Factorial validity of the Movement Assessment Battery for Children-2 (age band 2)

Matthias Oliver Wagner a,*, Julia Kastner b, Franz Petermann b, Klaus Bös c

a Institute for Health Science, Department of Sports and Movement, University of Education Schwäbisch Gmünd, Oberbettinger Str. 200, 73525 Schwäbisch Gmünd, Germany
b Centre for Clinical Psychology and Rehabilitation, University of Bremen, 28359 Bremen, Germany
c Institute for Sports and Sports Science, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany

ABSTRACT

The Movement Assessment Battery for Children 2 (M ABC 2) is one of the most commonly used tests for the diagnosis of specific developmental disorders of motor function (F82). The M ABC 2 comprises eight subtests per age band (AB) that are assigned to three dimensions: manual dexterity, aiming and catching, and balance. However, while previous exploratory findings suggested the correctness of the assumption of factorial validity, there is no empirical evidence that the M ABC 2 subtests allow for a valid reproduction of the postulated factorial structure. The purpose of this study was to empirically confirm the factorial validity of the M ABC 2. The German normative sample of AB2 (7-10 years; N = 323) was used as the study sample for the empirical analyses. Confirmatory factor analysis was used to verify the factorial validity of the M ABC 2 (AB2). The incremental fit indices (χ² 28.675; df 17; Bollen-Stine p value 0.318; RMSEA 0.046 [0.011 0.075]; SRMR 0.038; CFI 0.960) provided evidence for the factorial validity of the M ABC 2 (AB2). However, because of a lack of empirical verification for convergent and discriminant validity, there is still no evidence that F82 can be diagnosed using M ABC 2 (AB2).

1. Introduction

1.1. Background

In recent years, developmental coordination disorders (DCD) have received increasing attention by the international research community (cf. Visser, 2003). While comorbidities are well documented and the prevalence of DCD is rising (e.g. Cermak, Gubbay, & Larkin, 2002; Kastner & Petermann, 2010), DCDs are often considered in therapeutic practice as being a temporary developmental phenomenon (see also Schott, 2010, p. 169). However, 46% of adolescents with a DCD diagnosis at the age of 5 years still suffered from related disorders at the age of 15 years (Cantell, Smith, & Ahonen, 1994) suggesting that children with deficits at a young age may not necessarily outgrow these deficits. One possible reason for the persistence of DCDs in some children is the observation that children with DCDs usually also consciously refrain from establishing a sufficient level of physical activity (Henderson, 1993). Hence, there is a great need to promote general healthy development in these children.

Successful treatment targeting necessitates a reliable diagnosis of DCDs (see Kastner & Petermann, 2009; Wilson, 2005). Diagnostic criteria for DCDs have been specified in the Diagnostic and Statistical Manual of Mental Disorders (DSM IV;
American Psychiatric Association, 2000) and in the International Classification of Mental Disorders (ICD 10; World Health Organisation, 1993). According to ICD 10, DCDs are termed specific developmental disorders of motor function (F82). One of the most commonly used tests for the diagnosis of F82 is the Movement Assessment Battery for Children (M ABC; Henderson & Sudgen, 1992) and its successor, the M ABC 2 (Henderson, Sudgen, & Barnett, 2007; see Geuze, Jongmans, Schoemaker, & Smits Engelsman, 2001). Both tests comprise eight subtests per age band (AB) and are assigned to three dimensions: manual dexterity (MD), aiming and catching (AC) and balance (BL). M ABC as well as M ABC 2 trace back on the revision (Stott, Moyes, & Henderson, 1984) of the Test of Motor Impairment (TOMI; Stott, Moyes, & Henderson, 1972).

1.2. Problem

Several studies have examined the validity of M ABC (overview, Henderson et al., 2007; in summary, Cools, De Martelaer, Samaey, & Andries, 2008). However, while the basic structure of M ABC remained unchanged in M ABC 2, the test modalities of the two tests differ, and hence reported validity of M ABC cannot be directly extended to M ABC 2 (Brown & Lalor, 2009). A comparison of both tests can be found in Henderson et al. (2007). The studies on the validity of M ABC 2 cited in the test manual have yet to be reviewed, and only relate to content (see Henderson et al., 2007) and criterion related (e.g. Kavazi, unpublished results; Siaperas et al., unpublished results) validity. To date, information on construct validity of M ABC 2, and likewise on that of M ABC (Venetsanou et al., 2011), is not available (see Brown & Lalor, 2009; see Cools et al., 2008). However, from a theoretical perspective, construct validity is of particular importance because (i) content validity does not represent an objective parameter and (ii) criterion related validity can only be examined on the basis of adequate external criteria. The concept of construct validity relates to inferring a theoretical and not directly observable construct from a test or any other empirical indicator. More specifically, components of construct validity apply to factorial, convergent and discriminant validity.

