An evaluation of the Movement ABC-2 Test for use in Italy:
A comparison of data from Italy and the UK

Stefania Zoia[*], Marina Biancotto[#], Marco Guicciardi[**], Romina Lecis[***], Fabio Lucidi[***], Giovanna M. Pelamatti[#], Marco Carrozzi[¶], Aldo Skabar[¶], David A. Sugden[^], Anna L. Barnett[°], Sheila E. Henderson[§]

[*]Struttura Complessa Tutela Salute Bambini Adolescenti Donne Famiglia, Azienda Sanitaria Universitaria Integrata di Trieste, Trieste, Italy
[**]University of Cagliari, Italy
[***]University of Rome, La Sapienza, Italy
[#]University of Trieste, Italy
[¶] Institute for Maternal and Child Health - IRCCS "Burlo Garofolo, Trieste
[^]University of Leeds, UK
[°]Oxford Brookes University, UK
[§]University College London, UK

Corresponding author:
Dr. Stefania Zoia
Struttura Complessa Tutela Salute Bambini Adolescenti Donne Famiglia
Azienda Sanitaria Universitaria Integrata di Trieste
via Sai 7, 34100 Trieste, Italy
E-mail: stefania.zoia@asuits.sanita.fvg.it
        zoia.stefania@libero.it

Abstract

Background. The standardized test within the Movement Assessment Battery for Children-2nd edition (MABC-2) is used worldwide to assess motor problems in children. Ideally, any country using a test developed in another country should produce national norms to ensure that it functions effectively in the new context.

Aim. The first objective of this study was to explore the differences in motor performance between Italian and British children. The second was to examine the structural validity of the test for the Italian sample.

Method. A total of 718 Italian (IT) and 765 British (UK) children, aged 3 to 10 years, were individually tested on the age-appropriate items of the MABC-2 Test.

Results. Developmental trends emerged on every task and differences between IT and UK children were obtained on 11 of 27 task comparisons. Interactions between age and country indicated that differences were not consistently in favor of one culture. Confirmatory factor analysis generally supported the proposed structure of the MABC-2 Test.

Conclusion. Although the differences between the IT and the UK children were relatively few, those that did emerge emphasize the need for population specific norms and suggest that cultural diversity in motor experiences should be considered when evaluating motor abilities in children.

What this paper adds
- The MABC-2 Test successfully discriminates among Italian children aged 3-10 years, demonstrating that the item content is suitable for an Italian population, as shown for every task by the age effect ($p < .001$, with a partial effect size ranging from moderate to high, η_p^2: .038 - .37 in most cases).
- Differences between Italian and British children in performing the MABC-2 Test emphasize the need for local norms as these variances occurred within all three areas of competence, although only 41% of comparisons between IT and UK children on the individual Test tasks were significant ($p<.01$) and showed a partial effect size from low to moderate (η_p^2: .014 - .09).
- The confirmatory factorial analysis (CFA) provides good support for the validity of the MABC-2 Test, corroborating the Test structure, with its three competence areas (for AB1: Satorra-Bentler X^2 (df = 23) = 57.42, $p < .01$; for AB2: Satorra-Bentler X^2 (df = 30) = 78.46, $p < .01$).

Highlights

- For the first time, the Italian and British raw scores of the MABC-2 Test were compared.
- This study starts to examine the properties of the MABC-2 Test for use in Italy.
- The Italian norms support the good structural validity of the MABC-2 Test.

Keywords

Movement Assessment Battery for Children-2
Manual dexterity
Aiming and Catching
Balance
Standardized test

Running title: The Movement ABC-2 Test for use in Italy

1. Introduction

The Movement Assessment Battery for Children-2nd edition (MABC-2; Henderson, Sugden & Barnett, 2007; Italian adaptation: Biancotto, Borean, Bravar, Pelamatti, & Zoia, 2013; Italian standardization: Biancotto, Guicciardi, Pelamatti, Santamaria, & Zoia, 2017) is one of the most commonly used tools for evaluating motor difficulties in children (see Blank, Smits-Engelsman, Polatajko, & Wilson, 2012), with ten published translations and two more in preparation. The standardized test included in the Battery (henceforth MABC-2 Test), along with a behavior checklist and an intervention manual, is widely used both in clinical and research settings, and has been shown to be effective with a very wide range of children with neurodevelopmental disorders. For example, recent studies have included children with developmental coordination disorder (DCD; Cheng, Ju, Chang, Chen, Pei, Tseng, & Cheng, 2014), children with attention problems (Howen, van der Veer, Visser, & Cantell, 2017), children within the autistic spectrum (Hanaie, Mohri, Kagitani-Shimono, & Taniike, 2016), children with Specific Language Impairment (SLI; Webster, Erdos, Evans, Majnemer, Kehayia, Thordardottir, Evans, & Shevell, 2006), children with dyslexia (Getchell, Pabreja, Neeld, & Carrio, 2007) and many other groups (e.g. Hill, Pratt, Kanji, & Jones Bartoli, 2017). In addition to its value as a measure of current difficulties, the MABC-2 Test has also been shown to have good predictive value. For instance, Griffiths et al. (2017) found that the Test results of preterm children, examined at 4 years of age, accurately predicted their motor status at 8 years. One of the reasons for the popularity of the MABC-2 Test is that its structure and mode of administration make it easy to use and enjoyable for children to perform. The test has three sub-components, labeled 'Manual Dexterity' (MD), 'Aiming & Catching' (AC) and 'Balance' (BAL). The three MD tasks focus on how hands are used to manipulate objects. These include a measure of speed and accuracy for each hand separately, a timed bi-manual task and an untimed drawing task. The AC component includes tasks of throwing an object towards a target and of catching an object with either one or two hands. The BAL component includes a static balance task and two dynamic balance tasks. The age range of the MABC-2 Test extends from 3 to 16 years, divided into three age bands (ABs). A similar set of 8 tasks is included within each AB across the three sub-components, with tasks increasing in difficulty according to age. The development of the rationale for this division and the selection of tasks within each AB has been a work in progress since the original publication of the test in 1972 and its subsequent revisions in 1984 (Stott, Moyes, & Henderson, 1972, 1984), in 1992 (Henderson & Sugden, 1992) and most recently in 2007 (Henderson, Sugden, & Barnett, 2007).

The aims of the MABC-2 Test are twofold. First, it provides an objective measure of motor performance, which can be used to diagnose conditions such as DCD or to recognize motor problems in different neurodevelopmental disorders. Second, it is designed to help clinicians in planning a motor intervention program. Following the World Health Organization guidelines, an effective intervention program for any child with a neurodevelopmental disorder should include an attempt to improve daily living activities, alongside or instead of, an individualized treatment provided by a healthcare professional (ICF-CY, WHO, 2012). The MABC-2 Test provides useful information for healthcare professionals and educators, both to plan individual treatment and to give parents, teachers, sport-trainers and other significant adults, information about daily living activities which can help children. The MABC-2 Test provides: (a) standardized scores for each of the components, thus showing a performance profile in addition to a single total score, (b) a record of qualitative observations on how the child performs each task and (c) an overview of a number of non-motor observations, which are useful when judging how a child might respond to intervention. The Ecological Intervention manual, included in the Battery, explains how these data can be used to teach or improve specific skills, as well as to promote positive motor experiences, based on environmental and social resources (Sugden & Henderson, 2007). Since its publication in 2007, the ten manuals translated into languages other than English have reported data on the psychometric properties of the MABC-2 Test, along with normative data specific for the countries concerned (e.g. Marquet-Doléac, Soppelsa, & Albaret, 2016; van Waelvelde, Peersman, & Smits-Engelsman, 2010). In addition, peer reviewed papers have appeared in various journals. Examples of studies assessing reliability come from at least three countries, in some cases leading to different conclusions. In the Netherlands, Smits-Engelsman et al. (2011) reported good test-retest and inter-tester reliability for 3 year olds, an age at which reliability can be difficult to test. In China, Hua et al. (2013) drew similar conclusions in a sample ranging from 36 to 72 months, as did Valentini et al. (2014) in Brazil, with children aged 3-13 years. In contrast, Holm et al. (2013) in Norway concluded that data drawn from the Test for AB2 were only suitable for clinical decisions and did not reliably measure change over time.

With regard to validity, once again, different approaches have been taken but direct comparisons between the original UK standardization data with data collected in another country are popular starting points. Often these comparisons are difficult to interpret, as the sampling techniques employed do not always lead to truly representative countrywide samples. However, when appropriate sampling methods and statistics are applied, they can be useful. For instance, Ellinoudis et al. (2011) found that the MABC-2 Test was suitable for use with Greek children aged 3-5 years.

Psotta et al. (2012) suggested that the UK norms could be used without alteration for 7- 8 year old Czech boys, while norm adjustments were required for 9-10 year old children. Niemeijer et al. (2015) compared data from a representative sample of Dutch and Flemish children with the UK data and found that new norms were essential for this population at all ages. Valentini et al. (2014) translated all MABC-2 items, examined them with respect to content, construct and criteria validity for a Brazilian-Portuguese population and suggested that the original standardized scores were valid for Brazilian children. However, these authors admit that their sample, although large (N=844), was not representative of the whole population. Finally, Kita et al. (2016) compared data from 132 typically developing Japanese children, aged 7-10 years, with the UK normative sample and found differences in MD and BAL. To date, no study that reports a comparison between raw data from British (UK) and Italian (IT) children has been published.

An important validity test of an assessment instrument involves examining its internal structure and the extent to which its items cluster the way the authors intended. To validate the MABC-2 Test, Schulz et al. (2011) described the structural validity for each AB separately. Their results supported the three-component structure of the Test, which originated from a theoretical analysis of the motor demands of everyday life (Rarick, Dobbins, & Broadhead, 1976). Their findings also showed that the Test structure remains identical across each AB. More recently, Ellinoudis et al. (2011) obtained similar results for AB1 using their data on Greek children aged 3-6 years. In contrast, Hua et al. (2013) obtained a slightly different solution for their sample of 1823 Chinese children, aged 3 to 6 years. Using confirmatory factor analysis, they obtained goodness-of-fit indices (all above the desirable level of .90) for an adjusted model of the Test, which excluded two tasks (MD3 'Drawing Trial' task and Dynamic BAL2 'Walking Heels Raised' task). In relation to AB2 of the MABC-2 Test, Wagner et al. (2011) described the factorial validity of the Test in a German sample as similar to that found in the UK sample. Even in Japan the confirmatory factorial validity revealed significant factor loadings on all three components (MD, AC and BAL) into the eight individual items (Kita, Suzuki, Hirata, Sakihara, Inagaki, & Nakai, 2016).

Considering the reported similarities and differences between children's performances on the MABC-2 Test in different countries around the world, the aim of the present study was to examine its suitability for use in Italy. Thus, our first objective was to explore possible differences in MABC-2 Test performance between two samples of children from Italy and the UK. Secondly, we set out to compare the factorial validity of the MABC-2 Test as it emerged from our IT data with that obtained in the UK. Since the Battery is most frequently used with children aged 3 to 10, our focus was on this age range for both aims.

2. Method

2.1. Participants

The IT sample consisted of 338 children between the ages of 3 and 6 years (AB1 of MABC-2 Test) and 380 children between the ages of 7 and 10 years (AB2). This group provided part of the data collected for the full IT standardization of the MABC-2 Test (Biancotto, Guicciardi, Pelamatti, Santamaria, & Zoia, 2017). Within each AB, the proportion of males and females was similar (162 females and 176 males for AB1; 199 females and 181 males for AB2). Geographically, the children came from four regions of Italy: Friuli-Venezia-Giulia and Veneto in the north-east, Lazio in the center and Sardinia, an island to the west of the mainland. Both urban and rural schools were involved in order to guarantee proper representativeness of the population. Overall, 11 schools were involved (5 kindergartens and 6 primary schools). All children were tested within their own schools, as in the UK.

Any child with a known neurological condition, psychiatric disorder, intellectual disability and/or behavioral problem was excluded from the sample. All other children were considered to be eligible, provided their parents agreed. Thus, children with learning difficulties and children with a specific learning disorder involving reading, spelling, handwriting and/or math were included.

The UK sample employed in this study was part of the full standardization covering the whole age range of the MABC-2 Test published in 2007 (Henderson, Sugden, & Barnett, 2007). Data were provided by the Test authors from a stratified sample of 431 children, aged between 3-6 years (AB1 of MABC-2 Test) and 333 children aged 7-10 years (AB2). Demographic information about the two samples is reported in Table 1.

Insert Table 1 about here

2.2. The MABC-2 Test

The MABC-2 Test is divided into three ABs: AB1 for 3-6 year olds; AB 2 for 7-10 year olds and AB 3 for 11-16 years olds. Within each AB, 8 tasks are grouped under the three headings: MD, AC and BAL. The items for each of the two ABs considered in this study, along with the respective measures obtained, are shown in Table 2. For some tasks, a separate assessment of both hands and of both legs is required, leading to a different raw score for each hand/leg. The raw scores available for each task vary and include 'time in seconds', 'number of errors' and 'number of correct attempts'. For some tasks, a higher score indicates better performance (e.g. number of seconds

balanced, number of correct jumps), whereas for others a lower score indicates better performance (e.g. completion time, number of errors). As mentioned above, in addition to quantitative data, the Test provides qualitative data on how each child performs a task. For the purposes of the present study, however, only the quantitative raw scores were considered.

Insert Table 2 about here

2.3. Recruitment procedure, tester training and administration.

Children aged between 3 and 10 years were invited to participate in the project. Families were contacted through local kindergartens and primary schools. Prior to testing, participants and their parents received a detailed explanation about the study and all parents provided written informed consent. The ethics board of the IRCCS *Burlo Garofolo* Mothers' and Children's Hospital in Trieste approved the research protocol, as well as the ethics committees of the Universities of Cagliari and Rome.

All children were individually assessed according to the instructions given in the MABC-2 Examiner's Manual (Henderson, Sugden, & Barnett, 2007) by psychologists and physiotherapists trained to use the Test. The training consisted of reading the Italian translation of the manual and practicing the Test administration with a group of children not included in the main sample. After each tester had scored the performance of their own "practice subjects", a team session took place in which experiences were shared and critical points discussed. Later, each examiner was requested to watch an instruction video of an expert administering the Test to children with typical and atypical motor abilities and then to score the observed performances. Any further problems or queries were addressed and the possible differences between typical and DCD children were discussed.

Each child was individually assessed in the school gym during two testing sessions, for a total of 60 minutes. The MD tasks were administered in the first session, while the BAL and AC tasks were performed in the second one. The order was counterbalanced between participants.

3. Data Analysis

Before proceeding with the analyses of variance, the underlying assumptions were tested with regard to skewness and kurtosis. Although some evidence of non-normality emerged for some tasks, these violations were sufficiently small to allow us to proceed with the Analysis of variance (ANOVA). In support of this decision, several studies assessing the robustness of ANOVA in

relation to violation of the assumptions in medium to large sample size were consulted. In particular, the study by Schmider et al. (2010) provides strong support for the robustness of applying ANOVA to non-normally distributed data and discourages the use of non parametric tests due to the loss of precision that comes along with transformations into rank data, and lower power (Tanizaki, 1997) and inaccuracy in case of multiple violations (Zimmerman, 1998).

In order to compare the IT and UK samples on each item of the MABC-2 Test, the raw scores obtained for each task for each age year within each AB were employed. ANOVAs were then conducted with age and country as between subject factors. Due to the large number of comparisons being conducted, Bonferroni correction was applied in each case and we considered only the main effects statistically significant at the 99% level (i.e., $p \leq .01$), along with the possible interaction effects. When the tasks had different requirements within one AB, separate one-way ANOVAs had to be performed. For example, the ′Posting Coins′ task (MD1) in AB1 requires children aged 3-4 years to insert 6 coins into a box, while 5-6 year olds have to insert 12 coins. These differences, along with the above mentioned distinct raw scores for each hand/leg for some tasks, led to 15 separate comparisons in AB1 and 12 in AB2. When an interaction effect emerged as significant, post-hoc comparisons were done by using LSD test (Fisher's Least Significant Difference) and Cohen's effect size was calculated for each comparison.

The raw data from the IT sample was also used to compute the covariance matrices, which formed the basis of the confirmatory factor analysis (CFA), separately for each AB. The aim then was to employ structural equation modeling (SEM), using LISREL 8.8, to describe the factor structure underlying the correlations between tasks within each of the two Abs, as had been done for the UK sample (Schulz, Henderson, Sugden, & Barnett, 2011). The method applied in this study strictly replicated what had been done by Schultz et al. (2011), with the aim of verifying the three-component design of the MABC-2 Test, as postulated by Henderson et al. (2007). Hence, for each AB, a CFA was first tested, followed by a more explorative phase of analysis.

4. Results

4.1. Comparisons between Italian and British children.

Tables 3 and 4 show the means and standard deviations for the IT and UK children's scores on each task in AB1 and AB2, respectively.

Insert Tables 3 and 4 here

ANOVAs performed on these data revealed that a main effect of Age was consistently present for all tasks (p < .001 in each case), except for Dynamic BAL2 task in AB1, which showed a ceiling effect from age five. The partial effect size for these significant main effects ranged from moderate to high (η_p^2: .038 - .37).

Considering the two ABs together, a main effect of Country emerged for 11 of the 27 comparisons made between IT and UK children's raw scores in Test tasks (p < .01), with partial effect sizes from low to moderate (η_p^2: .014 - .09). There were 5 significant interactions between Age and Country, which are further explored below.

For AB1, examination of the individual task differences showed statistically significant comparisons in all three components of the Test. On MD1 for 5-6 year old children, a main effect of Country on the preferred hand emerged ($F_{(1,348)} = 7.626$; p = .006), along with a significant Age X Country interaction ($F_{(1,348)} = 4.853$; p = .028). Post-hoc comparisons then revealed that only 5 year old IT children were slower than their UK peers in inserting the coins ($F_{(1,348)} = 15.129$, p < .001; Cohen's d: .5). On MD3, a main effect of Country also emerged, with IT children making more errors than UK children in the 'Drawing Trail' task ($F_{(1,762)} = 10.774$; p = .001) at all four ages. Moving to the AC subcomponent, a main effect of Country was found for AC1 in the 3-4 years range ($F_{(1,414)} = 9.536$; p = .002), along with an interaction effect ($F_{(1,414)} = 4.103$; p = .043): post-hoc comparisons revealed that, at 3 years of age, IT children made fewer catches than their UK peers ($F_{(1,414)} = 10.985$, p = .001; Cohen's d: .48). Finally, a significant difference between IT and UK children was found for Dynamic BAL1 ($F_{(1,762)} = 42.76$; p < .001), where IT children made less correct steps than their UK counterparts. Overall, examination of these few significant differences in AB1 revealed that IT children obtained lower scores than UK children.

In AB2 (7-10 years), the effect of Country was found to be significant for the MD1, both for the preferred ($F_{(1,705)} = 7,581$; p = .006) and the non-preferred hand ($F_{(1,705)} = 23,003$; p = .001), along with the corresponding interaction between Age and Country (preferred hand: $F_{(3,705)} = 8.631$; p < .001; non-preferred hand: $F_{(3,705)} = 4.868$; p = .002). Post-hoc comparisons showed that IT children were faster at placing pegs compared to their UK peers only at the age of 7 with the preferred hand ($F_{(1,705)} = 31.873$; p < .001; Cohen's d: .56, see Figure 1 panel a) and at the ages of 7 and 9 with the non-preferred hand ($F_{(3,705)} = 4,868$; p= .002; see Figure 1 panel b; Cohen's d: .61 for 7 years and .57 for 9 years).

Insert Figure 1 panel a and b here

Regarding the AC subcomponent of the MABC-2 Test, the Country effect was found only for the AC1 task at age 9-10 years ($F_{(1,319)} = 31,659$; $p < .001$): the IT children achieved a higher number of correct catches compared to their UK peers.

With respect to the BAL subcomponent, the Country effect proved to be significant for Static BAL on both legs (best leg: $F_{(1,705)} = 13,581$; $p < .001$; other leg: $F_{(1,705)} = 21,675$; $p < .001$), revealing that IT children maintained balance longer than their UK peers. A Country main effect also emerged for the Dynamic BAL1 task ($F_{(1,705)} = 32,423$; $p < .001$), along with an interaction effect Age X Country ($F_{(3,705)} = 4.270$; $p = .005$): the post-hoc comparisons revealed a better result for the UK children (higher number of correct steps in walking forward heel-to-toe) compared to the IT children only at the age of 7 ($F_{(1,705)} = 35.28$, $p < .001$, Cohen's d: .66) and 8 years ($F_{(1,705)} = 11.331$, $p = .001$, Cohen's d: .47; see Figure 2).

Insert Figure 2 here

Finally, a main effect of Country was found for the Dynamic BAL2 task with 'other leg' ($F_{(1,705)} = 12,768$; $p < .001$), in which the IT children made a higher number of correct hops compared to UK peers.

4.2. Confirmatory Factor Analysis (CFA)

The preliminary correlation matrices between the tasks for each AB are displayed in Table 5.

Insert Table 5 here

11

According to the method used by Schultz et al. (2011), a three correlated factors model, with a simple structure loading matrix, no double loadings allowed and no correlated measurement errors, was first tested for each AB as a CFA. This model was rejected for both ABs, due to the results of the global goodness of fit test and several fit indices.

Moving from a strict confirmatory phase to more explorative testing, as done by Schulz et al. (2011) for the UK data, in AB1 a model with a general factor, as well as task specific factors, was tested. However, this model, which resulted in the best-fit model for the UK data, had to be rejected for the IT data, due to its poor fit. Therefore, the best-fit model for the IT data was found by introducing post-hoc modifications that allowed double loadings and error correlations. This model identified the three specific motor ability factors, without a general factor (see Figure 3), producing an acceptable fit model, Satorra-Bentler X^2 (df = 23) = 57.42, p < .01, with all the fit indices ranging from acceptable to excellent: RMSEA = .067 (the p-value for test of close fit was high: p = .096), NNFI = .96, AGFI = .93, SRMR = .054; the largest standardized residual was 3.46 and all model parameters were significant (t-value > 1.96). In this case, all items had a modest to substantial loading on the relevant factor, with the exception of the MD3 task, which was excluded because of its weak correlations with the other two MD tasks. Furthermore, since double loadings had been allowed, MD2 had a substantial loading on its relevant factor and a modest negative loading on the BAL factor. A measurement error correlation was found between the Dynamic BAL tasks, pointing toward another systematic source of variation. Lastly, the correlations between the three factors ranged from null (MD-AC) to substantial (AC-BAL).

Insert Figure 3 here

For AB2, the IT model structure was similar to the UK one (see Figure 4). The best fit model for the IT data was characterized by no double loadings and two measurement error correlations for the MD1 tasks and the Dynamic BAL tasks. All items had a modest (Dynamic BAL tasks) to substantial loading on their respective factor. As for the UK sample, the IT sample model identified the three specific motor ability factors (MD, AC and BAL), and diversely from the UK sample, a unique BAL factor. Furthermore, the correlations among the three factors were substantial, suggesting sufficient common variance to assume a higher order general factor, which in turn strongly correlated with each specific factor. This second order CFA model produced an acceptable model fit, Satorra-Bentler X^2 (df = 30) = 78.46, p < .01, with all the fit indices ranging from acceptable to excellent: RMSEA = .065 (the p-value for test of close fit was high, p = .073), NNFI

= .95, AGFI = .92, SRMR = .067; the largest standardized residual was 3.15 and all model parameters were significant (t-value > 1.96).

Insert here Figure 4 here

----

5. Discussion

The MABC-2 Test was designed to identify children with movement difficulties between the ages of 3 and 16 years. Based on the principle that the Test items should be manageable by the majority of children within this age range, the tasks deemed suitable for a particular AB remain the same but increase in difficulty between ABs. Within each AB, developmental trends are expected and are indeed found in most studies. In the present study, significant and strong main effects of age were found for each task in both AB1 and AB2 suggesting that the motor performance of IT children between the ages of 3 and 10 is similar to that of UK children. When the differences between the two countries were examined, however, the overlap was not perfect and cultural differences were found on tasks within both AB1 and 2. The nature and direction of these differences, along with any implications for the development of country-specific norms, are addressed in the paragraphs below.

5.1. Item by item review of AB1 tasks (3-6 years)

Statistically significant differences between groups of children on a single task may or may not be clinically significant. For example, with a big enough sample, a mean difference of a couple of seconds on a timed task might reach statistical significance, as was the case for many timed tasks at AB1. However, when there are a number of significant differences in one direction, this is likely to indicate a real cultural difference with children's performance being under or over estimated when their scores are aggregated. As long as the tasks are deemed suitable for the culture involved, this problem can be solved by establishing country-specific norms. In this study, all of the IT children between the ages of 3 and 6 years enjoyed the tasks but, overall, their performance was slightly less proficient. For example, on the timed MD tasks (MD1 and 2), UK children between 3 and 6 years were faster when posting coins in a box and, except for the 3 year olds, also faster when threading beads. At all ages, the IT children made more errors when completing a drawing task requiring precision (MD3). Interestingly, the difference on this last task was similar for Dutch, Chinese and Brazilian children (Smits-Engelsman, Niemeijer, & van Waelvelde, 2011; Hua, Gu, Meng, & Wu, 2013; Valentini, Ramalho, & Oliveira, 2014). To account for the number of errors made by the 3 year old children, Smits-Engelsman et al. (2011) noted the high number of specific instructions

involved, e.g. not lifting the pen, not turning the paper more than 45° and not changing drawing direction. These researchers also suggested that a limited attention span might influence the overall performance on the MABC-2 Test at this age. In IT schools, drawing activities are included but creativity and use of color are given more emphasis than line precision, which becomes relevant only when children start primary schools at the age of six.

In the areas of ball skills and balance, cultural differences were fewer but, once again, the direction of differences between IT and UK children pointed to the IT children being less proficient. For instance, in the beanbag catching task (AC1) there was an interaction between age and country with 3 to 4 year old IT children being less proficient and IT 5 to 6 year olds being more similar to their UK peers. When asked to walk along a line with heels raised (Dynamic BAL1), the IT children were less proficient and made fewer correct steps. Our observation of activities in schools suggest that IT children are less familiar with dynamic activities which emphasize accuracy rather than speed and consequently showed less integration of proprioception and on-line visual feedback.

In summary, our data on IT children aged 3 to 6 support the idea that country specific norms should be produced and that these in turn would allow practitioners to identify children with difficulties more accurately.

5.2. Item by item review of AB2 Tasks (7-10 years)

At AB2, the differences between UK and IT children showed a different pattern from AB1. Among MD tasks, for example, only the 'Placing Pegs' (MD1) task differed between IT and UK. In fact, IT children in AB2 performed significantly faster when using both the preferred and the non preferred hand than their UK peers, a change which might be due to increased emphasis on activities that require a precision grip between preschool and primary school in Italy. In the AC sub-component, the IT children obtained a higher number of catches with two hands (AC1) but only at 9 and 10 years, a result which is likely linked to the kind of sport experiences IT children are exposed to. At the age of 9, around 60% of IT children in our sample played ball games at least three times per week (volleyball, basketball, rugby, baseball, football, etc.). In addition, IT children often practice ball games during physical educational classes and they frequently spend their leisure time in these activities. In the BAL sub-component, the direction of the differences between countries was also different from AB1. More precisely, IT children of 7-10 years were able to balance longer on one foot (Static BAL) than UK peers on both legs. Similarly, on the BAL task labeled 'Hopping on Mats' (Dynamic BAL2), IT children aged 7-10 years were better, with a slightly more pronounced difference on the 'other leg'. Conversely, as in AB1, IT children aged 7 and 8 found the

BAL task requiring precision, i.e. 'Walking heel-to-toe forwards along a line' (Dynamic BAL1), more difficult than UK children. Although walking along a line represents an ecological task also for IT children, the requested precision in performance is quite far from their daily living experience. When the action requires a good trade-off between speed and accuracy (stressing particularly the last), the IT children seemed to be less used to integrating proprioceptive information and visual feedback. These findings highlight the different functionally relevant aspects of motor control which characterized each static and dynamic task in the BAL subcomponent, as well as the differences linked to the leg used (the MABC-2 Test doesn't emphasize foot dominance but distinguishes between best and other leg only through the quality of performance).

Recently, two studies have reported differences similar to those found for AB2 of the MABC-2 Test in our study. For example, in a Japanese study, differences for the 'Placing Pegs' (MD1), 'Throwing Beanbag onto Mat' (AC2) and 'Hopping on Mats' (Dynamic BAL2) tasks were reported (Kita, Suzuki, Hirata, Sakihara, Inagaki, & Nakai, 2016) and a Czech study found differences for the 'Drawing Trail' item (MD3) and all BAL tasks in AB2, mainly in the 'One Board Balance' (Static BAL) item and the 'Hopping on Mats' (Dynamic BAL2) item (Psotta, Hendl, Frömel, & Lehnert, 2012). Although these studies are not directly comparable to our own since they used standard scores as opposed to the raw scores used here, they do support the idea that cultural differences are present in various parts of the world and that country specific norms are essential if the Test aims to yield reliable data in clinical settings.

5.3 Construct validity of the Test for AB1 and AB2

One of the main approaches to verify the validity of a test is to examine its factor structure (Yun, & Ulrich, 2002). Among the numerous ways possible, the most informative is to test the correspondence between the theoretical assumptions proposed by the test's authors, in this case Henderson et al. (2007), and the data obtained on a suitable population. Using the original data from the UK standardization of the MABC-2 Test, a CFA conducted by Schultz et al. (2011) demonstrated that the three-domain model proposed by Henderson and colleagues appeared to fit sufficiently at all three ABs. In addition, however, a general factor was obtained which was present between the ages of 3 and 10, but absent for the AB designed for children aged 11 to 16. In this study, the approach of Schultz et al. (2011) was followed as closely as possible for ABs 1 and 2 and a three-domain solution was confirmed for children aged 3 to 10. In contrast, however, a general factor clearly emerged at AB2, but was absent for AB1.

In addition to the present study, four other studies examined the factor structure of the MABC-2 Test but none have covered the entire age range of the Test. A Greek (Ellinoudis, Evaggelinou,

Konstantinidou, Venetsanou, & Kambas, 2011) and Chinese (Hua, Gu, Meng, & Wu, 2013) study focused on AB1, while a Japanese and a German study examined AB2. Although all four provide support for the three-domains solution originally proposed, there were differences, which require consideration. For example, at AB1 the small study of 183 Greek children by Ellinoudis et al. (2011) contrasted with the much larger study of 1823 Chinese children of the same age by Hua et al. (2013). While the first replicated the results of Schultz et al (2011), the second found that some modifications to the original model were necessary to obtain a stable solution. Specifically, removing the 'Drawing Trail' (MD3) item and the 'Walking with Heels Raised' (Dynamic BAL1) task yielded better results. In the present study, a better fit to the model investigated was also obtained when the MD3 task was removed. Considering the findings from the Chinese and the IT samples together, interestingly, the two tasks that are the most difficult to score at this age, emerge as separate from the other tasks since their lower reliability may affect the factor solution. More generally, however, the MD3 task, focused on accuracy rather than speed, often stands out from other manual tasks, and difficulties in graphic tasks often require careful evaluation with specific tests, such as the DASH test (Detailed Assessment for Speed in Handwriting, Barnett, Henderson Scheib, & Schulz, 2007), designed by the MABC-2 authors as a supplement, or the DGM-P test, devised for Italian school-aged children (Borean, Paciulli, Bravar, & Zoia, 2012).

At AB2 of the Test, both the present study and that of Kita et al. (2016) obtained a factorial structure similar to that of the UK data, with three specific factors and one general factor. However, in the BAL area, Kita et al. (2016) found that the static and dynamic components were more closely related than they were in the original UK study or in this study. Once again, the separation of these two types of tasks varies from study to study. Many textbooks on motor development deal with these separately, yet clinically we tend to think that children with poor static balance are also more likely to have problems in dynamic balance tasks.

In addition to the within domain differences, the present analysis reveals a different solution regarding a general factor. When interpreting their finding of a general factor at ABs 1 and 2 that disappears at AB3, Schultz et al. (2011) suggested that early motor development is less differentiated than it is later, with specificity emerging as the dominant feature with age. However, since data from Italy, Greece and China now show a different pattern for 3-6 year old children, perhaps culture specific factors lead to different developmental trajectories. Thus, children in these cultures could be driven to develop specific motor domains (MD, AC, BAL) rather than a general motor competence, because cultural-specific constraints may shape or model the acquisition of certain motor skills. In this sense, the identification of a general factor may vary between population-based samples and different age ranges because different educational experiences

interact with the organization of motor cognitive functions, especially at younger ages. Noticeably, there are some limitations in this study and further research is necessary, considering that this work only focused on children between the ages of 3 and 10. Given the differences in the factor structure reported, it will be important to perform the same analyses on older children between the ages of 11 through 16. Educational systems differ considerably from country to country as do students' interests and participation in sports or other motor activities. The normative data for the whole MABC-2 Test reported by Biancotto et al. (2017) does show differences between students at the secondary school level but the factor structure of the test for this population has not yet been tested. Alongside this research, which uses a cross sectional approach to data collection, future longitudinal studies are needed to examine the complex relationships between motor proficiency and its impact on the quality of life, from a developmental and across lifespan viewpoint, examining these relationships in both non-DCD and DCD samples.

6. Conclusions

Motor development used to be viewed as the "poor cousin" of other areas of child development, such as cognitive and language development. Recently, however, emphasis has changed and the ability to act physically on the world is now viewed as underpinning cognitive, language and social development (von Hofsten, 2013; Kenny, Hill, & Hamilton, 2016). In parallel, interest in children with atypical motor development has increased. The present study employs data collected during the IT standardization of the MABC-2 Test and provides two sources of validity data which explain the characteristics of the Test. A detailed comparison between IT and UK children revealed differences explicable in terms of educational, social and cultural experiences, highlighting the need for local norms. These, in turn, underpin the usefulness of determining country specific cut-off points for the identification of children with significant motor difficulties, as recently done for the IT population (Biancotto, Guicciardi, Pelamatti, Santamaria, & Zoia, 2017). Furthermore, the replication of the MABC-2 Test three-factors structure for the IT sample supports the view that the IT standardization norms, drawn from the same sample, could be used both for research and clinical purposes throughout Italy. On the other hand, the variable presence of a general motor factor between the two ABs underlines the advantages of considering not only the total Test score as a means of identifying significant motor problems (if ≤ 5° percentile; or a condition "at risk for motor problems" if between 6° and 16° percentile), but also the profile of a child's performance across the three domains (MD, AC and BAL). In particular, at 3-6 years (AB1), IT MABC-2 Test users should exert caution when interpreting the result of the MD sub-component, and in considering specific sub-component results in comparison with the total Test score.

In summary, the findings of this study support the view that the IT version of MABC-2 Test has an important role to play both in the diagnosis of DCD and in the identification of movement difficulties in children with other neurodevelopmental disorders. Its use in the diagnostic process can reduce the frequency of misdiagnosis of DCD and promote the identification of children with motor problems from the age of three years.

7. Acknowledgments

8. References

American Psychiatric Association (2013). DSM-5. Diagnostic and statistical manual of mental disorders. Washington, DC: American Psychiatric Association (Italian edition: DSM-5. Manuale diagnostico e statistico dei disturbi mentali, quinta edizione. 2014. Milano: Raffaello Cortina).

Barnett, A., Henderson S.E., Scheib, B., & Schulz, J. (2007). Detailed Assessment of Speed of Handwriting (DASH). Oxford: Pearson Assessment.

Blank, R., Smits-Engelsman, B., Polatajko, H., &Wilson, P. (2012). European Academy for Childhood Disability (EACD): Recommendations on the Definition, Diagnosis and Intervention of Developmental Coordination Disorder. Developmental Medicine and Child Neurology, 54 (1), 54-93.

Biancotto, M., Guicciardi, M., Pelamatti, G.M., Santamaria, T., & Zoia, S. (2017). Movement Assessment Battery for Children-Second Edition. Standardizzazione Italiana. Firenze: Giunti O.S. Psychometrics. ISBN: 978-88-09-99492-B.

Biancotto, M., Borean, M., Bravar, L., Pelamatti, M.G., & Zoia, S. (2013). Movement Assessment Battery for Children-Second Edition. Adattamento italiano (Manuale per l'Esaminatore e Manuale per l'Intervento Ecologico). Firenze: Giunti O.S. Psychometrics. ISBN: 978-88-09-77052-2.

Borean, M., Paciulli, G., Bravar, L., & Zoia, S. (2012). Test DGM-P: Test per la valutazione delle Difficoltà Grafo-Motorie e Posturali della Scrittura. Trento: Erickson. ISBN: 978-88-6137-974-9.

Cheng, C.H., Ju, Y.Y., Chang, H.W., Chen, C.L., Pei, Y.C., Tseng, K.C., & Cheng, H.Y.K. (2014). Motor impairments screened by the Movement Assessment Battery for Children-2 are related to the visual-perceptual deficits in children with Developmental Coordination Disorder. Research in Developmental Disabilities, 35, 2172–2179.

Ellinoudis, T., Evaggelinou, C., Konstantinidou, Z., Venetsanou, F., & Kambas, A. (2011). Reliability and validity of age band 1 of the Movement Assessment Battery for Children- Second

Edition. Research in Developmental Disabilities, 32, 1046–105.

Getchell, N., Pabreja, P., Neeld, K., & Carrio, V. (2007). Comparing children with and without dyslexia on the Movement Assessment Battery for Children and the Test of Gross Motor Development. Perceptual Motor Skills, 105 (1), 207-14.

Griffiths, A., Morgan, P., Anderson, P.J., Doyle, L.W., Lee, K.J., & Spittle, A.J. (2017). Predictive value of the Movement Assessment Battery for Children - Second Edition at 4 years, for motor impairment at 8 years, in children born preterm. Developmental Medicine and Child Neurology, 59, 490-496.

Hanaie, R., Mohri, I., Kagitani-Shimono, K., & Taniike, M. (2016). White matter volume in the brainstem and inferior parietal lobule is related to motor performance in children with autism spectrum disorder: A voxel-based morphometry study. Autism Research, 9 (9), 981–992.

Henderson, S.E., & Sugden, D.A. (1992). Movement Assessment Battery for Children: Manual. London: Psychological Corporation.

Henderson, S.E., Sugden, D.A., & Barnett, A.L. (2007). Movement Assessment Battery for Children-second edition. London: Pearson Psychological Corporation.

Hill, E.L., Pratt, M.L., Kanji, Z., & Jones Bartoli, A. (2017). Motor and coordination difficulties in children with emotional and behavioural difficulties. Emotional and Behavioural Difficulties. ISSN 1363-2752.

Holm, I., Tveter, A.T., Aulie, V.S., & Stuge, B., (2013). High intra- and inter-rater chance variation of the movement assessment battery for children 2, age band 2. Research in Developmental Disabilities 34, 795-800.

Howen, S., van der Veer, G., Visser, J., & Cantell, M. (2017). The relationship between motor performance and parent-rated executive functioning in 3- to 5-year-old children: What is the role of confounding variables? Human Movement Science, 53: 24-36. doi:10.1016/j.humov.2016.12.009.

Hua, J., Gu, G., Meng, W., & Wu, Z. (2013). Age band 1 of the movement assessment battery for children-second edition: exploring its usefulness in mainland China. Research in Developmental Disabilities, 34, 801-808.

Kenny, L., Hill, E., & de C. Hamilton, A.F. (2016). The Relationship between Social and Motor Cognition in Primary School Age-Children. Frontiers on Psychology, 7, 1-12. doi:10.3389/fpsyg.2016.00228

Kita, Y., Suzuki, K., Hirata, S., Sakihara, K., Inagaki, M., Nakai, A. (2016). Applicability of the Movement Assessment Battery for Children-Second Edition to Japanese children: A study of the Age Band 2. Brain & Development, 38 (8), 706-713.

Marquet-Doléac, J., Soppelsa, R., & Albaret, J.M. (2016). Batterie d'Evaluation du Mouvement chez l'Enfant – 2ème édition (MABC-2) – adaptation française. Montreuil: Editions du Centre de Psychologie Appliquée.

Niemeijer, A.S., van Waelvelde, H., & Smits-Engelsman, B.C.M. (2015). Crossing the North Sea seems to make DCD disappear: Cross-validation of Movement Assessment Battery for Children-2 norms. Human Movement Science, 39, 177-188.

Psotta, R., Hendl, J., Frömel, K., & Lehnert, M. (2012). The Second Version of the Movement Assessment Battery for Children: a comparative study in 7-10 year old children from the Czech Republic and the United Kingdom. Acta Universitatis Palackianae Olomucensis: Gymnica, 42 (4), 19-27.

Rarick, G.L., Dobbins, D.A., & Broadhead, G.D. (1976). The motor domain and its correlates in educationally handicapped children. Engelwood Cliffs, NJ: Prentice-Hall.

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. Methodology 6 (4), 147-151. doi:10.1027/1614-2241/a000016.

Schulz, J., Henderson, S.E., Sugden, D.A., & Barnett, A.L. (2011). Structural validity of the Movement ABC-2 test: Factor structure comparisons across three age groups. Research in Developmental Disabilities, 32 (4), 1361-1369.

Smits-Engelsman, B.C.M., Niemeijer, A.S., & van Waelvelde, H. (2011). Is the Movement Assessment Battery for Children-2nd edition a reliable instrument to measure motor performance in 3 year old children? Research in Developmental Disabilities, 32, 1370-1377.

Stott, D.H., Moyes, F.A., & Henderson, S.E. (1972). The test of motor impairment. San Antonio, TX: The Psychological Corporation.

Stott, D.H., Moyes, F.A., & Henderson, S.E. (1984). The test of motor impairment - Henderson revision. San Antonio, TX: The Psychological Corporation.

Sugden, D.A., & Henderson, S.E. (2007). Ecological Intervention for children with movement difficulties. London: Pearson Education (Italian version: M. Biancotto, M. Borean, L. Bravar, G.M. Pelamatti, & S. Zoia. Movement ABC-2. Un Intervento Ecologico per bambini con difficoltà di movimento, 2013. Firenze: Giunti O.S. Psychometrics).

Tanizaki, H. (1997). Power comparison of non-parametric tests: small-sample properties from Monte Carlo experiments. Journal of Applied Statistics, 24, 603-632.

Valentini, N.C., Ramalho, M.H., Oliveira, M.A. (2014). Movement Assessment Battery for Children-2: Translation, reliability, and validity for Brazilian children. Research in Developmental Disabilities, 35, 733-740.

van Waelvelde, H.V., Wim Peersman, W., & Smits-Engelsman, B. (2010). Factoranalytische validatie van de Movement ABC-2 test. Nederlands Tijdschrift Voor Kinderfysiotherapie, 22 (1), 1-19.

von Hofsten, C. (2013). Action in infancy: a foundation for cognitive development. Published in Wolfgang Prinz Miriam Beisert Arvid Herwig (eds.) Tutorials in Action Science, 255-279. MIT Press.

Wagner, M.O., Kastner, J., Petermann, F., & Bös, K. (2011). Factorial validity of the Movement Assessment Battery for Children-2 (age band 2). Research in Developmental Disabilities, 32, 674-680.

Webster, R.I., Erdos, C., Evans, K., Majnemer, A., Kehayia, E., Thordardottir, E., Evans, A., & Shevell, M.I. (2006). The Clinical Spectrum of Developmental Language Impairment in School-Aged Children: Language, Cognitive, and Motor Findings. Pediatrics, 118 (5), 1541-1549.

doi:10.1542/peds.2005-2761.

World Health Organisation (2007). ICF-CY. International classification of functioning, disability and health: Children and youth version. Genève: Organizzazione Mondiale della Sanità (Italian edition: ICF-CY. Classificazione internazionale del funzionamento, della disabilità e della salute – Versione per bambini e adolescenti, 2007. Trento: Erickson).

Yun, J., & Ulrich, D.A. (2002). Estimating measurement validity: A tutorial. Adapted Physical Activity Quarterly, 19 (1), 32-47.

Zimmerman, D.W. (1998). Invalidation of parametric and nonparametric statistical tests by concurrent violation of two assumptions. Journal of Experimental Education, 67, 55-68.