Convergent Validity Between Two Motor Tests: Movement-ABC and PDMS-2

Hilde Van Waelvelde and Wim Peersman
Arteveldehogeschool and Ghent University

Mattieu Lenoir
Ghent University

Bouwien C.M. Smits Engelsman
K.U. Leuven

The aim of this study was to investigate the convergent validity of the Movement Assessment Battery for Children (M-ABC) and the Peabody Developmental Motor Scales – 2 (PDMS-2). Thirty-one 4- and 5-year-old children (mean age 4 years 11 months, SD 6 months), all recruited from a clinical setting, took part in the study. Children were tested on the M-ABC and the PDMS-2 in a counterbalanced order on the same occasion. The results showed that the total scores on the two tests correlated well ($r_s = .76$). However, when the ability of the two tests to identify children with difficulties was examined, agreement between them was low ($K = .29$), with the PDMS-2 being less sensitive to mild motor impairment in this population. Taken together, these findings suggest that clinicians need to be aware that, although measuring a similar construct, these tests are not interchangeable.

All of us are clumsy at times, but there exists a small group of children whose coordination is so poor that they find it hard to learn the movement skills that a typically developing child would learn with ease. Although this disorder appears mild when compared to neuro-motor disorders such as Cerebral Palsy or Muscular Dystrophy, the impact of these difficulties on the lives of the children concerned is known to be severe and often long-lasting (Polatajko, 1999; Henderson & Henderson, 2002). Now formally labelled by the American Psychological Association...
Van Waelvelde et al. (DSM-IV, 1994) as Developmental Coordination Disorder (DCD), the etiology of this condition remains unclear. However, the need to identify and follow children with mild to moderate movement difficulties has been discussed since the seventies (Wedell & Raybounc, 1976). Early recognition of this group of children is, however, very important, not only to provide support for the child and his/her parents as early as possible, but also to prevent the development of concomitant problems, e.g., psychosocial problems (Jongmans, 2005).

A number of authors have noted the need for a multilevel approach to assessment for children with mild to moderate motor impairment (Henderson & Sugden, 1992; Kaplan, Dewey, Crawford, & Wilson, 2001; Larkin & Rose, 2005; Schoemaker, Smits-Engelsman, & Jongmans, 2003). In other words, assessment should encompass multiple sources of information beginning with a detailed assessment of the motor impairment per se and extending to the impact on activities of daily living and school performance as well as social and family life. As part of a wide-ranging assessment of this kind, a formal standardized motor test should always be included. As Tieman, Palisano, and Sutlive (2005) point out, valid and reliable motor tests are essential tools for clinicians, therapists, and teachers, not only as part of the identification process but also as a means of evaluating progress in motor development and assessing the efficacy of interventions.

Movement skill assessment is a relatively new topic of research and only a few tests fulfilling the minimal criteria of standardization, reliability, and validity are available, especially for pre-school children (Barnett & Peters, 2004; Burton & Miller, 1998; Burton & Rodgerson, 2003; Tieman et al., 2005). To date, therefore, we are still unsure as to which tests to use and how valid the tests are to identify and follow-up motor problems at preschool age. The aim of the present study was to evaluate two tests, both of which are currently used with this particular population of children.