1.3. Desideratum and objective

To date, there is no empirical evidence that the subtests of M ABC 2 facilitate a valid reproduction of the factorial structure as postulated by Henderson et al. (2007). Previous exploratory findings (Petermann, 2009, p. 135f) suggest the correctness of the assumption of factorial validity. The purpose of this study was to empirically confirm the factorial validity of M ABC 2.

2. Materials and methods

2.1. Sample

The German normative sample of age band 2 (AB2; 7-10 years; see Petermann, 2009) was used as study sample for the empirical analyses. The sample consisted of 323 children (male: N = 169, female: N = 154) with a mean age of 8.96 years (min: 7.02, max: 10.98). Data was collected between July 2007 and January 2008. The tests were mainly carried out in primary schools. To ensure the representativeness of the sample, age, gender, geographic region and the urban/rural distribution were included in the sampling plan. A comprehensive description of the sample can be found in Petermann (2009, p. 127ff).

2.2. Measurements

Measurements for eight subtests assigned to the three dimensions manual dexterity (MD), aiming and catching (AC), and balance (BL) were recorded. Manual dexterity was assessed using subtests placing pegs (MD1), threading lace (MD2) and drawing trail 2 (MD3), Aiming and catching was assessed using subtests two hand catch (AC1) and throwing bean bag onto mat (AC2). The balance scale comprised subtests one board balance (BL1), walking heel to toe forwards (BL2) and hopping on mats 2 (BL3). Table 1 shows a brief subtest description. A comprehensive subtest description can be found in Henderson et al. (2007). For subtest MD1 the attempt with the dominant hand and for subtest AC2 the better attempt was scored, respectively.

2.3. Data analysis

To verify the factorial validity of M ABC 2 (AB2), a confirmatory factor analysis (CFA) was conducted using AMOS 18 (Arbuckle, 2003). The Mardia Test was used for the assessment of multivariate normal distribution. One factor loading each was fixed to one in order to scale the latent variables (Fig. 1). The non normalized model parameters were estimated by maximum likelihood using the covariance matrix. Standardized partial regression weights and error variances (Table A.1), the intercorrelation matrix (Table A.2) as well as relevant descriptive characteristics of all subtests (Table A.3) are listed in Appendix A. To evaluate the model fit, selected incremental fit indices were referenced, including root mean square error of approximation (RMSEA), standardized root mean residual (SRMR) and comparative fit index (CFI). The selection of the indices is oriented on Beauducel and Wittmann (2005); their evaluation is based on the recommendations of Hu and Bentler.
Despite a good global fit, structural equation models can still be problematic within their substructures. Therefore, the model's local fit was assessed more comprehensively in a second step where convergent and discriminant measures were evaluated. The convergent measures assessed were indicator reliability (IR) and factor reliability (FR) as well as the average assessed variance (AAV). Fornell Larcker Ratios (FLR) and $\chi^2$ difference tests (CMIN) were used as discriminant measures. The selection and evaluation of the convergent and discriminant measures was based on previous recommendations (Homburg, Klarmann, & Pflesser, 2008). Modification indices were not included in the analysis because the CFA was intentionally used for model verification and not for exploratory model optimization. Finally, verification of an equivalent hierarchical model was not performed because the total impairment score does not represent a theoretically deduced motor construct.

Table 1
Subtest description of M-ABC-2 (AB2).

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Subtest</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD$^a$</td>
<td>MD1$^a$</td>
<td>The child holds a box steady with one hand and picks up the pegs, one at a time, and inserts them into the board as quickly as possible</td>
</tr>
<tr>
<td></td>
<td>MD2$^b$</td>
<td>The child inserts a lace through the holes of a plastic board</td>
</tr>
<tr>
<td></td>
<td>MD3$^b$</td>
<td>The child draws a single continuous line between two lines without crossing its boundaries</td>
</tr>
<tr>
<td>AC$^c$</td>
<td>AC1$^c$</td>
<td>The child throws a tennis ball from behind a marked distance at the wall and catches it with both hands</td>
</tr>
<tr>
<td></td>
<td>AC2$^c$</td>
<td>The child throws a bean bag targeting an orange circle on a mat</td>
</tr>
<tr>
<td>BL$^d$</td>
<td>BL1$^b$</td>
<td>The child balances on one foot on a balance board for up to 30 s</td>
</tr>
<tr>
<td></td>
<td>BL2$^b$</td>
<td>The child walks along a line, placing the heel of one foot against the toe of the other with each step</td>
</tr>
<tr>
<td></td>
<td>BL3$^b$</td>
<td>From a unilateral stance position, the child makes five continuous hops forward from mat to mat, stopping on a target mat</td>
</tr>
</tbody>
</table>

Adapted from Henderson et al. (2007).

