Validity and Reliability Determination of Denver Developmental Screening Test-II in 0-6 Year-Olds in Tehran

Soheila Shahshahani¹, MD; Roshanak Vameghi², MD; Nadia Azari¹, MD; Firoozeh Sajedi*², MD; and Anooshirvan Kazemnejad³, PhD

1. Pediatric Neurorehabilitation Research Center, University of Social Welfare & Rehabilitation Sciences, Tehran, IR Iran
2. Clinical Sciences Department, Pediatric Neurorehabilitation Research Center, University of Social Welfare & Rehabilitation Sciences, Tehran, IR Iran
3. University of Tarbiat Modaress, Tehran, IR Iran

Abstract

Objective: This research was designed to identify the validity and reliability of the Persian version of Denver Developmental Screening Test II (DDST-II) in Iranian children, in order to provide an appropriate developmental screening tool for Iranian child health workers.

Methods: At first a precise translation of test was done by three specialists in English literature and then it was revised by three pediatricians familiar with developmental domains. Then, DDST-II was performed on 221 children ranging from 0 to 6 years, in four Child Health Clinics, in north, south, east and west regions of Tehran city. In order to determine the agreement coefficient, these children were also evaluated by Ages and Stages Questionnaires (ASQ) test. Because ASQ is designed to use for 4-60 month-old children, children who were out of this range were evaluated by developmental pediatricians. Available sampling was used. Obtained data was analyzed by SPSS software.

Findings: Developmental disorders were observed in 34% of children who were examined by DDST-II, and in 12% of children who were examined by ASQ test. The estimated consistency coefficient between DDST-II and ASQ was 0.21, which is weak, and between DDST-II and the physicians’ examination was 0.44. The content validity of DDST-II was verified by reviewing books and journals, and by specialists’ opinions. All of the questions in DDST-II had appropriate content validity, and there was no need to change them. Test-retest and Inter-rater methods were used in order to determine reliability of the test, by Cronbach’s α and Kauder-Richardson coefficients. Kauder-Richardson coefficient for different developmental domains was between 61% and 74%, which is good. Cronbach’s α coefficient and Kappa measure of agreement for test-retest were 92% and 87% and for Inter-rater 90% and 76%, respectively.

Conclusion: Persian version of DDST-II has a good validity and reliability, and can be used as a screening tool for developmental screening of children in Tehran city.

Key Words: Development; Developmental screening; DDST-II; Denver; ASQ test
**Introduction**

Developmental disabilities can be seen in 10-15% of children in different populations. Early detection and appropriate referral of children with developmental delays or disorders is important in Pediatrics. This is only possible by continuous developmental monitoring and assessment. Developmental assessment is made by early detection of problems through developmental surveillance and screening, precise evaluation by using standardized and formal diagnostic tools as well as evaluation of the medical, social, family history and physical examination of the child\[1,2\]. Developmental screening must be repeated periodically and incorporated into pediatrics practice\[3,4\].

Developmental screening test is a brief standardized tool that is used for identifying children who need more detailed evaluation\[5\] and if used appropriately is useful and cost benefit effective\[6\]. Because screening is used for identifying the children who will receive the benefits of more professional evaluation or treatment, it is recommended that all children be screened for developmental delays\[5\].

There are many developmental screening tools. The base of all of them is achieving developmental milestones at specific chronological ages. Denver Developmental Screening Test II (DDST-II) and Bayley are examples for such formal tools. For having ability to differentiate between abnormal children from those normal children who have slower rate of achieving developmental skills, these developmental screening tools must be reliable and valid, have acceptable sensitivity and specificity, be easy to perform and not expensive\[1,6,7\].

DDST-II is a formal developmental screening tool that assesses children from birth to 6 years of age. First it was standardized on 1036 children (543 boys and 493 girls) from 2 weeks old to 6/4 years of age in Denver, Colorado as DDST\[8\]. Then in 1992 it is revised and restandardized on 2096 children and is known as DDST-II. Test reliability on test-retest is 90% and its inter-rater reliability is 80-95%\[9\]. The test is valid and there is a strong relationship between classification on the DDST and scores on the Stanford-Binet intelligence scales and the Previous edition of Bayley infant scales\[10\].