The first test, the Movement Assessment Battery for Children (M-ABC; Henderson & Sugden, 1992), is used extensively in many European countries. In a review paper examining diagnostic criteria for Developmental Coordination Disorder, the worldwide use of the M-ABC was confirmed in both clinical and research contexts (Geuze, Jongmans, Schoemaker, & Smits-Engelsman, 2001). The M-ABC has been translated into several different languages and validated for different cultures (e.g., Chow, Hsu, Henderson, Barnett, & Kai Lo, 2006). The test is not time consuming and children are likely to participate willingly. The M-ABC contains two assessment instruments: a standardized performance test and a criterion-referenced teachers’ observational checklist. In this study, we will focus on the performance test only. The M-ABC test contains eight items, divided into three major performance areas: manipulative skills, ball skills, and balance skills. The scoring system is a specific feature of the M-ABC test. The raw performance score of each item is converted into a scaled score ranging from 0 to 5. Lower scores indicate a better performance. All children performing better than the 25th percentile of the standardization group for an item receive a zero score for that specific item. The M-ABC was specifically developed for identifying and evaluating children with mild to moderate motor impairment (Henderson & Sugden, 1992). The M-ABC was not designed for identification of well-coordinated children, that is children who have above average profiles, as it is not sufficiently discriminating at that end of the distribution (Larkin & Rose, 2005).
Another measure of motor development commonly used for preschool-aged children is the Peabody Developmental Motor Scales-2 (PDMS-2; Folio & Fewell, 2000). This test includes two comprehensive norm-referenced scales, one focusing on gross motor performance, the other on fine motor. In both cases, the scales are intended to determine whether children of age 6 years and younger have delayed motor development. The PDMS-2 is a more extensive test than the M-ABC, covering a wide variety of skills organized in a developmental sequence.

Examination of the manuals of these two tests reveals that they were developed from a different theoretical framework and with a slightly different population of children in mind. Folio and Fewell (1983), for example, began by focusing on children with severe movement difficulties. They described a variety of purposes for the PDMS, including estimating a child’s motor competence relative to his or her peers, determining the necessity or eligibility for intervention, planning an intervention program, and evaluating changes over time (Burton & Miller, 1998). With the development of the subsequent PDMS-2, Folio and Fewell (2000) incorporated a qualitative approach along with a quantitative approach, but the developmental framework remained. On the other hand, the M-ABC is mainly concerned with the identification and description of mild impairment of motor function, although the authors also address intervention planning, program evaluation, and research purposes. This test, too, includes a qualitative element, which extends beyond the motor domain to encompass other difficulties, such as lack of attention, which might affect motor performance. The different theoretical frameworks of the two tests are expressed in different test structures. The developmental sequence of the PDMS-2 test items cannot be compared to the eight M-ABC test items constructed to identify the children failing to perform the tasks; however, in clinical practice, both are used to identify and follow-up preschool children with mild motor disorders.

One important topic in the debate regarding the most appropriate means of assessment involves evaluating the validity of the instruments concerned. Contemporary measurement specialists define validity as “the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of the test” (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1999). Yun and Ulrich (2002), for example, stress that establishing validity is an ongoing process, being determined by criteria related to the inference or decision made from the measurement. Thus, validity is not a property of a test, as such, but rather of the meaning of the test scores (Yun & Ulrich, 2002). The traditional concept of validity divides it into three separate types—content, criterion, and construct validity. Messick (1995) proposes a new concept of validity that unites these issues. Validity can be defined as an evaluative summary of both the evidence for and the actual—as well as potential—consequences of score interpretation and use. This unified validity concept does not imply that validity cannot be differentiated into distinct aspects, such as external validity, including evidence of convergent and discriminant validity gathered from multimethod comparisons. The constructs represented in the assessment should rationally account for the external pattern of correlations. Both convergent and discriminant correlation patterns are important, the convergent patterns indicating a correspondence between measurers of the same construct and the discriminant pattern indicating a distinctness from measures of other constructs (Messick, 1995).
Validity data for the M-ABC can be found in the manual and in more recently published studies (e.g., Smits-Engelsman, Henderson, & Michels, 1998; Croce, Horvat, & McCarthy, 2001; Van Waelvelde, De Weerdt, De Cock, & Smits-Engelsman, 2004). However, relatively few of these studies address the first age band of the M-ABC (for or use with 4-6 year old children), and none included 4-year-old children. Van Hartingsveldt, Cup, and Oostendorp (2005) have recently investigated the convergent validity of the standard fine motor score of the PDMS-2 and the manual dexterity cluster score of the M-ABC in 18 children between the ages of 4 and 5 years with and without presumed fine motor problems. They found a correlation of .69 between tests.