(a) Placing pegs.
(b) Manual dexterity.
(c) Threading lace.
(d) Drawing trail 2.
(e) Two-hand-catch.
(f) Aiming and catching.
(g) Throwing bean-bag onto mat.
(h) One-board balance.
(i) Balance.
(j) Walking heel-to-toe forwards.
(k) Hopping on mats 2.

(1999). Despite a good global fit, structural equation models can still be problematic within their substructures. Therefore, the model's local fit was assessed more comprehensively in a second step where convergent and discriminant measures were evaluated. The convergent measures assessed were indicator reliability (IR) and factor reliability (FR) as well as the average assessed variance (AAV). Fornell Larcker Ratios (FLR) and $\chi^2$ difference tests (CMIN) were used as discriminant measures. The selection and evaluation of the convergent and discriminant measures was based on previous recommendations (Homburg, Klarmann, & Pflesser, 2008). Modification indices were not included in the analysis because the CFA was intentionally used for model verification and not for exploratory model optimization. Finally, verification of an equivalent hierarchical model was not performed because the total impairment score does not represent a theoretically deduced motor construct.

Fig. 1. Reflective model for the assessment of factorial validity of M-ABC-2 (AB2). MD, manual dexterity; AC, aiming and catching; BL, balance; MD1, placing pegs; MD2, threading lace; MD3, drawing trail 2; AC1, two-hand-catch; AC2, throwing bean-bag onto mat; BL1, one-board balance; BL2, walking heel-to-toe forwards; BL3, hopping on mats 2; e1–e8, error variables.
3. Results

3.1. Model assumptions

Applying maximum likelihood requires that the included indicators are interval scaled and multivariate normally distributed. Interval scale level was assumed for all indicators. However, the Mardia Test (Table A.3) showed a clear violation of the multivariate normal distribution assumption. Skewness and kurtosis were particularly problematic for subtests MD3, BL2 and BL3. The Bollen-Stine bootstrap procedure (200 samples) was applied to achieve necessary correction of p-values.

3.2. Global analyses

The Bollen-Stine corrected $\chi^2$ statistic revealed no significant differences between the theoretical and the empirical covariance matrix ($\chi^2 = 28.675; \text{df}=17; \text{Bollen Stine} p \text{ value} = .318$). The incremental fit indices ($\text{RMSEA} = 0.046 [0.011; 0.075], \text{SRMR} = 0.038; \text{CFI} = 0.960$) supported this finding because they were below the upper limit of RMSEA (≤0.08) and SRMR (<0.11) and above the lower limit of CFI (>0.95). In conclusion, the incremental fit indices provided evidence for the factorial validity of M-ABC-2 (AB2).

3.3. Substructure analyses

For convergent measures (Table 2), each estimated factor loading differed significantly from zero ($t \geq 1.645$). However, factor reliability of latent variables AC and BL was well below the minimum requirement of $r \geq 0.60$ (see FR, Table 2). In addition, an average of less than 50% of the total variance of each indicator block was explained by the superordinated latent variables MD, AC and BL (see AAV, Table 2). Hence, the latent variables only showed limited explanatory power for the reflective indicators or, in other words, the sum of all indicators showed insufficient measurement accuracy for the superordinated latent variables. Indicators drawing trail 2 (MD3), two hand catch (AC1), walking heel to toe forwards (BL2), and especially hopping on mats 2 (BL3) are particularly unreliable because less than 40% of the variance of these indicators was explained by the superordinated latent variables (see IR, Table 2).

