Predictive value of Ages & Stages Questionnaires for cognitive performance during early child education

Validez del Ages & Stages questionnaires para predecir el desempeño cognitivo en los primeros años de educación escolar

Luisa Schonhaut B.a, Marcela Pérez R.a, Ana María Castilla F.b, Sonia Castro M.a,b, Patricia Salinas A.a, Iván Armijo R.c

aServicio de Pediatría, Clínica Alemana, Facultad de Medicina, Universidad del Desarrollo, Santiago, Chile
b“La Casa Azul”, Centro Integral de Atención Psiquiátrica y Psicológica, Santiago, Chile
cFacultad de Psicología, Universidad Gabriela Mistral, Santiago, Chile

Received: 20-11-2015; Accepted: 18-8-2016

Abstract

Introduction: The Ages and Stages questionnaires (ASQ) has been recently validated in our country for developmental screening. The objective of this study is evaluate the validity of ASQ to predict low cognitive performance in the early years of schooling. Patients and method: Diagnostic test studies conducted on a sample of children of medium-high socioeconomic level were evaluated using ASQ at least once at 8, 18 and/or 30 months old, and later, between 6 and 9 years old, reevaluated using the Wechsler Intelligence Scale for Children-third edition (WISC-III). Each ASQ evaluation was recorded independently. WISC-III was standardized, considering underperformance when the total score was under -1 standard deviation. Results: 123 children, corresponding to 174 ASQ assessments (42 of them were 8 months old, 55 were 18 months and 77 were 30 months of age) were included. An area under the ROC curve of 80.7% was obtained, showing higher values at 8 months (98.0%) compared to 18 and 30 months old (78.1 and 79.3%, respectively). Considering different ASQ scoring criteria, a low sensitivity (27.8 to 50.0%), but a high specificity (78.8 to 96.2%) were obtained; the positive predictive value ranged between 21 and 46%, while the negative value was 92.0-93.2%. Conclusion: ASQ has low sensitivity but excellent specificity to predict a low cognitive performance during the first years of schooling, being a good alternative to monitor psychomotor development in children who attend the private sector healthcare in our country.

Introduction

Identifying psychomotor development disorders is a fundamental component of health supervision for children to receive timely treatments. Given that the clinical impression is rather ambiguous, in particular when assessing children with no identifiable risk factors, the American Society of Pediatrics recommends a psychomotor development screening through standardized tests at 9, 18 and/or 30 months old, while a
similar practice is observed in the Primary Health Care of our country.4

There are different systems for measuring a psychomotor assessment; some must be applied by a trained professional, with the drawbacks of being a time-consuming task for the professional and a different experience for the child, while others can be implemented by the parents or caregivers in the natural environment of the child. The advantages of the latter are low cost, smooth implementation and parents’ involvement.5,7

One of the reporting methods is the Ages and Stages Questionnaires (ASQ), which was developed at the University of Oregon in the 1980s, and subsequently updated and validated in many countries.3,9 Over the last decade, it has gained widespread popularity in the US. ASQ consists of a series of questionnaires targeting different age groups, which screen 5 domains, communication, gross motor, fine motor, problem solving and personal-social.

Based on the original validation, a child is at risk of developmental problem, and referred to a more specialized assessment, in case of a performance below -2 standard deviations (SD) in at least one of the questionnaires’ segments.8 However, there is debate regarding this in recent research because with a performance below the cut-off in two or more domains, the ASQ specificity increases, while efficiency overall improves when the total score is considered. Independent from the approach, it has been observed that when parents conduct the assessment, the search for possible development disorders significantly increases, anticipating up to 30% referral increase among evaluated children. The current evidence is unclear as most of the research corresponds to concurrent studies; therefore, there is a need to design prospective research that helps to define a deficit criterion with a real clinical impact.

In our country, ASQ was validated in a broad sample community. In a sub-sample of medium-high socioeconomic level children with a seemingly typical development, the concurrent validity of ASQ was assessed with sensitivity and specificity values of 73% and 81% respectively. A positive screening was observed in 14% of the children; however, the future development of these children is unknown.

The objective of this study is evaluate the validity of ASQ to predict low cognitive performance in the early years of schooling in a sample of children of medium-high socioeconomic level, and to compare the diagnosis accuracy of many ASQ criteria as predictors of below average cognitive performance.