DDST-II is a brief and validated screening tool that many of pediatricians are familiar with it. Although there is doubt about its limited specificity (43%) and risks of over referral\[5,11\], it has high rate of sensitivity (83%) and identifies children with developmental delays\[5,12\]. DDST-II assesses child’s development in 4 general areas: 1) personal–social (25 items), 2) fine motor- adaptive (29 items), 3) language (39 items), and 4) gross motor (32 items)\[9,13\]. Screening by it produces 3 scores: normal, suspect and untestable\[9\] (these children refused participating in some items that 95% of age matched children could pass them). Sometimes DDST results are interpreted as normal, suspect, questionable (these children cannot pass some items that 75-95% of age matched children could pass them) and untestable. A study found sensitivity of 80% if "questionable scores" were included with abnormal scores but specificity of 46%. Alternatively, if "questionable scores" were included with normal scores, sensitivity was 46% and specificity 80\[11\].

By considering the importance of early detection of developmental disabilities and absence of an Iranian developmental screening test, this study was planned to determine the validity and reliability of Persian version of DDST-II (by translating to Persian and evaluating the cultural adaptation of the items) in Iranian children in order to provide an appropriate developmental screening tool for Iranian child health workers.

**Subjects and Methods**

This research is an action research that was performed from January to August 2008 in 4 Child Health Care centers located in north, south, east and west regions of Tehran city.

These are primary health care centers which provide mainly general health services for people including children from different socio-economical classes of general population. Usually normal children visit such centers and services for growth monitoring, vaccination, vitamin supplements, etc). These centers are
under the supervision of Deputy of Health, or Shahid Beheshti Medical University.

At first test form and guiding sheet was translated precisely by 3 specialists familiar with English. Then the research team (4 pediatricians) read all 3 translated versions and for each item in form and sheet we chose the best translation (simple, short, easy to understand and culturally compatible). Then we sent them along with original version to 3 other pediatricians who were familiar with developmental domains. The research team discussed their view points and implemented their opinions in the final form. Healthy newborns, infants and children, 0-6 years old, in Tehran city could participate in this study. The inclusion criteria were: 1) age between birth to 6 years, 2) Iranian nationality, 3) living in Tehran city, and 4) parental cooperation. Exclusion criteria were: 1) having obvious developmental delay or disability (because including children with gross developmental disorders would lower the cutoff point for each developmental item in Iranian children), 2) parental refusal.

The study was approved by the research committee and thereafter by the ethical committee of University of Social Welfare & Rehabilitation Sciences. Consent for participation was obtained from parents. The parents whose children had developmental problems were informed and guided.

Convenient sampling was used and 221 children (100 girls and 121 boys) in 13 age groups (0 to 2, 2.1 to 4, 4.1 to 6, 6.1 to 9, 9.1 to 12, 12.1 to 15, 15.1 to 18, 18.1 to 24, 24.1 to 30, 30.1 to 36, 36.1 to 48, 48.1 to 60 and 60.1 to 72 months), each age group containing 17 children, were examined. Demographic items included date of birth, sex, birth order, maternal education level, gestational age at birth (preterm or term; for preterm children up to 2 years we calculated corrected age), and history of disability of the child.

Eight examiners were trained in a 1 day workshop for performing the DDST-II. A demographic questionnaire was completed for each child by parents and then DDST-II was done by the examiners. In order to determine the reliability of DDST-II, 25% of children in each age group (small children after 30-60 minutes, older children up to 2-3 days later) were re-examined by the same examiners (test-retest).