The assessment of discriminant measures was less consistent. The average variance of each latent variable captured (see AAV, Table 3) was lower than the highest squared intercorrelation (see MSI, Table 3) with all other latent variables in the model ($\text{̄FLR} > 1.00$, Table 3). This result is a clear indication for a lack of discrimination. In contrast, fixing the correlation of two factors to one (Fig. 2) resulted in a significantly inferior fit of all three possible restricted models ($\text{CMIN a, b, c} > 3.84$). Consequently, a superior two way nested model does not appear to exist.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Subtest</th>
<th>IR$^e$</th>
<th>$t$</th>
<th>FR$^c$</th>
<th>AAV$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD$^g$</td>
<td>MD1$^f$</td>
<td>0.44</td>
<td>-f</td>
<td>0.62</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>MD2$^h$</td>
<td>0.45</td>
<td>7.51</td>
<td>0.53</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>MD3$^i$</td>
<td>0.18</td>
<td>5.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ac$^k$</td>
<td>AC1$^j$</td>
<td>0.15</td>
<td>-f</td>
<td>0.43</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>AC2$^l$</td>
<td>0.44</td>
<td>2.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BL$^m$</td>
<td>0.46</td>
<td>-f</td>
<td>0.53</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>BL1$^o$</td>
<td>0.29</td>
<td>6.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BL2$^p$</td>
<td>0.08</td>
<td>4.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- $^a$ Indicator reliability.
- $^b$ $t$-Value of factor loading.
- $^c$ Factor reliability.
- $^d$ Average assessed variance
- $^e$ Placing pegs.
- $^f$ Standardized path.
- $^g$ Manual dexterity.
- $^h$ Threading lace.
- $^i$ Drawing trail 2.
- $^j$ Two-hand-catch.
- $^k$ Aiming and catching.
- $^l$ Throwing bean-bag onto mat.
- $^m$ One-board balance.
- $^n$ Balance.
- $^o$ Walking heel-to-toe forwards.
- $^p$ Hopping on mats 2.
4. Discussion

The purpose of this study was to empirically confirm the factorial validity of the Movement Assessment Battery for Children 2 (M-ABC 2) using the German normative data for age band 2 (AB2). The global fit measures provided evidence for the assumption of factorial validity. However, the model proved to be problematic within its sub structures. The results of our study confirm doubts on the discriminant validity but even more so on the convergent validity of the M-ABC 2 (AB2). However, the examination of the inner method convergence/divergence is not sufficient for a final assessment of convergent and discriminant validity (see Steenkamp & van Trijp, 1991), because no other construct close or construct remote methods were used for comparison. Regarding a comprehensive assessment of the construct validity of M-ABC 2 (AB2), we thus propose to examine the convergent and discriminant validity of subtests based on the multitrait multimethod (MTMM) approach. A revision or elimination of the affected subtests drawing trail 2 (MD3), two-hand-catch (AC1), throwing bean-bag onto mat; one-board balance; walking heel-to-toe forwards; hopping on mats 2 should only be considered if the convergence and divergence problems found in this study are confirmed by MTMM based analysis. In addition, factorial, convergent and discriminant validity should be verified for the subtests of AB1 and AB3.

5. Conclusion

The proof of evidence for the factorial validity of M-ABC 2 (AB2) shown in this study supports their use in therapeutic practice. Nevertheless, because of the identified lack of empirical evidence for the convergent and discriminant validity, to
date M ABC 2 (AB2) cannot be used for conclusively diagnosing F82. Therefore, we support the previous recommendation by Brown and Lalor (2009) that other methods are still needed in addition to M ABC 2 in clinical practice.

Disclosure statement

The authors declare that no actual or potential conflict of interest exists.

Acknowledgements

The authors would like to thank Dr. Annegret Mündermann (ABR Solutions) who provided scientific writing services on behalf of the authors. Furthermore, the authors would like to thank Prof. Dr. Martin Klarmann (University of Mannheim) and Prof. Dr. Hans Christian Waldmann (University of Bremen) for the productive and collegial knowledge exchange.

Appendix A

See Tables A.1 A.3.

Table A.1

Standardized partial regression weights and error variances of subtests of MABC-2 (AB2).

<table>
<thead>
<tr>
<th>Subtest</th>
<th>MD</th>
<th>AC</th>
<th>BL</th>
<th>Error variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD1a</td>
<td>0.66</td>
<td>–</td>
<td>–</td>
<td>0.56**</td>
</tr>
<tr>
<td>MD2a</td>
<td>0.67</td>
<td>–</td>
<td>–</td>
<td>0.55**</td>
</tr>
<tr>
<td>MD3a</td>
<td>0.42</td>
<td>–</td>
<td>–</td>
<td>0.82**</td>
</tr>
<tr>
<td>AC1a</td>
<td>–</td>
<td>0.39</td>
<td>–</td>
<td>0.85**</td>
</tr>
<tr>
<td>AC2a</td>
<td>–</td>
<td>0.66</td>
<td>–</td>
<td>0.56**</td>
</tr>
<tr>
<td>BL1a</td>
<td>–</td>
<td>–</td>
<td>0.68</td>
<td>0.54**</td>
</tr>
<tr>
<td>BL2a</td>
<td>–</td>
<td>–</td>
<td>0.54</td>
<td>0.71**</td>
</tr>
<tr>
<td>BL3a</td>
<td>–</td>
<td>–</td>
<td>0.29</td>
<td>0.92**</td>
</tr>
</tbody>
</table>

a Manual dexterity.

b Aiming and catching.

c Balance.

d Placing pegs.

e Threading lace.

f Drawing trail 2.

g Two-hand-catch.

h Throwing bean-bag onto mat.

i One-board balance.

j Walking heel-to-toe forwards.

k Hopping on mats 2.

**p < 0.01.

Table A.2

Intercorrelations of dimensions and subtests of MABC-2 (AB2).

<table>
<thead>
<tr>
<th>Ac</th>
<th>MD</th>
<th>BL</th>
<th>BL3</th>
<th>BL2</th>
<th>BL1</th>
<th>AC2</th>
<th>AC1</th>
<th>MD3</th>
<th>MD2</th>
<th>MD1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ac</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>–0.39</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL</td>
<td>0.53</td>
<td>–0.74</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL3</td>
<td>0.15</td>
<td>–0.22</td>
<td>0.29</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL2</td>
<td>0.29</td>
<td>–0.40</td>
<td>0.54</td>
<td>0.16</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL1</td>
<td>0.36</td>
<td>–0.50</td>
<td>0.68</td>
<td>0.20</td>
<td>0.37</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC2</td>
<td>0.66</td>
<td>–0.26</td>
<td>0.35</td>
<td>0.10</td>
<td>0.19</td>
<td>0.24</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC1</td>
<td>0.39</td>
<td>–0.15</td>
<td>0.21</td>
<td>0.06</td>
<td>0.11</td>
<td>0.14</td>
<td>0.26</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD3</td>
<td>–0.16</td>
<td>0.42</td>
<td>–0.31</td>
<td>–0.09</td>
<td>–0.17</td>
<td>–0.21</td>
<td>–0.11</td>
<td>–0.06</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>MD2</td>
<td>–0.26</td>
<td>0.67</td>
<td>–0.50</td>
<td>–0.14</td>
<td>–0.27</td>
<td>–0.34</td>
<td>–0.17</td>
<td>–0.10</td>
<td>0.28</td>
<td>1.00</td>
</tr>
<tr>
<td>MD1</td>
<td>–0.26</td>
<td>0.66</td>
<td>–0.49</td>
<td>–0.14</td>
<td>–0.27</td>
<td>–0.33</td>
<td>–0.17</td>
<td>–0.10</td>
<td>0.28</td>
<td>0.44</td>
</tr>
</tbody>
</table>

a Aiming and catching.

b Manual dexterity.

c Balance.

d Hopping on mats 2.

e Walking heel-to-toe forwards.

f One-board balance.

g Throwing bean-bag onto mat.

h Two-hand-catch.

i Drawing trail 2.

j Threading lace.

k Placing pegs.
Table A.3
Descriptive characteristics (N = 323) of subtests of MABC-2 (AB2).

<table>
<thead>
<tr>
<th>Subtest</th>
<th>M</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD1f</td>
<td>26.15</td>
<td>4.50</td>
<td>14.00</td>
<td>46.00</td>
<td>0.94</td>
<td>6.88</td>
</tr>
<tr>
<td>MD2g</td>
<td>23.54</td>
<td>7.38</td>
<td>12.00</td>
<td>53.00</td>
<td>1.34</td>
<td>9.83</td>
</tr>
<tr>
<td>MD3h</td>
<td>0.52</td>
<td>1.03</td>
<td>0.00</td>
<td>7.00</td>
<td>2.82</td>
<td>20.67</td>
</tr>
<tr>
<td>AC1i</td>
<td>7.57</td>
<td>2.28</td>
<td>0.00</td>
<td>10.00</td>
<td>-0.97</td>
<td>-7.12</td>
</tr>
<tr>
<td>AC2j</td>
<td>6.66</td>
<td>1.96</td>
<td>1.00</td>
<td>10.00</td>
<td>-0.38</td>
<td>-2.81</td>
</tr>
<tr>
<td>BL1k</td>
<td>22.73</td>
<td>9.33</td>
<td>1.00</td>
<td>30.00</td>
<td>-0.82</td>
<td>-6.05</td>
</tr>
<tr>
<td>BL2l</td>
<td>14.50</td>
<td>1.96</td>
<td>2.00</td>
<td>15.00</td>
<td>-4.38</td>
<td>-32.17</td>
</tr>
<tr>
<td>BL3m</td>
<td>4.94</td>
<td>0.34</td>
<td>2.00</td>
<td>5.00</td>
<td>-6.38</td>
<td>-46.79</td>
</tr>
</tbody>
</table>

Multivariate

90.43 64.25

References