The aim of this study was to extend the study of van Hartingsveldt, Cup, and Oostendorp (2005) to the gross motor scale and to the total scores yielded by each test. First, the convergent validity between the total scores as well as between both fine and gross motor parts of the M-ABC and the PDMS-2 was investigated. Second, the discriminant validity between the fine and gross motor parts of the tests was evaluated. Since validity is related to the social consequences (Messick, 1995), this study also examined whether the two assessment tools identify the same children as being motor impaired.

Method

Participants

For this study, 31 children (4 girls and 27 boys) between the ages of 4 and 5 years were selected from a population of children attending a Center for Developmental Disabilities and a Center for Ambulant Rehabilitation in Flanders (Belgium). Children who met the following criteria were included in the study: (a) no severe neuro-motor disorder such as Cerebral Palsy, according to the medical records and (b) a level of cognitive development that allowed them to understand the test instructions. This judgment was made by the therapists or educators from the Center most closely involved with the child. For 17 children, IQ data were also available in the records and the IQ varied between 63 and 130. Since neither instruments provides separate norms for males and females, both boys and girls were considered for inclusion. Only four girls met the inclusion criteria. Permission for the study was given by the local ethics committee. Informed consent was obtained from the parents and assent was obtained from the children.

Materials

Movement Assessment Battery for Children. The M-ABC consists of two independent instruments: a test and a checklist. This study focused exclusively on the standardized test, using the Dutch version produced by Smits-Engelsman, 1998. The test consists of eight items that evaluate fundamental movement skills. The test accommodates four age bands designed for use with children age 4 to 12 years. The items are different for each age band but cover the same type of skills. All children in the present study were assessed on the first age band, for use with 4-6 year olds, which includes three manual dexterity items: (a) posting coins in
a bank box—time needed in sec; (b) threading beads—time needed in sec; (c) drawing a line into a trail—number of errors; two ball skills: (d) catching a bean bag—number of correct catches; (e) rolling a ball into a goal—number of goals; three balance items: (f) standing on one leg—maximal time in sec; (g) jumping over a cord—succeeding at the first, second, or third trial, or not succeeding; and (h) walking heels raised on a line—number of steps. Test administration takes 20 to 30 min. Each item score is converted into a scaled score ranging from 0 to 5. The total impairment score is the sum of 8 scores, which when summed produces a score between 0 and 40. Subscores can be calculated by adding up the 3 scores for manual dexterity (range: 0-15), 2 scores for ball skills (range: 0-10), and three balance items (range: 0-15). Lower scores indicate a better performance. The total motor impairment score is then interpreted in the light of percentile norm tables. Norms are based on a USA standardization sample. For the Dutch version, a limited sample of children from the Netherlands was assessed. The results prompted Smits-Engelsman (1998) to conclude that the USA norms could be used for children from the Netherlands.

In the manual the 5th and the 15th percentile are suggested as critical cut-off scores, with a score at or below the 5th percentile being indicative of a definite motor problem. Scores between the 5th and 15th percentile suggest a degree of difficulty that is borderline. In addition to reliability data presented in the manual, test-retest reliability was recently investigated in a large sample of Chinese preschool children. An ICC of .77 suggested acceptable test-retest reliability (Chow & Henderson, 2003). Recently, our research group also studied the test-retest reliability for this age group, but within a group of children with low motor performance and obtained an ICC of .88 (Van Waelvelde, Peersman, Lenoir, & Smits-Engelsman, in press).

**Peabody Developmental Motor Scale-2 (PDMS-2).** The PMDS-2 by Folio and Fewell (2000) evaluates the early motor milestones and fundamental movement skills of children from birth to 5 years and 11 months of age. The items are organized further into six skill categories. The Gross Motor Scale contains the categories reflexes (only for children below one year), static balance (30 items), locomotion (89 items), and ball skills (24 items). The Fine Motor Scale contains the categories grasping (26 items) and visual-motor integration (72 items). For each item in the test, performance criteria are specified and scored on a 3-point scale, from zero to two. Administration starts in each category at a basal age level, defined as the first level at which three consecutive items get a 2-score. Then the testing steps up to the ceiling age level, defined as the level at which 3 consecutive items get a 0-score. The raw score is the sum of the item scores and can be transformed into standardized scores, percentiles, and standardized developmental motor quotients for the Gross Motor Scale, the Fine Motor Scale, and the Total Score. Test administration takes 45 to 60 min. Norms, based on a USA standardization sample, are available. To the best of our knowledge, the PDMS-2 norms have not been checked for their suitability for other populations.

Folio and Fewell (2000) report in the manual a study evaluating PDMS-2 test-retest reliability for children younger than 17 months. Reliability coefficients varied from .73 to .96. Test-retest and interrater reliability of the fine motor scale of the PDMS-2, specifically for 4- and 5-year-old children, varied from .84 to .99.
(Spearman correlation coefficients) in a study by Van Hartingsveldt, Cup, and Oostendorp (2005).

**Procedure**

Each child was tested individually in a quiet room on the M-ABC and the PDMS-2 on the same day. The order of administering the two tests was counterbalanced. At least one break was offered to the children between tests. If necessary, a second break was offered between the categories of the PDMS-II. The tests were conducted and scored in accordance with instructions provided in the respective test manuals. All children were tested by the same two trained examiners, working together, to evaluate the children. Alternating, one examiner was instructing and evaluating the child while the other examiner wrote down the results.

**Data Analysis**

To evaluate the convergent and discriminant validity, the PDMS-2 quotient scores and the sum of the scaled M-ABC scores, the total impairment score, were used. As not all data were normally distributed, Spearman rank correlation coefficients were calculated. The 95% confidence intervals for the Spearman rank correlation coefficients were estimated using a bias-corrected bootstrap procedure (analyses were carried out in Stata 9.0). To be able to compare the gross and fine motor quotients of the PDMS-2 with the M-ABC, the cluster scores ball skills and balance of the M-ABC were summed to produce a M-ABC gross motor score (M-ABC GM), and the dexterity cluster score of the M-ABC was considered as a M-ABC fine motor score (M-ABC FM). In addition, the children were divided in two groups on each test separately: (a) children with a score at or below the 15th percentile and (b) children with a score above the 15th percentile. The agreement between the two tests in this dichotomy was evaluated with a Kappa statistic.

**Results**

The children varied in age between 4 years 0 months and 5 years 11 months, with a mean of 4 years 11 months (SD = 6 months). The median total M-ABC score, expressed as a percentile, was 26, with a range of 1 to 78. The median PDMS-2 score was 39 with a range of 7 to 77 (Figure 1). The absolute value of the Spearman correlation coefficient between the total score of the M-ABC and the total score of the PDMS-2 was .76.

Correlations between the various subscores of the M-ABC and of the PDMS-2 are reported in Table 1. As the table shows, the correlation coefficient between the gross motor subscores for each test was higher than between their fine motor equivalents, 0.71 versus .48. However, the correlation between the M-ABC manual dexterity and the PDMS-2 gross motor quotient was slightly higher than the correlation with the PDMS-2 fine motor subscores (0.51 versus .48). The correlation between M-ABC subscore manual dexterity and the sum of the gross motor subscores of the M-ABC was higher than the correlation between the fine and gross motor quotient of PDMS-2.
Figure 1 — Box plots of the percentile scores of the M-ABC and PDMS-2 (n = 31).

Table 1 Absolute Values of Spearman Correlation Coefficients Between the M-ABC and the PDMS-2 (95% Confidence Interval Between Parentheses)

<table>
<thead>
<tr>
<th></th>
<th>M-ABC GM</th>
<th>M-ABC FM</th>
<th>PDMS-2 Total</th>
<th>PDMS-2 GMQ</th>
<th>PDMS-2 FMQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-ABC Total</td>
<td>.91**</td>
<td>.78**</td>
<td>.76**</td>
<td>.71**</td>
<td>.47**</td>
</tr>
<tr>
<td></td>
<td>(.78-.97)</td>
<td>(.57-.90)</td>
<td>(.51-.90)</td>
<td>(.42-.88)</td>
<td>(.14-.70)</td>
</tr>
<tr>
<td>M-ABC GM</td>
<td>.49**</td>
<td>.67**</td>
<td>.71**</td>
<td>.36**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.16-.76)</td>
<td>(.44-.85)</td>
<td>(.48-.87)</td>
<td>(.003-.66)</td>
<td></td>
</tr>
<tr>
<td>M-ABC FM</td>
<td></td>
<td>.64**</td>
<td>.50**</td>
<td>.48**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.35-.83)</td>
<td>(.09-.79)</td>
<td>(.12-.72)</td>
<td></td>
</tr>
<tr>
<td>PDMS-2 Total</td>
<td></td>
<td></td>
<td>.86**</td>
<td>.69**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.65-.95)</td>
<td>(.34-.86)</td>
<td></td>
</tr>
<tr>
<td>PDMS-2 GMQ</td>
<td></td>
<td></td>
<td></td>
<td>.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.09-.60)</td>
<td></td>
</tr>
</tbody>
</table>

Note. M-ABC GM = sum of Gross Motor subscores of the M-ABC (subscore ball skills + balance); M-ABC FM = Fine Motor subscore of the M-ABC (subscore dexterity); PDMS-2 GMQ = Gross Motor Quotient of the PDMS-2; PDMS-2 FMQ = Fine Motor Quotient of the PDMS-2.

** = coefficient is significant at .01 level.
* = coefficient is significant at .05 level.
Van Waelvelde et al.

Table 2 reports a crosstab of both tests when the children were divided in two groups. The agreement between tests in this dichotomy resulted in a $\kappa$ of .29, with a 95% CI between 0.15 and 0.43. The percent agreement between tests was 68%.

Table 2 Frequencies of Children Scoring at or Below the 15th Percentile (Pc) and Above the 15th Percentile on the M-ABC and on the PDMS-2

<table>
<thead>
<tr>
<th></th>
<th>$\leq$ Pc15</th>
<th>$&gt; \text{Pc15}$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Movement $\leq \text{Pc15}$</td>
<td>3</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>Movement $&gt;$ Pc15</td>
<td>0</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>ABC Total</td>
<td>3</td>
<td>28</td>
<td>31</td>
</tr>
</tbody>
</table>

Table 2 reports a crosstab of both tests when the children were divided in two groups. The agreement between tests in this dichotomy resulted in a $\kappa$ of .29, with a 95% CI between 0.15 and 0.43. The percent agreement between tests was 68%.

Discussion

Early intervention requires early identification and follow-up of developmental motor disorders. In this context, questions regarding the most appropriate means of assessment should be considered. The main purpose of this study was to evaluate the convergent validity of two tests frequently used for early identification and follow-up of motor disorders in preschool children.

The results of the study suggested that the M-ABC and PDMS-2 do indeed measure a similar construct, the correlation between the total scores on the two tests being 0.76. However, while a correlation of this magnitude supports the convergent validity of both tests, it should also be noted that a substantial part of the variance on one test cannot be explained by the scores on the other test. In this context, clinicians should be aware that test results do not reflect “the” motor ability or “the” motor performance of a child and can only be interpreted in relation to the specific tasks used in the assessment.

When the various components of the two tests were examined, the correlation of 0.71 between their respective gross motor components also provided evidence of convergent validity for each instrument. In contrast, the fine motor scales were less highly correlated (.48). There are two possible explanations of this outcome. First, the MD subsection of the M-ABC consist of only three manual dexterity items and consequently does not cover the fine motor construct completely. Second, the items in this subsection are more time constrained than those contained in the PDMS-2. Against these arguments must be set the finding of Hartingsveldt, Cup, and Oostendorp (2005), who found a higher correlation of .69 between fine motor scores on the PDMS-2 and manual dexterity scores on the M-ABC in 18 children between the ages of 4 and 5 years with presumed fine motor problems. Differences in the investigated sample and/or sampling error might explain these differences. However, it should also be noted that with the small samples involved in both of these studies and therefore the large confidence intervals of the correlation coefficients, the interpretation of the differences between correlation coefficients should be made with caution.
In contrast to positive evidence of convergent validity, the correlation coefficients between the gross and fine motor subscores of the two tests do not provide evidence for the discriminant validity between the subscores. Again, given the large confidence intervals, interpretation of these differences between correlation coefficients is difficult. The fine motor quotient of the PDMS-2 seems to measure a different concept when compared to the gross motor scores of PDMS-2 and M-ABC. But the correlation coefficients between the M-ABC manual dexterity items and the gross motor scores of respectively M-ABC and PDMS-2 were higher than might be expected. This could be related to the small number of items of the M-ABC dexterity score. In general, in movement skill assessment, there is still some confusion about allocating tasks in terms of their function or attribute (Burton & Rodgerson, 2001; Larkin & Rose, 2005).

Turning now to the question of early identification of children who have or are suspected of having a motor disorder, the main question addressed was whether the two tests identified the same children. Although the overall correlation between them was substantial, there was actually rather low agreement in the extent to which they identified individual children as impaired. The M-ABC scored more stringently than the PDMS-2, and its sensitivity was higher. In contrast, although the investigated sample was a group of children at risk for motor impairment, the PDMS-2 only identified three children at risk and no children with a definite disorder. This outcome was poor in comparison to other studies investigating the agreement between different tools to identify children with mild motor impairment (Crawford, Wilson, & Dewey, 2001; Smits-Engelsman, Henderson, & Michels, 1998). One possible explanation of the low agreement between PDMS-2 and M-ABC in this study might be found in the way the tests were standardized. In the standardization sample for the PDMS-2, 10% of children with disabilities were included, and 2% of them were physically disabled (Folio & Fewell, 2000). In the standardization sample of the M-ABC, all children were physically unimpaired in the sense that none should have obvious physical impairments (Henderson & Sugden, 1992). None of the manuals explicitly describe how the children were recruited. The suitability of the M-ABC norms for Flemish children has been investigated and only small differences have been detected between the motor performance of American, Dutch, and Flemish 4- to 5-year-old children (Van Waelvelde, Peersman, Lenoir, & Smits-Engelsman, 2007). The suitability of the PDMS-2 norms for identification of Flemish 4- and 5-year-old children with mild motor impairment can strongly be doubted.

Administration of the PMDS-2 was reported by the examiners to be more difficult than administration of the M-ABC. Extended training was necessary before they could begin administration. We found it useful to administer the PDMS-2 together. The PDMS-2 offers a more complete assessment and allows making more qualitative observations during the test. The M-ABC also offers a list of qualitative observations for each item. However, the list does not allow a score but helps the clinician in the evaluation of movement quality along with some behaviors. For screening and identification the M-ABC is more convenient as the test takes less time and is easier to administer.

An important aspect of the construct validity of a diagnostic test is its predictive validity. Longitudinal research is strongly needed to validate the M-ABC and PDMS-2 for identifying children with mild motor impairment.
Conclusions and Recommendations

Although both the M-ABC and PDMS 2 appear to measure similar constructs, practitioners need to be aware that they are not interchangeable. Moreover, diagnosis of a child should always be part of an overall assessment plan. Tests can help clinicians in their decisions, but they must be aware that test results can only be generalized with extreme caution. The choice of one test over another should be made carefully, and limits of reliability and validity of each test should be known before decisions can be made. The PDMS-2 is less sensitive for identification of Flemish children between the ages of 4 and 5 years with mild motor impairment. Without suitable PDMS-2 norms, the M-ABC may be considered a better choice for screening of preschool children suspected of mild motor impairment. The M-ABC is less time consuming and easier to administer. On the other hand, the PDMS-2 evaluates different and more motor tasks, enhancing the validity for follow-up and development of therapeutic programs. The PDMS-2 seems to be suited for children with a physical disability.

References


**Acknowledgments**

The authors would like to thank the children, their parents, the participating schools, Carolien Bombaert, and Elisabeth Rys for collecting data and Prof. Dr. Sheila Henderson for her valuable assistance in the preparation of the manuscript.