Another 25% of children were retested by another examiner (inter-rater reliability). In order to determine agreement coefficient, these children were also evaluated by ASQ (Ages and Stages Questionnaires) test. ASQ is not a diagnostic gold standard test. It is a developmental screening tool. Because we had no accessibility to any diagnostic tests we compared these two developmental screening tools to determine their agreement coefficient. Anyway, by another research team, ASQ was translated into Persian and was standardized on 11000 Iranian children. The results have not been published yet, but the general report exists and we have used the translated forms. Because ASQ is designed to use for 4-60 month-old children and each questionnaire can be used for one month before or after the specific age, children who were out of this range (3-61 months) were evaluated by developmental pediatricians. Also 10% of children of other age groups, after examining by DDST-II and completing the ASQ by parents, were evaluated by developmental pediatricians.

As mentioned above, test-retest and inter-rater methods were used in order to determine reliability of the test by Cronbach’s α and Kauder-Richardson coefficients. We use Cronbach’s α for reliability determining of each test item and Kauder-Richardson coefficient in 4 developmental domains. In test-retest and inter-rater tests we measured Cronbach’s α and kappa measure of agreement for comparison of each developmental domain and final results of each test respectively. Content validity of the test was verified by reviewing texts and related articles, and by specialists’ opinions. Data was analyzed by SPSS software.

**Findings**

In this study 221 children were evaluated by DDST-II (100 girls and 121 boys) in 13 age groups (Appendix). Birth order of children were 73% first, 22% second, 3% third and 2% fourth
Table 1: Interclass Correlation Coefficients for test – retest examination by DDST-II

<table>
<thead>
<tr>
<th>Developmental domain</th>
<th>ICC*</th>
<th>ICC interval (95% confidence interval)</th>
<th>Cronbach’s α coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal-social</td>
<td>0.95</td>
<td>0.92-0.97</td>
<td>0.96</td>
</tr>
<tr>
<td>Fine motor-adaptive</td>
<td>0.90</td>
<td>0.84</td>
<td>0.95</td>
</tr>
<tr>
<td>Language</td>
<td>0.93</td>
<td>0.89-0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>Gross motor</td>
<td>0.91</td>
<td>0.86-0.95</td>
<td>0.95</td>
</tr>
</tbody>
</table>

ICC: Interclass Correlation Coefficients/ DDST-II: Denver Developmental Screening Test II

The reliability of DDST-II was evaluated by the Kuder-Richardson coefficients determination. The estimated coefficients were 0.74 for personal-social, 0.63 for fine motor-adaptive, 0.63 for language and 0.61 for gross motor domains. Test-retest and inter-rater methods were also used as other ways for reliability determination.

Table 2: Interclass Correlation Coefficients for inter-rater examination by DDST-II

<table>
<thead>
<tr>
<th>Developmental domain</th>
<th>ICC*</th>
<th>ICC interval (95% confidence interval)</th>
<th>Cronbach’s α coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal-social</td>
<td>0.88</td>
<td>0.79-0.93</td>
<td>0.94</td>
</tr>
<tr>
<td>Fine motor-adaptive</td>
<td>0.89</td>
<td>0.82-0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>Language</td>
<td>0.96</td>
<td>0.93-0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>Gross motor</td>
<td>0.86</td>
<td>0.77-0.92</td>
<td>0.93</td>
</tr>
</tbody>
</table>

ICC: Interclass Correlation Coefficients/ DDST-II: Denver Developmental Screening Test II

Discussion

In our study the content validity of DDST-II was verified by reviewing books and journals, and by specialists’ opinions. All of the questions in DDST-II had appropriate content validity, and there was no need to change them.
Table 3: Comparison between the results of DDST-II with ASQ and pediatricians' evaluation

<table>
<thead>
<tr>
<th>Test or clinical evaluation</th>
<th>Result</th>
<th>DDST-II results (No. of children)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Failed</td>
<td>Passed</td>
</tr>
<tr>
<td>ASQ (No. of children)</td>
<td>Failed</td>
<td>21</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Passed</td>
<td>49</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>70</td>
<td>123</td>
</tr>
<tr>
<td>Pediatricians' evaluation</td>
<td>Failed</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>(No. of children)</td>
<td>Passed</td>
<td>10</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>14</td>
<td>42</td>
</tr>
</tbody>
</table>

