills. Parents are instructed to try activities with their child to facilitate accurate assessment. A pass/fail score was assigned for each area of development. The presence of any domain screened <2 SDs below the mean area score was considered a positive screen. The ASQ-3 was validated by using large, standardized samples of children from diverse ethnic and socioeconomic backgrounds. Although there is a Spanish version that has been used in other countries, we generated a linguistic adaptation for Chile before the initiation of the study.

**Bayley-III**

The Bayley-III is a comprehensive developmental assessment, for children ages 1 to 42 months. Three subscales were administered (cognitive, language, and motor) by an accredited occupational therapist, who was blinded to the ASQ-3 results. The Bayley-III is not available in Spanish, but the therapist administering the test was fluent in English and Spanish and was able to communicate adequately with the child when presenting the stimuli and observing responses. A child was considered to have developmental delay if he or she scored ≥1 SD below the mean in at least 1 domain on the Bayley-III.

### Statistical Analysis

Analysis of variance was used to compare demographic characteristics of children at the 3 different assessment ages. To evaluate the correlation between diagnostic categories of the ASQ-3 and Bayley-III, a Fisher’s exact test was computed. Also, a t test was used to evaluate differences in proportions of children with delay detected by using each test. A Pearson’s correlation between raw test scores was also calculated. Sensitivity, specificity, and predictive values were calculated for the ASQ-3 by using as a reference the gold standard Bayley-III. All data were analyzed for different CGA group populations and for different ages at testing (8, 18, and 30 months). SPSS version 18 (IBM SPSS Statistics, IBM Corporation, New York, NY) was used to analyze the data.
RESULTS

The ASQ-3 and Bayley-III were administered to 306 children who met the inclusion criteria (Table 1); there were no significant statistical or clinical differences between tested age groups. Although we found statistical differences in maternal age at testing time, there were no differences in maternal age at the time of delivery. All the families had higher socioeconomic status as measured by years of maternal education and income level of eighth and ninth deciles of income than the general Chilean population and were representative of the private medical center where the study was conducted.

The ASQ-3 allocated 88 (28.8%) children as having a suspected developmental delay and the Bayley-III identified 56 (18.3%) children as having a developmental delay (P < .05). Statistical significance was evaluated as the proportion of low scores and the association between each measure and detection of developmental delay at different testing ages (Table 2).

The ASQ-3 showed an overall agreement of \( r = 0.56 \) with the Bayley-III. This correlation increased with testing age (\( r = 0.75 \) at 30 months), as shown in Fig 1, and with the degree of prematurity (\( r = 0.65 \) for extremely preterm), as shown in Fig 2.

The psychometric characteristics of the ASQ-3 included a sensitivity of 75%, specificity of 81%, positive predictive value of 47%, and negative predictive value of 9%; it had a positive likelihood ratio (LR) of 4 and a negative LR of 0.31. The psychometric properties of the ASQ-3 showed a tendency to improve with testing age and degree of prematurity (Tables 3 and 4). There were no differences in ASQ psychometric characteristics and/or test correlations when twin populations were analyzed separately.

DISCUSSION

In this population, the ASQ-3 had adequate psychometric properties and concurrent agreement compared with the Bayley-III. We detected a tendency to improve both mean sensitivity and mean specificity with testing age and when comparing term versus extremely preterm children. Our findings are in accordance with those reported by other authors. Agreement seems to be higher in those studies that included higher risk children such as those who were born extremely preterm. However, a previous study questioned the capacity of ASQ for detecting developmental delay in extremely preterm children, when separate area scores were analyzed (instead of total scores) and when parents did not have time to observe and assess their child per authors’ recommendations.

The improved psychometric properties of the measure in the extremely preterm birth infants can be partly explained by the higher prevalence of developmental deficits in this group. However, mothers of preterm infants may be more aware and concerned, and they may somehow look at developmental issues of their offspring in more critical ways after their previous experience caring for the child in the NICU and participating in follow-up programs. Also, the fact that psychometric properties of the ASQ-3 were similar in twins may reflect the fact that mothers are able to focus, individualize, and identify the capabilities of each child independently.

The ASQ-3 identified more children at risk for developmental delays than
those diagnosed according to the Bayley-III. This result was reflected in low positive predictive values, consistent with other studies.\textsuperscript{12,30,31} It has been reported that psychomotor developmental screening measures may overdiagnose in 15% to 30% of tested children,\textsuperscript{32,33} and the “risk” for potential overidentification and over-referral has been widely debated in the literature. As with any screening test, over-referral may increase costs and is associated with increased parental anxiety. However, children identified as false-positives (children with suspected delays per the ASQ but typical per the Bayley-III) may represent a group at-risk for future academic difficulties. This at-risk group likely requires more support systems, as has been demonstrated in other studies.\textsuperscript{34,35} The conundrum of false-positive developmental screens may partially explain the suboptimal sensitivity (59%) of the ASQ in term children. Also, the theoretical “risk” of over-referrals has not yet been quantified with empirical research. In primary care settings, this risk could be ameliorated by the evidence-based benefits of developmental/behavioral promotion.\textsuperscript{35}

Despite its status as a reference test, use of the Bayley-III as a standardized tool may be questioned because it has a number of limitations. For example, some recent studies have reported that it underestimates language deficits in toddlers.\textsuperscript{36} These considerations are to be taken into account when interpreting our results.

One study limitation is that the population screened with the ASQ-3 represented only those middle-class parents who attended private medical clinics in Chile; however, this study is a first step to investigating developmental screening policies for improved pediatric care in Chile and other middle-income countries. In addition, a possible effect of small sample bias.

![FIGURE 2](image-url)

Correlations between the ASQ-3 and Bayley-III according to gestational age.

### TABLE 3 ASQ-3 Psychometric Values Compared With Bayley-III for Different Age Groups

<table>
<thead>
<tr>
<th>Values</th>
<th>Age Group</th>
<th>Total (N = 306)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8 Months</td>
<td>18 Months</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>71 (53–84)</td>
<td>79 (52–92)</td>
</tr>
<tr>
<td>Specificity</td>
<td>76 (66–84)</td>
<td>84 (75–90)</td>
</tr>
<tr>
<td>PPV</td>
<td>54 (39–68)</td>
<td>44 (27–63)</td>
</tr>
<tr>
<td>NPV</td>
<td>87 (77–93)</td>
<td>96 (89–99)</td>
</tr>
<tr>
<td>Positive LR</td>
<td>2.9 (1.9–4.6)</td>
<td>4.8 (2.8–8.4)</td>
</tr>
<tr>
<td>Negative LR</td>
<td>0.4 (0.22–0.67)</td>
<td>0.26 (0.09–0.7)</td>
</tr>
</tbody>
</table>

Data are presented as % (95% CI) or n (95% CI). NPV, negative predictive value; PPV, positive predictive value.

### TABLE 4 ASQ-3 Psychometric Values Compared With Bayley-III According to Gestational Age Group

<table>
<thead>
<tr>
<th>Value</th>
<th>Term (n = 119)</th>
<th>Late Preterm (n = 124)</th>
<th>Extreme Preterm (n = 63)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>59 (38–78)</td>
<td>80 (61–91)</td>
<td>86 (60–96)</td>
</tr>
<tr>
<td>Specificity</td>
<td>87 (73–92)</td>
<td>73 (65–81)</td>
<td>86 (73–93)</td>
</tr>
<tr>
<td>PPV</td>
<td>44 (26–65)</td>
<td>43 (30–57)</td>
<td>63 (40–81)</td>
</tr>
<tr>
<td>NPV</td>
<td>93 (86–96)</td>
<td>94 (86–97)</td>
<td>96 (85–99)</td>
</tr>
<tr>
<td>Positive LR</td>
<td>4.0 (2.4–8.8)</td>
<td>2.9 (2.0–4.3)</td>
<td>6.0 (2.9–12.3)</td>
</tr>
<tr>
<td>Negative LR</td>
<td>0.4 (0.27–0.83)</td>
<td>0.27 (0.1–0.6)</td>
<td>0.17 (0.05–0.6)</td>
</tr>
</tbody>
</table>

Data are presented as % (95% CI) or n (95% CI). NPV, negative predictive value; PPV, positive predictive value.
reflected by the large confidence intervals, may have affected the results. Further study with a more heterogeneous sample in public health clinics is necessary to replicate these results.

CONCLUSIONS

Given its psychometric properties and correlations with the Bayley-III, the ASQ-3 may be recommended for routine pediatric developmental screening, particularly in children with high biological risks for developmental delay. Continued research with diverse populations in Chile and in other middle-income countries is recommended.

ACKNOWLEDGMENTS

We thank the Ambulatory Pediatric Group and the Preterm Follow-up team at Clínica Alemana–Universidad del Desarrollo for their continuous support and collaboration, particularly Dr Marcela Perez and Dr Patricia Salinas. We thank all the families and children who participated in our study, for their patience and trust. We also thank Jane Squires and Teresa Cole for their support, advice, and help writing the manuscript.

REFERENCES


25. Gollenberg AL, Lynch CD, Jackson LW, McGuinness BM, Msall ME. Concurrent validity of the parent-completed Ages and Stages Questionnaires, 2nd Ed. with the Bayley Scales of Infant Development II in
Validity of the Ages and Stages Questionnaires in Term and Preterm Infants
Luisa Schonhaut, Iván Armijo, Marianne Schönstedt, Jorge Alvarez and Miguel Cordero

*Pediatrics* 2013;131:e1468; originally published online April 29, 2013; DOI: 10.1542/peds.2012-3313

Updated Information & Services
including high resolution figures, can be found at:
/content/131/5/e1468.full.html

References
This article cites 27 articles, 7 of which can be accessed free at:
/content/131/5/e1468.full.html#ref-list-1

Citations
This article has been cited by 14 HighWire-hosted articles:
/content/131/5/e1468.full.html#related-urls

Subspecialty Collections
This article, along with others on similar topics, appears in the following collection(s):
- Developmental/Behavioral Pediatrics
  /cgi/collection/development:behavioral_issues_sub
- Growth/Development Milestones
  /cgi/collection/growth:development_milestones_sub
- Fetus/Newborn Infant
  /cgi/collection/fetus:newborn_infant_sub
- Neonatology
  /cgi/collection/neonatology_sub

Permissions & Licensing
Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at:
/site/misc/Permissions.xhtml

Reprints
Information about ordering reprints can be found online:
/site/misc/reprints.xhtml
Validity of the Ages and Stages Questionnaires in Term and Preterm Infants
Luisa Schonhaut, Iván Armijo, Marianne Schönlstedt, Jorge Alvarez and Miguel Cordero

*Pediatrics* 2013;131:e1468; originally published online April 29, 2013;
DOI: 10.1542/peds.2012-3313

The online version of this article, along with updated information and services, is located on the World Wide Web at:
/content/131/5/e1468.full.html