Patients and Method

This diagnostic test study was conducted on a sample of children of medium-high socioeconomic level who were evaluated using ASQ at least once (at 8, 18 and/or 30 months old), between 2008 and 2011, and later, between 6 and 9 years old, re-evaluated using the Wechsler Intelligence Scale for Children-third edition (WISC-III). Each ASQ evaluation was recorded independently.

Inclusion Criteria

Previously evaluated using ASQ at 8, 18 and/or 30 months old, without history of neurological, congenital or metabolic disease, or with a non-fluent Spanish speaker caregiver. In case of children born at less than 37 weeks gestational age, their 8 and 18 month-old ages were corrected, but at 30 months old, their chronological age was considered.

Exclusion Criteria

History of a breakthrough condition which may affect the development, such as meningitis, CNS tumor disease, vascular accident, severe head injury, diagnosed after the ASQ. Children that were living overseas or were 9 years old at the time of evaluation were also excluded.

Children with serious developmental disorders (who did not have a diagnosis at the time of ASQ), who were not in school attendance and/or able to participate in the evaluation, were not examined by WISC-III, but considered underperformance in the final analysis.

The parents of the children were contacted by phone and agreed to participate through an informed consent. Subsequently, they filled in a form regarding bio-demographic, neonatal and academic antecedents, plus possible developmental and cognitive-related diagnosis and interventions of the children. All children had at least one parent that attended college and belonged to families in the fifth income quintile.

Instruments

Ages and Stages questionnaires

A series of 21 questionnaires for children aged 2 months-5 years. Each questionnaire screens 5 domains, 6 questions each, evaluating different aspects of the psychomotor development, such as: comunicación, gross motor, fine motor, problem solving and personal-social. They are multiple choice questions with answers «yes» (10 points), «occasionally» (5 points) or «not yet» (0 point); the scores of each domain are added for a possible total of 60 points per area. At the end, a separate section of 7 open questions to explore possible parents’ concerns can be found; however, they were not considered in this analysis. The parents filled up the ASQ forms based on the age group. The questionnaires were validated in our country by Schonhaut and Armijo (2013 y 2015).
**Wechsler Intelligence Scale for Children-third edition**

It is an individually administered clinical instrument to evaluate the cognitive performance of children between the ages of 6 and 16 and 11 months, using 13 subtests, 6 verbal scale tests and 7 performance scales. Children performance is summarized in 3 composite scores, verbal, performance and total IQ coefficients. This Scale was standardized and adapted in Chile by Ramírez and Rozas in 2007. Only the total IQ coefficient was considered in this analysis.

WISC-III was implemented by a group of ASQ result-blind trained psychologists. The evaluations where performed either at the doctor’s office or at the patients’ home, and the scores were given after a consensual review by the participating professionals. In the end, the parents received a report of the children performance.

**Definition of psychomotor development delay risk according to Ages and Stages questionnaires**

Total ASQ scores were standardized and transformed into a Z-score in each age group to be analyzed as a whole. Various delay criteria were analyzed: performance < –2 standard deviation (SD) at least in one developmental area; performance < –2 standard deviation (SD) in 2 or more areas; total score < –2 DS.

**Definition of low performance according to the Wechsler Intelligence Scale**

The WISC-III score obtained standardized for the simple under study and a low performance was defined as the total score was 1 SD below average.

**Statistical Analysis**

The normality of the distribution of WISC-III and ASQ scores was analyzed using the Shapiro-Wilk test. WISC-III standardized scores, separated in two groups: underperformance and normal performance were used as development consolidation indicators. ROC curves were created to determine ASQ scores according to age group, which improve diagnosis accuracy in low performance cases.

Although there is no absolute standard to interpret the area under the curve (AUC), it is widely accepted that an AUC of 0.5 is a test without predictive value; 0.6 and 0.7 values are considered acceptable; 0.7 to 0.9 as good and higher than 0.9 as excellent. DeLong test was used to compare the relative precision obtained among different age groups when identifying cases of low performance, comparing the AUC of all ROC curves obtained.

ASQ psychometric properties were analyzed (sensitivity, specificity and predictive values) using various developmental delay risk criteria previously defined.

Analyses were perform using R Statistical, specifically with the implementation of the pROC module.

---

**Results**

123 children with history of at least one previous ASQ assessment were included. The WISC test was used on 121 children, and 2 children with diagnosis of developmental disorders were not evaluated, but considered in the final analysis, therefore, a total of 174 ASQ evaluations were analyzed. Among these children, 73 were full term babies, 34 had been born preterm between 32 and 36 weeks and 16 had been born before 32 weeks of gestational age (table 1).

Of the total number of ASQ evaluations, 42 were performed at 8 month old, 55 at 18 months old and 77 at 30 months old. 7 children had 3 evaluations, 37 had 2 and the rest had been evaluated once (table 1).

The total score obtained using the WISC-III scale presented a normal distribution, with an average value of 114.8 and a standard deviation of 12.8 (Shapiro-Wilk statistics test for normality, p = 0,19). 15 children met the low performance criteria with the WISC-III scale (13 with performance below –1 SD and two children could not be evaluated), corresponding to a 12.2% of the sample.

**Table 1.**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Número de niños incluidos</strong></td>
<td>123</td>
<td>(100)</td>
</tr>
<tr>
<td><strong>Género</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Masculino</td>
<td>70</td>
<td>(56,9)</td>
</tr>
<tr>
<td>Femenino</td>
<td>53</td>
<td>(43,1)</td>
</tr>
<tr>
<td><strong>Edad gestacional</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNT</td>
<td>73</td>
<td>(59,3)</td>
</tr>
<tr>
<td>PMT</td>
<td>34</td>
<td>(27,6)</td>
</tr>
<tr>
<td>PE</td>
<td>16</td>
<td>(13,0)</td>
</tr>
<tr>
<td><strong>Gemelar</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sí</td>
<td>33</td>
<td>(26,8)</td>
</tr>
<tr>
<td>No</td>
<td>90</td>
<td>(73,2)</td>
</tr>
<tr>
<td><strong>Número de evaluaciones con ASQ por niño</strong></td>
<td>174</td>
<td>(100)</td>
</tr>
<tr>
<td>1</td>
<td>79</td>
<td>(64,2)</td>
</tr>
<tr>
<td>2</td>
<td>37</td>
<td>(30,0)</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>(5,7)</td>
</tr>
<tr>
<td><strong>Total evaluaciones</strong></td>
<td>174</td>
<td>(100)</td>
</tr>
<tr>
<td><strong>Edad evaluación</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 meses</td>
<td>42</td>
<td>(24,1)</td>
</tr>
<tr>
<td>18 meses</td>
<td>55</td>
<td>(31,6)</td>
</tr>
<tr>
<td>30 meses</td>
<td>77</td>
<td>(44,3)</td>
</tr>
</tbody>
</table>

Note: PE: prematuros extremos (edad gestacional < 32 semanas de gestación); PMT: prematuros moderados y tardíos (32-36 semanas de edad gestacional); RNT: recién nacidos a término (37-42 semanas de gestación).
In ROC curve analysis, the total ASQ scores produced an area under the curve of 80.7% (CI 95%: 71.8 to 89.5). Figure 1 shows ROC curves at ages when ASQ took place with the low WISC-III performance as reference; a great precision was observed at 8 months old with an area under the ROC curve of 98.0% (CI 95%: 94.3 to 100), compared to 18 (78.1%) (CI 95%: 60.1-96.0) and 30 months (79.3%) (CI 95%: 67.3 to 91.4) where precision was overall good. A comparison of curve precision using DeLong test showed significant differences between: 8 versus 18 months (D = –2.13, gl = 58.7, p = 0.037); 8 versus 30 months (D = –2.90; gl = 90.0; p = 0.004); but no significant differences between 18 versus 30 months were observed (D = –0.12; gl = 99.16; p = 0.907).

A comparison of different psychomotor development delay risk criteria showed no significant differences among ASQ psychometric properties. When considering total score < –2 SD, sensitivity values of 33.3% and specificity values of 92.9% were obtained; when at least 2 areas below the cutoff point were considered, sensitivity was 27.8% and specificity 96.2, while with at

### Table 2.

<table>
<thead>
<tr>
<th>Criterio ASQ</th>
<th>Sensibilidad</th>
<th>Especificidad</th>
<th>VPP</th>
<th>VPN</th>
<th>Sobre-referencia</th>
<th>Infra-referencia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Análisis puntaje total Z &lt; –2</td>
<td>33,3 (14,4-58,9)</td>
<td>92,9 (87,4-96,3)</td>
<td>35,3 (15,3-61,3)</td>
<td>92,4 (86,7-95,8)</td>
<td>64,7 (32,9-82,5)</td>
<td>7,6 (4,4-14,0)</td>
</tr>
<tr>
<td>Al menos un área &lt; –2 DS</td>
<td>50,0 (26,8-73,2)</td>
<td>78,8 (71,4-84,8)</td>
<td>21,4 (10,8-37,2)</td>
<td>93,2 (87,1-96,6)</td>
<td>78,5 (62,8-89,2)</td>
<td>6,8 (3,4-12,9)</td>
</tr>
<tr>
<td>Dos o más áreas &lt; –2 DS</td>
<td>27,8 (10,7-53,6)</td>
<td>96,2 (91,4-98,4)</td>
<td>45,5 (18,1-75,4)</td>
<td>92,0 (86,4-95,5)</td>
<td>54,5 (24,6-81,9)</td>
<td>8,0 (4,5-13,5)</td>
</tr>
</tbody>
</table>

DS: desviación estándar; VPN: valor predictivo negativo; VPP: valor predictivo positivo.
least one area below the cutoff point, sensitivity raised to 50.0% and specificity decreased to 78.8%. Positive predictive value ranged from 21.4% to 45.5%, and over-referral percentages varied from 54.5% to 78.5%. Negative predictive values ranged from 92.0% to 93.2% in the 3 analysis criteria (table 2).

Discussion

An area under the ROC curve of 80.7% was observed in this study, a good overall discriminative capacity of the ASQ to predict a lower cognitive performance during early child education. This was reflected in adequate specificity values (78.8% to 96.2%), but low sensitivity (27.8% to 50.0%), regardless of the risk criterion used.

These results are not coincident with the studies by Halbwachs et al., who reported a high sensitivity (80%) and moderate specificity (54%) in a sample of premature infants26, whereas Kerstjens et al., based on the area under the cut-off point criterion, obtained 89% and 80% sensitivity and specificity, respectively, to predict the need for special education within a year of follow-up27. No other studies that show the predictive validity of ASQ to foresee lower cognitive performance in the long term have been published.

In our analysis, low positive predictive values (21.4% to 45.5%) were overall observed, with a high percentage of over-referral, 2 areas under the cut-off point criterion produced the closest rate of over-referral to the expected 30%6,16. The problem of over-referral is its high economic and family costs involved28, and by Doyle using the Bayley scale29,30, a tendency to improve the ASQ psychometric properties as children age increased was not observed. At 8 months, a ROC curve with an excellent discriminative capacity (AUC 98%) was obtained compared to 18 and 30 months where AUC was overall good (78.1% and 79.3%, respectively). Considering that the lower limit of confidence interval exceeded 50% in all cases, it is suggested that ASQ nonrandomly identifies children with possible developmental disorders. At 8 months old, gross motor development disorders were predominant compared to other domains31, which shows how intertwined the different areas of development are.

In order to obtain a correct analysis of our results, the fact that the evaluated children grew up in a culturally enriched environment and had received other evaluations and interventions, not analyzed in this study, should be considered as it could be associated with better cognitive and/or academic performance32,33. The high overall performance of children, with a low prevalence of real developmental problems, could explain the low positive predictive values and the high over-referral obtained in our study. Another limitation is the fact that it is an opportunity sample, not allowing statistical inference. Also, the fact that only cognitive development is measured without evaluating other domains of development, such as motor or social-emotional performance can be questionable, aspects not considered in the objectives of this study and coincident with other publications on the subject26,27.

We concluded that, in the sample under study, ASQ has low sensitivity but excellent specificity to predict a lower cognitive performance during early child education. Considering the high percentage of over-referral, we recommend that a child be referred for a more comprehensive assessment if he or she has 2 or more areas below the cut-off point. Children who have a delay in one area, a stimulation plan and a subsequent reassessment with ASQ are suggested. Our results confirm the importance of early developmental assessment, and ASQ is a good alternative to monitor the psychomotor development of children who attend the private sector healthcare in our country.

Ethical Responsibilities

Protection of people and animals: The authors reported that the procedures followed are in accordance to the ethical standards of the responsible human experimentation committee and in agreement with the World Medical Association and the Declaration of Helsinki.

Confidentiality of the data: The authors confirmed that they have followed the protocols of their center regarding the publication of patient data.

Privacy rights and informed consent: The authors have obtained the informed consent of the patients and/or subjects referred to in the article. These documents are in the possession of the corresponding author.

Funding

Project funded by Internal Project Contests of Universidad del Desarrollo (ID 20131009161129950210) and Project Contests of Sochipe 2012.

Conflict of interests

The authors declare no conflict of interest.

Acknowledgements

We thank the psychologists Loreto González, Constanza Osorio, María Teresa Burich, Pía Margate, Daniela Werner, Josefina Jirón and Bernardita Martínez for their support, love and dedication evaluating the children.
References