DDST-II: Denver Developmental Screening Test II / ASQ: Ages and Stages Questionnaires

We evaluated the reliability of the test by the Kauder-Richardson coefficients determination. Determined Kauder-Richardson coefficients for all of developmental domains were “good”. Test-retest and inter-rater methods were also used as other ways for reliability determination. In test-retest examination the Cronbach’s α coefficients for all developmental domain is very good and kappa measure of agreement is 87% (P<0.001). In Inter-rater examination the Cronbach’s α coefficients for all developmental domain were very good and kappa measure of agreement was 76% (P<0.001). Thus DDST-II has very good reliability in test-retest and Inter-rater examination.

Sensitivity of DDST-II in different references ranges from 40-83%[14,15] and its specificity is reported from 40-80%[16,17]. In this study sensitivity and specificity of DDST-II by comparing the results of DDST-II and ASQ were 60% and 69% respectively. Of course this cannot be considered as the actual validity of the test, because, as explained before, ASQ is not a diagnostic gold standard test. We found that children passed the ASQ (88%) more than DDST-II (65%) and consistency coefficient of the two tests was poor (0.21). Therefore, either ASQ may be undersensitive and/or DDST-II oversensitive. Which of these is true? It has to be investigated by comparing the results of these tests with the results of a developmental diagnostic test.

It is possible that in comparison with the Denver sample, Iranian children have a slower rate of development. One study conducted in Shiraz (Iran) showed that 3-6 year-old Iranian children have slower rate of development by DDST-II in fine and gross motor domains[16]. It is worthy to mention that DDST-II has subgroup standards based on sex, race, maternal education and place of residence that are presented in Denver-II technical manual[18].

Some of our other findings revealed no relationships between sex, maternal education and place of residence with children's developmental status in Tehran. In another native study in Shiraz city, gross and fine motor performance of 1524 children aged 3-6 years was evaluated by DDST-II in 2005-2008. In this evaluation girls had better performance[16]. Yalaz and Epir tested 1176 Turkish children aged 2 weeks to 6 yr-4 mo by DDST in 1984. They found that girls’ development is better than that of boys[17].

Durmazlar and Anlar also evaluated 1091 Turkish children aged 0-72 months by DDST-II in 1998. In their study few and inconsistent differences were observed between boys and girls[19]. Bryants and Stark. evaluated the achievement of test items in the first year of life of Cardiff infants by DDST in 1974. They concluded that in the first year of life, there are no developmental differences between boys and girls[20].

In this study maternal education had no effect on children’s developmental status. In South Okanagan of Canada, Barnes and Stark evaluated 206 children 2 weeks to 6 yr-4 mo old by DDST. In their study, there was no relationship between maternal education and child’s developmental level except in 10-12 month-old infants[21]. William's study in Philippine showed that maternal education and her birth place (rural or urban) had relation to DDST results. Children of mothers with higher educational level and who were born in urban regions had better results[22]. Lejaraga and co-workers studied 0-5 year-old children in Argentina in 2002. In their research, after first year of life, sex
and maternal education were related to children’s developmental status[23].

Bryant et al examined 686 infants aged 2 weeks to 12 months in Cardiff by DDST in 1974. They found that children of Cardiff in fine and gross motor domains had slower developmental rate[20]. Later they standardized DDST on 1574 under 6 year-old children in Cardiff in 1976. They showed that Cardiff children had slower developmental rate, but under 18 months of age, they were better than Denver children in language domain. There was no difference between the two groups in fine motor development and Cardiff children had better performance in personal-social domain [24]. Oeda in one DDST study on 1171 children showed that children of Tokyo had better results in some items of personal-social domain but in infancy they had slower developmental rate in gross and fine motor domains[25]. Gross and fine motor performance of 78 healthy Swedish children aged 15 to 18 months were examined with DDST by Lundberg. He showed that Swedish children had slower rate of fine and gross development[26]. Shapiro and Harles in Israel examined 2248 children 2 weeks to 6.5 years old with DDST. They found that Israeli children in comparison to Denver children had slower rate of development in gross and fine motor domains[27]. A review article showed that by using DDST, children of Japan, Philippine, Tokyo, Okinawa, Netherland and Bangkok had slower rate of motor development[28].

