Original Article

Preliminary Results of the Macedonian-Adapted Version of Ages and Stages Developmental Questionnaires

Blagica Stanoevska, Luis Anunciação¹, Jane Squires², Ajay Singh³, Vladimir Trajkovski⁴

Context: Early detection of developmental problems is critical, and interventions are more effective when they are carried out early in a child’s life. In Macedonia, there are only four centers providing early intervention services. Aims: In this research, we determined the reliability of the translation and adaptation of Ages and Stages Questionnaires 3rd edition (ASQ-3-M) for assessment of children aged 3–5 years old in Macedonia, and reported preliminary results of the gender differences in the development. Materials and Methods: ASQ-3-M was completed by 165 parents and 40 educators in seven kindergarten classrooms. Children were 3–5 years old. Statistical Analysis Used: Cronbach’s alpha, Intraclass Correlation coefficient (ICC), and interrater reliability (IRR) were used to assess ASQ-3-M psychometric properties. The Bayesian t-test was performed to estimate the difference in means between males and females. Results: The Cronbach’s alpha ranged from 0.65 to 0.87. The overall ICC was 0.89 (ranged from 0.8 to 0.95), which indicates a strong to almost perfect strength of agreement between test-retest. IRR correlation revealed an average of 0.88 (ranged from 0.74 to 0.95), suggesting that ASQ-3-M is reliable and stable. Conclusions: The results from the comparison between males and females on all dimensions of ASQ-3-M were not statistically significant (BF10 <3), indicating no significant gender difference. That said, the ASQ-3 is recommended for routine use in screening children aged 3–5 years old.

Keywords: Ages and stages questionnaire, Bayesian analysis, child development, developmental assessment, Macedonia

INTRODUCTION

Early experiences affect the development of the brain and have a direct impact on how children develop their sociability, self-expression, independence, initiative, and social and emotional skills.¹-³ Once infancy is a crucial time for both physically and mentally in every human’s life, measuring the development in young children promotes information that can be used to determine which groups of children are eligible for receiving further assistance and provides the opportunity for children to benefit from early intervention.

Several approaches have been used to assess young children’s developmental outcomes and the results of psychometric testing, in general, and screening tools, in specific, accurately reflect children development, when the tool administered has adequate psychometric properties. From this standpoint, psychometric instruments must guarantee their validity and reliability. Validity refers to the accuracy of measurement for a specific purpose and is the extent to which the test measures what it says it measures. Reliability refers the degree to which an assessment instrument produces...
stable and consistent results each time it is used in the
same setting with the same type of subjects.[4]

In spite of that, there is a lack of a screening and
eye identification system for preschool children in
the Republic of Macedonia. Children are often not
identified as having disabilities until they are 3 years
or older and sometimes, not even until they start school
(Dimitrova-Radojichikj, Chichevska-Jovanova, and
Rashikj-Canevska, 2016). This situation can be partially
explained because in Macedonia, there are only four
centers that provide early intervention services and
because of the absence of psychometric screening tools:
Even if a wide range of screening tools is available to
assess child development, before a local adaptation, their
use is limited to the population from whom they were
developed.[5]

The process of developing, adapting, and implementing
screening tools to use with Macedonian children is vital
to improve services for young children and help them
to reach their full potential. With that said, the purpose
of this study is to verify the reliability of the translated
and adapted version of Ages and Stages Questionnaires,
3rd Edition (ASQ-3-M) with data gathered in Republic
of Macedonia.

MATERIALS AND METHODS

Participants

This study included 165 Macedonian children at the
age of 36, 48, and 60 months, their parents and their
kindergarten teachers. Participants were randomly
selected from 7 kindergartens in different parts in
Skopje.

Measures: Ages and Stages Questionnaires,
3rd Edition

ASQ-3, was chosen as a study target because is currently
the most used parent-completed developmental screener
in the world[6] and has been successfully studied in
different countries and cultures.[7,9]

The ASQ-3 consists of 21 questionnaires, for children
from 2 to 66 months. Each questionnaire contains 30
questions composed by 6 questions for each domain:
communication, gross motor, fine motor, problem
solving, and personal-social. ASQ questions are
written at the fourth to sixth grade reading level and
can be administered as an interview for parents with
low literacy levels and include illustrations that assist
parents in understanding items. Parents indicate “yes,”
“sometimes,” or “not yet” in response to each item, with
point values of 10, 5, or 0, respectively.[10,11]

Each domain score is independently and there are
empirically derived cutoff scores to indicate whether the
questionnaire falls within a normal developmental range
based on chronological age, or if the score represents “at
risk” or delayed development. The ASQ requires about
15 min to complete and 2–3 min to score.[12]

Children whose scores fall within the “typical” range or
above the cutoff scores are considered to be developing
appropriately and should continue the screening
process at regular intervals. Children whose scores fall
below the cut off score in any developmental area are
recommended to receive further assessment. If a child’s
scores fall within the monitoring zone (1–2 standard
deviations below the mean), specialized activities and
repeat screening are recommended.[11,12]

Procedures

The ASQ was translated into Macedonian
(i.e., ASQ-3-M) and necessary cross-cultural adaptations
were made, and then it was back translated. The
accuracy of the translation was evaluated and changes
were made, when necessary, by members of the research
team who were very proficient in both languages.

After this, the translated version was distributed and
completed by 165 parents and by 40 teachers in their
seven kindergarten classrooms over a 6 months’ time
period. Questionnaires were distributed to teachers and
parents at the same time. A demographic form attached
to the questionnaire asked for general family information
on the child’s gender, date of child’s birth, parents’ age,
education level, gender, and nationality. The researcher
personally traveled to all kindergartens and had meetings
with directors, teachers, and the staff to explain the
study, its goals, procedures, and steps. After teachers had
given permission to participate, the study was presented
to parents, and they gave their permission if they were
willing to participate.

When parents returned the questionnaire, they received
a duplicate copy of the same ASQ-3-M corresponding to
their child’s age to complete. They were distributed by
researcher and teacher only in hard copy (some of the
questionnaires for parents were distributed via teachers
to the parents). Teachers and the researcher collected the
questionnaires and remained parents to complete both
questionnaires throughout the process. The researcher
entered all study data into a secure database.

Statistical analysis

Internal consistency reliability was assessed by
calculating the Cronbach’s $\alpha$ coefficient and results
equal or above 0.7 are considered acceptable.[13] To
assess test-retest reliability, the parents were asked
to complete the ASQ-3-M on their child and then, to
complete the same questionnaire 2 weeks apart, blind
to the outcomes of the first result. The intraclass
correlation coefficient (ICC) was computed with the results. An ICC between 0.75 and 0.90 indicates good test-retest reliability. Interrater reliability (IRR, a.k.a., Interobserver reliability) was estimated by using Pearson Product Moment Correlation in the results from the questionnaires completed by parents and kindergarten teachers for the same children.

To investigate how much males and females’ results were different one from another in all domains of ASQ-3-M, since the small sample in this research, a Bayesian t-test was performed using with a prior width set at $r = 0.707$, which is recommended as the default value for this test. Bayes factors provide a numerical value that quantifies how well a hypothesis predicts the empirical data relative to a competing hypothesis. The Bayes Factor ($BF_{10}$) expresses the probability of the data in favor of $H_1$ relative $H_0$. If the $BF_{10}$ is equal to 1, it indicates that the observed finding is equally likely under both hypotheses, if $BF_{10}$ is >1 then the data provide support for the alternative hypothesis and if $BF_{10}$ is <1 then the data provide support for the null hypothesis.

The significance threshold for Bayesian inference was set at $BF_{10}>3$ and for frequentist inference was set at $P < 0.05$. There was no missing value in dataset and data were analyzed by using R statistical software version 3.4.4 with tidyverse, BayesFactor and psych packages.

**RESULTS**

Of the 210 participants’ approaches to take part in this study, 165 (78.6%) completed the ASQ: 3-M. These results are presented in Table 1, which show the absolute frequency and percentage of each included sociodemographic variable.

The reliability was estimated by three different ways: Internal consistency with Cronbach’s alpha, test-retest reliability by ICC and IRR by Pearson Product-Moment correlation. All results were significant, and the outcomes are reported in Table 2.

Cronbach alpha is a test of internal consistency and the higher is the score, the better is the internal consistency. Test-retest provides clinicians with the assurance that the instrument measures the outcome the same way each time it is performed. Better reproducibility, measured by higher ICC, suggests better precision of single measurements, which is a requirement for better tracking of changes in measurement. IRR assess the degree to which different raters/observers give consistent estimates of the same phenomenon.

To decide which of two hypotheses is more likely given an experimental result, one must consider the ratios of their likelihoods. This ratio is called the Bayes Factor (BF), and $BF_{10}$ reflects the likelihood of $H_1$ compared to $H_0$ given a set of data. BF is a

### Table 1: Reporting of demographic variables

<table>
<thead>
<tr>
<th></th>
<th>36 months, n (%)</th>
<th>48 months, n (%)</th>
<th>60 months, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>24 (50)</td>
<td>28 (44)</td>
<td>27 (50)</td>
</tr>
<tr>
<td>Female</td>
<td>24 (50)</td>
<td>35 (56)</td>
<td>27 (50)</td>
</tr>
<tr>
<td>Total</td>
<td>48</td>
<td>63</td>
<td>54</td>
</tr>
<tr>
<td>Education of who completed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>17 (35)</td>
<td>20 (32)</td>
<td>17 (31)</td>
</tr>
<tr>
<td>4 years college</td>
<td>25 (52)</td>
<td>33 (52)</td>
<td>31 (57)</td>
</tr>
<tr>
<td>Above college</td>
<td>6 (12)</td>
<td>10 (16)</td>
<td>6 (11)</td>
</tr>
<tr>
<td>Nationality of who completed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macedonian</td>
<td>43 (90)</td>
<td>58 (92)</td>
<td>48 (89)</td>
</tr>
<tr>
<td>Non-Macedonian</td>
<td>5 (10)</td>
<td>5 (8)</td>
<td>6 (11)</td>
</tr>
</tbody>
</table>

### Table 2: Reliability results for all 5 dimensions

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Internal consistency</th>
<th>Intra-rater reliability</th>
<th>Inter-rater reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cronbach's alpha</td>
<td>ICC (3)</td>
<td>Pearson correlation</td>
</tr>
<tr>
<td></td>
<td>36 48 60</td>
<td>36 48 60</td>
<td>36 48 60</td>
</tr>
<tr>
<td>Communication</td>
<td>0.83 0.86 0.77</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>Gross motor</td>
<td>0.83 0.87 0.77</td>
<td>0.82</td>
<td>0.95</td>
</tr>
<tr>
<td>Fine motor</td>
<td>0.65 0.78 0.72</td>
<td>0.85</td>
<td>0.95</td>
</tr>
<tr>
<td>Problem solving</td>
<td>0.68 0.76 0.69</td>
<td>0.86</td>
<td>0.95</td>
</tr>
<tr>
<td>Personal/social</td>
<td>0.77 0.7 0.69</td>
<td>0.86</td>
<td>0.95</td>
</tr>
</tbody>
</table>

ICC: Intraclass correlation coefficient
continuous measure of evidence and results with values between 1 and 3 are seems as inconclusive, values between 3 and 10 are weak and >10 are generally seem as strong evidence. Results are reported in Table 3.

Despite the numerical difference, there were no evidence of a statistically difference between males and females in terms of their abilities. Figure 1 displays the mean scores of all variables of ASQ-3-M for males and females.

**DISCUSSION**

To the best of our knowledge, this is the first study of the Macedonian version of ASQ-3, as well the first to investigate the gender differences in response to it. In terms of cultural and linguistic appropriateness for Macedonian children, some items underwent modifications to ensure words and phrases are semantically similar, with a minimum of wording changes. A similar process of item adaptation happened in other countries, such as Brazil, Chile and China, which points to the importance of the cross-cultural adaptation of psychological instruments before its implementation.

The internal consistency of the questionnaire ranged between 0.65 and 0.87. These results were higher than studies in Portugal, but were similar to other international studies, such as those conducted in China and in Arabic countries. This outcome indicates that although some items were changed, they retained their concepts and did not lose their original meanings. Test-retest reliability measures the stability of test outcomes over time and was estimated by having parents complete two ASQ-3-M on the same child within a 2 weeks’ interval. The ICC between scores of the two ASQ-3-M concludes the stability of the scores of ASQ-3-M. Similar findings were found in other studies.

The IRR refers to the agreement of test outcomes and was examined using Person Product Moment. The results between parents and kind garden teachers were lower than the original ASQ-3 study, but showed a strong agreement between both parents and teachers. The literature about the agreement between professionals and parents led both results (high and low agreement) and this evidence and this results provides supporting to evidence of the agreement regarding child development.

There were no significant differences between the groups with respect to all scales of ASQ-3-M. However, although these results did not show a statistically significant difference, the scores for males and females were within the expected range.

**Table 3: Mean, standard deviation and hypothesis testing**

<table>
<thead>
<tr>
<th>Months</th>
<th>Communication</th>
<th>Gross motor</th>
<th>Fine motor</th>
<th>Problem solving</th>
<th>Personal/social</th>
</tr>
</thead>
<tbody>
<tr>
<td>36 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>53.96±8.21</td>
<td>55.21±7.73</td>
<td>55.33±5.2</td>
<td>56.25±4.95</td>
<td>58.12±4.62</td>
</tr>
<tr>
<td>Female</td>
<td>56.67±5.84</td>
<td>56.67±5.84</td>
<td>56.67±5.34</td>
<td>56.88±4.38</td>
<td>58.75±2.66</td>
</tr>
<tr>
<td>BF (%)</td>
<td>0.58±0.01</td>
<td>0.36±0.02</td>
<td>0.29±0.02</td>
<td>0.31±0.02</td>
<td>0.329±0.02</td>
</tr>
<tr>
<td>P</td>
<td>0.19</td>
<td>0.46</td>
<td>0.83</td>
<td>0.65</td>
<td>0.57</td>
</tr>
<tr>
<td>48 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>57.5±5.69</td>
<td>57.5±5.69</td>
<td>52.5±7.01</td>
<td>55.18±6.73</td>
<td>54.46±5.67</td>
</tr>
<tr>
<td>Female</td>
<td>57.43±5.99</td>
<td>57.43±5.99</td>
<td>53±8.51</td>
<td>56±5.79</td>
<td>52.71±7.41</td>
</tr>
<tr>
<td>BF (%)</td>
<td>0.26±0</td>
<td>0.26±0</td>
<td>0.27±0</td>
<td>0.29±0</td>
<td>0.40±0</td>
</tr>
<tr>
<td>P</td>
<td>0.96</td>
<td>0.96</td>
<td>0.80</td>
<td>0.61</td>
<td>0.29</td>
</tr>
<tr>
<td>60 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>57.41±4.01</td>
<td>57.41±4.01</td>
<td>54.81±5.09</td>
<td>55.56±4.46</td>
<td>55.19±4.9</td>
</tr>
<tr>
<td>Female</td>
<td>55.93±6.66</td>
<td>55.93±6.66</td>
<td>53.3±7.47</td>
<td>55±6.04</td>
<td>53.33±6.2</td>
</tr>
<tr>
<td>BF (%)</td>
<td>0.41±0.01</td>
<td>0.41±0.01</td>
<td>0.38±0.01</td>
<td>0.29±0.01</td>
<td>0.51±0.02</td>
</tr>
<tr>
<td>P</td>
<td>0.33</td>
<td>0.33</td>
<td>0.40</td>
<td>0.70</td>
<td>0.23</td>
</tr>
</tbody>
</table>

SD: Standard deviation, BF: Bayes factor
significant difference between genders on any of the ASQ-3-M domains, this could be seen as result of having small sample. Variations between men and women development can be found in the literature[12] and in other studies that used ASQ-3 was measurement tool.[23,24]

Some limitations are presented in these results. This study included only children at preschool age (36, 48, and 60 months) and only in Skopje. For a valid normative study, a larger sample of at least 100 children at each age interval is needed, stratified on a national Macedonia sample. In addition, only paper questionnaires were used in this study and to truly reach diverse young children; an online system would be preferable, as this would reduce the time and energy needed for data collection in individual kindergarten classrooms throughout the country. However, the impact of the limitations was small, as can be seen in the findings of this study.

These issues aside, as mentioned earlier, the basis of adult health and well-being are built prenatally and during early childhood.[1,3,25] The availability of ASQ: 3-M opens up possibilities for examining child development in diverse cultural groups, as well as for effective cross-cultural comparisons of development.

It was concluded that the Macedonian version of the ASQ-3 questionnaire was successfully translated and adapted, with good internal consistency and reliability, which provide solid support for the use of the scale to measure domains of child development among children in Macedonia. While this study has demonstrated the potential ASQ-3-M, it also being extended in longitudinal and comparative ways.

Financial support and sponsorship
Authors would like to say thanks to "Singh & Yeh Foundation, USA" for partial financial support to this research study.

Conflicts of interest
Dr. Squires is an author of the ASQ and receives some royalties from its publication.

REFERENCES