This study has some limitations. First, some parents that their children had to be re-examined by DDST-II, did not return to clinic. So we chose other cases and there was wasting of time and resources. Second, developmental screening tools are not diagnostic and their results must be followed by a more intensive evaluation. The sensitivity and specificity of DDST-II must be determined by comparison of the test results with a developmental diagnostic test. Because there was no standardized diagnostic test in Iran, we compared the DDST-II with ASQ and pediatricians’ evaluation of children’s development. Third, by considering the results of similar researches, and on the base of results of this study, delays in fine and gross motor areas are more than in other developmental domains. It is recommended to design another study on larger samples in order to standardized DDST-II and determine the norms of Iranian children.

Early detection and intervention of children with developmental delays or disorders is an important issue in pediatrics medicine. Early detection and intervention in developmental problems can reduce their impacts on the well-being and functioning of child and his/her family. American Academy of Pediatrics recommended that pediatricians use standardized developmental screening test regularly at the 9, 18 and 30 (or 24) month visits[29].

Conclusion

This research showed that Persian version of DDST-II has a good validity and reliability, and can be used as a screening tool for developmental screening of children in Tehran city. For determining the sensitivity and specificity of the test, it is suggested that the results of each of the two screening tests DDST-II and ASQ are compared with a standard diagnostic test in future studies.

Acknowledgment

This research was supported by a grant from Pediatric Neurorehabilitation Research Center of University of Social Welfare and Rehabilitation, Tehran. We wish to thank Dr N Hatamizadeh, Dr H Karimi and Dr F Soleimani for their assistance.

Conflict of Interest: None

References


دستورالعمل انجام آزمون

- سعی کنید با لبخندزدن، صبح کردن پاره‌کردنی‌های کودک را بخشنده‌بندی نمایید.
- کودک با پای چند ثانیه به دستش خود شود.
- والدین مو تنبلندی به حرکت‌های کودک را برای متغیرهای گفتنی‌های خود هر روز مورد کمک کنند.
- لازم نیست کودک نیاز به کمک یا بیدریگی پشت لباسی را بپندد.
- کاماو این به همگامی و حدود ۲۰ ساعت متر بالا را از صورت کودک و به صورت یک پیم داره از یک طرف به طرف دیگر حرکت دهد.

در صورتی که کودک بتواند پس از مدتی جمع‌بندی به تأثیر با پشت انگشتان آرای بگیرد مثبت ثبت شود.
- اگر کودک سعی کند پیم گلره کاموا کجا رفته است، مثبت ثبت است. آزمونگر یا پیدا حرکت دادن باروی خود کاموا را سریعاً ببیند.
- بیان‌زاد آن‌که از دید کودک خارج شود.
- کودک با پای بدون کمک گرفتن از بعد دهان یا میز مکعب را از سمت دست به دست دهد.

در صورتی که کودک بتواند با هر سرعتی از ست قسمتی از ست قسمتی از ست پشت انگشت دیگرش کمشش را بدرد مثبت ثبت مثبت شود.
- جهت می تواند فقط ۴۰ درجه با کمر با خط آزمون‌گر فرق نماید.
- دست خود را مشت کنید و شست را به طرف بالا بگیرید و فقط شست خود را دکر دهد. در صورتی که کودک بتواند تفکید کرده و غیر از ست قسمتی از ست قسمتی شود.

- هرگونه شکل پشت قابل قبول است. حرکات مدور و مند قابل قبول نیست.

- کدام خط بلندتر است (نه بزرگتر) کافی را وارونه کرده و سوال را تکرار کنید (یا بیان درست ۳ بار از ۵ بار ۵ بار از ۴ بار ۴ بار از ۴ بار)

- هر دو خطی که حداکثر دو و سه یک‌دیگر را قطع کنند قابل قبول است.

- او اجازه دهید کودک از روز شکل کی که کودک اگر نتوانست شما انگام دهد تا بپندد.

- وقتی عارضه ۱۲ و ۳۱۴ و ۱۵ و ۱۶ را انگام می‌دهید، شکل یا رایکر. کشیدن ۱۲ و ۱۴ و ناشان دهید.

- هنگام شروران اعضا هر جفت عضو (۱ دست، ۲ پا و غیره) را یکی حساب کنید.

- یک مکعب را در فنگان گذاشته و فنگان را به آرامی نزدیک گوش کودک و از خارج از دید او نکان دهید. برای گوش دیگر نکرار.

- به تصویر اشاره کنید و از کودک بخواهید نام آن را برد (فقط صدای آن درآورده قابل قبول نیست) اگر کمتر از ۲ تصویر را نام یاد از این بخواهید تصویری را که نوع می‌بیند نشان دهد.
19- عروسک را بپردازید که کودک گوش‌پیچی-چشم‌پیچی-گوش‌پیچی-چشم‌پیچی-کهکشان درون دهد. پايد
6 مورد از 8 مورد را نشان دهد.

20- یا استفاده از تصاویر از کودک پرنسس که پروری که مشکل؟ کدام می‌باشد کدام؟-
کدها زرد؟ کدام پارس می‌باشد کدام؟-
چهار نول می‌رود 2تا5تا 2تا 5تا 2تا 3تا 2تا 3تا 2تا 3تا 2تا 3تا 2تا 3تا 2تا 3تا 2تا 3تا 2تا 3تا 2تا 3تا 2تا
21- از کودک پرنسس یا دریه سردرد است چکار می‌کنی؟ وظیفه خسته ای چکار می‌کنی-
وظیفه گردن می‌شهری چکار می‌کنی۲تا ۳تا ۳تا ۳تا ۳تا ۳تا ۳تا ۳تا ۳تا ۳تا ۳تا ۳تا ۳تا ۳تا ۳تا ۳تا ۳تا ۳تا ۳تا
3 مورد است.

22- از کودک پرنسس یا چنگال کیک می‌کنی؟ از صندلی چه استفاده می‌کنی؟ از مسله‌ی چه استفاده ای-
که می‌کننت؟ کلماتی که دلتنک منابع کارداری باید در پایین ها وجود داشته باشد.

23- اگر کودک بی‌دتای مکعب ها را درست روی کاغذ گذاشته و گوید که چند چند می‌گردد؟
به کودک گوئید مکعب را روز می‌میزد، زیر میز، جلوی من و پشت من گذاش (۴تا ۴مورد است) (با اشاره و یا حرکت سر یا چشم
به یا کمک تکیه).

25- از کودک پرنسس (تپ چسبی- دریا چسبی- میت چسبی- خانه چسبی- موز چسبی- پره چسبی-
نرد چسبی- سفید چسبی؟
اگر پرندگان مورد صحیف مشکل، جنس یا دسته بندی کلی آن را گوئید می‌شود. (مثلاً) گوئید موز میوه است نه اینکه فقط گوئید
وزر است) ۵تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا ۸تا
4 مورد است.

27- کودک می‌کند است فقط آن یا ننده که گیاه به اندازه می‌گردد. می‌کند است چهار دست و یا نرود.

29- کودک باید بتواند تپوا تپ را حداقل ۰۵سانتی‌متر بلوک آزمونگر برآورد کند بلوک‌کوپه در دسترس یا قرار گیرد (۳فوت).

30- بی کودک گوئید به طرف جلو راه بروید بلوک‌کوپه پاشته پک‌پای خود را حدود ۷۵سانتی‌متر از شست پای دیگر قرار دهد.
آزمونگر می‌تواند نحوه انجام کار را به او نشان دهد.
کودک باید ۴کام پشت سر یا برد. در سال دوم زندگی نمی‌از کودکان طبیعی همکاری کافی ندارند.

31- مشاهدات: