Which items of the movement assessment battery for children are most sensitive for identifying children with probable developmental coordination disorder? Results from a large-scale study

Peer-reviewed author version

VERBECQUE, Evi; Denysschen, Marisja; Coetzee, Dane; Valtr, Ludvik; Bonney, Emmanuel & Smits-Engelsman, Bouwien (2025) Which items of the movement assessment battery for children are most sensitive for identifying children with probable developmental coordination disorder? Results from a large-scale study. In: Research in developmental disabilities, 157 (Art N° 104904).

DOI: 10.1016/j.ridd.2024.104904
Handle: http://hdl.handle.net/1942/45340

Which items of the Movement Assessment Battery for Children are most sensitive for identifying children with probable developmental coordination disorder? Results from a large-scale study.

Evi Verbecque[1*], Marisja Denysschen[2], Dané Coetzee[2], Ludvik Valtr[3], Emmanuel Bonney[4,5], Bouwien Smits-Engelsman[2,6]

[1]Rehabilitation Research Centre - REVAL, Faculty of Rehabilitation Sciences and Physiotherapy, Hasselt University, Hasselt, Belgium

[2] Department of Physical Activity, Sport and Recreation, Faculty Health Sciences, North-West University, Potchefstroom, South Africa

[3] Faculty of Physical Culture, Palacký University, Olomouc, Czechia

[4] School of Kinesiology, University of Minnesota, Minneapolis, Minnesota, USA.

[5]Masonic Institute for the Developing Brain, University of Minnesota, Minneapolis, Minnesota, USA.

[6] Department of Health and Rehabilitation Sciences, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

* corresponding author: Evi Verbecque, evi.verbecque@uhasselt.be, Wetenschapspark 7, 3590 Diepenbeek, Belgium.

Funding sources: The Czech data collection was supported by the Czech Science Foundation [GAČR 21-15728X]. The UK, Dutch and Flemish data collection was supported by Pearson.

Conflict of interest: The authors declare no conflict of interest. The funders had no role in the analyses, interpretation of data; in the writing of the manuscript; nor in the decision to publish the results.

Ethics committee approval: Anonymized data from previous studies were merged. For each of the original studies ethical clearance was gained from the local ethical committees. The data used in this study are anonymized and therefore do no longer require ethical approval, as the data cannot be linked to any of the participants.

Abstract

Introduction: Despite the widespread use of the Movement Assessment Battery for Children, 2nd edition (MABC-2), little is known about the sensitivity or specificity of the individual items to detect probable Developmental Coordination Disorder (p-DCD). This study examined which specific MABC-2 items were most sensitive to identify children with p-DCD and which items would predict p-DCD.

Methods: Based on a large dataset including European and African children aged 3-16 years (n=4916, typically developing (TD, 49.6% boys); n=822 p-DCD (53.1% boys), Hedges' g was calculated to establish the standardized mean difference (SMD) between p-DCD/TD. SMDs were considered substantial when absolute values at or above 1.4. Sensitivity and specificity of the raw MABC-2 item scores predicting p-DCD/TD per age band (AB) were established with logistic regression analysis.

Results: AB1: Children with p-DCD performed substantially poorer on threading beads (SMD: -1.61) and jumping on mats (SMD: 1.61). By combining all items and the country of origin, the sensitivity was 61.7% and specificity 98.6%. AB2: Walking heel-to-toe forwards (SMD: 1.65) was substantially poorer in p-DCD. By combining all items and the country of origin, the sensitivity was 79.0% and specificity 97.6%. AB3: Catching a ball with the preferred (SMD: 1.8) or non-preferred (SMD: 1.61) hand, and for walking heel-to-toe backwards (SMD: 1.78) were substantially poorer in p-DCD. All items combined resulted in a sensitivity of 94.4% and specificity of 99.6%.

Conclusion: Not all MABC-2 items are equally sensitive to distinguish between performances of p-DCD and TD. Despite the good specificity, the sensitivity was only moderate in AB1-2, the age at which children learn culturally influenced motor skills.

Keywords: Movement Assessment Battery for Children, Developmental Coordination Disorder, Motor competence, Sensitivity

1. Introduction

Developmental Coordination Disorder (DCD) occurs in 5-6% of the school-aged children and significantly impacts their ability to learn and perform motor skills (Association., 2022). Movement difficulties in affected children are often manifested as clumsiness (e.g. dropping or bumping into objects), slowness, and inaccuracies in skill performance (e.g. catching an object, using scissors, or participating in sports) (Association., 2022). The 5th edition of the Diagnostic and Statistical Manual of Mental Disorders text revision (DSM-5-TR) provides diagnostic guidelines for identifying DCD in clinical and research settings (**Appendix 1**).

In 2019, Blank et al. (Blank, et al., 2019) published the International clinical practice recommendations on the definition, diagnosis, assessment, intervention, and psychosocial aspects of DCD. Specifically for criterion A, the DCD guideline recommends an appropriate motor test that measures different areas of motor competence, has good psychometric properties, and has population-based standardization (Blank, et al., 2019): the Movement Assessment Battery for Children, second edition (MABC-2) or the Bruininks Oseretsky Test of Motor Performance, second edition (BOT-2) (Blank, et al., 2019). The frequency of using the MABC in publications to confirm criterion A in DCD increased from 50% in 2001 (Geuze, Jongmans, Schoemaker, & Smits-Engelsman, 2001) to 73% in 2015 (B. Smits-Engelsman, Schoemaker, Delabastita, Hoskens, & Geuze, 2015) and to 87% in 2024 (B. Smits-Engelsman, et al., 2024). Although the MABC-2 is not agreed upon as the gold standard for verifying criterion A (Venetsanou, et al., 2011), it unofficially is (Eddy, et al., 2024).

The MABC-2 measures motor skill performance across three motor domains (manual dexterity (MD), aiming and catching (AC), and balance (BAL)) using eight age-specific standardized tasks (**Table 1**). The raw item scores are converted to item standard scores (ISS, range: 1-19), and then summed to yield domain scores. The domain scores can be converted to percentile rank scores for interpretation at domain level or domain standard scores, which in turn can be summed to determine the total score. The total score is interpreted using a traffic light classification system: the green zone (scores > percentile 16 (P16)) implies 'typical motor development', the orange zone (P5<scores≤P16) 'being at risk of motor impairment' and the red zone (scores≤P5) 'severe motor impairment'. Particularly scores in the red zone are often used to label children as probable DCD (p-DCD) if a diagnosis has not been established yet (Geuze, et al., 2001). Although the MABC-2 is designed to discriminate between performances at the lower end of the scale, usually, the total scores and the domain scores are reported, whereas item scores are seldomly documented (Hill, Bishop, & Nimmo-Smith, 1998; B. Smits-

Engelsman, Coetzee, Valtr, & Verbecque, 2023; B. Smits-Engelsman, Verbecque, Denysschen, & Coetzee, 2022). A total score in the red zone also indicates poor performance on the items. However, investigations into the most difficult items for children with DCD are very scarce, despite their potential to build specific profiles in these children.

Profiling or subtyping children with DCD gained much interest over the last two decades, because of the high heterogeneity in the symptom presentation (Aertssen, Bonney, Ferguson, & Smits-Engelsman, 2018; Asonitou & Koutsouki, 2016; Gras, et al., 2023; Green, Chambers, & Sugden, 2008; Lust, et al., 2022; Macnab, Miller, & Polatajko, 2001; Pieters, Roeyers, Rosseel, Van Waelvelde, & Desoete, 2015; Vaivre-Douret, Lalanne, Cabrol, Golse, & Falissard, 2012; Vaivre-Douret, Lalanne, Cabrol, et al., 2011; Vaivre-Douret, Lalanne, Ingster-Moati, et al., 2011; Van Dyck, Baijot, Aeby, De Tiège, & Deconinck, 2022; Visser, 2003). Interestingly, the subtypes differ in each study and depend upon the outcomes used in the model. In a recent meta-analysis (B. Smits-Engelsman, et al., 2024), we attempted to identify key features in children with DCD, instead the large heterogeneity within 'the DCD group' was reconfirmed. The magnitude of differences between DCD children and their typically developing (TD) peers fluctuates across domains of functioning, such as executive functioning, academic skills, quality of life, self-perceived competence, behaviors, sensory functions, physical fitness and activity, and motor skills (B. Smits-Engelsman, et al., 2024). Although in essence, DCD is a motor skill disorder, the individual child's motor profile strongly varies (B. Smits-Engelsman, et al., 2024). Therefore, at a group level, the overall standardized mean difference (SMD) between children with DCD and age-matched controls approximated one standard deviation but with substantial statistical heterogeneity among results, that could not be resolved through subclassification in distinct motor areas (B. Smits-Engelsman, et al., 2024). A major shortcoming of any meta-analysis concerns the differences in study samples and methodologies. The extent to which the phenotypic heterogeneity in DCD reflects variability in sample characteristics or assessment tools therefore remains unclear. To better understand which motor tasks are most difficult for children with DCD and are therefore expected to be most sensitive to detect a child with DCD, a comparison of similar assessment items ought to be used across samples.

Because of its popularity in both clinical practice and research, most children suspected of DCD will be tested with the MABC-2 and are therefore selected based on their poor performance on only eight activities (per age band, AB) which allows us to answer the question which are the most difficult items for children with DCD. The selection of the MABC-2 items arose from knowledge of motor learning, which has shown over the past decades that exposure and experience—both culturally dependent—affect motor performance, and consequently, the level, quality, or competency of that performance. To mitigate the impact of these confounding factors, the developers of the MABC-2 selected fundamental motor skills that are expected to develop universally in children, regardless of cultural background. As a result, the MABC-2 items are easily performed by typically developing children but remain challenging for those who struggle with everyday motor tasks. Exploring children's MABC-2 item performance helps elucidate the most sensitive items, and therefore identify key activities that are most difficult for children with DCD.

To fully understand the extent to which children with p-DCD are performing poorer than their TD peers, their raw scores should be compared. As the MABC-2 uses ISS, it assumes a normal distribution, but this assumption is not data-driven (B. Smits-Engelsman, et al., 2023), which may distort the magnitude of the differences between children with p-DCD/TD. To address some of these challenges, we analyzed a large MABC-2 dataset, including European and African samples, to answer the following questions:

1. Which raw item scores of the MABC-2 discriminate most between children with p-DCD and their TD peers?
2. How sensitive and specific are the raw item scores in children with p-DCD and TD?
3. To what extent do the raw item scores predict the group assignment (p-DCD versus TD)?

We hypothesize that there are specific MABC-2 items that will reveal substantial differences between p-DCD and TD (i.e. SMD > 1.4 indicating percentile grade of 91.9% and an overlap between both groups of 31.9% (Alwahaibi, Alhidabi, & Alkharusi, 2020), that the prevalence of poor performance in the p-DCD group will be high for these specific items (>33%) and low for the TD group (<5%), suggesting these items to be general characteristics of the p-DCD group that predict the total MABC-2 score (p-DCD (≤P5) versus TD (>P16)). The existence of such items may simplify the diagnostic process for those affected and their families (Steenbergen, et al., 2024).

2. Methods

2.1. Participants

In this retrospective study, we performed a secondary analysis on anonymized data collected across multiple projects for which random samples of children were tested on the MABC-2 in different countries including the United Kingdom (n=380, age 3-6), the Netherlands (n=2547, age 3-16), Flanders (n=923, age 3-16), Czech Republic (n=504, age 6-11), South Africa (n=971, age 6-16), Ghana (n=302, age 7-16) and Nigeria (n=111, age 7-10) (Abdollahipour, et al., 2023; Niemeijer, van Waelvelde, & Smits-Engelsman, 2015; B. C. Smits-Engelsman, Niemeijer, & van Waelvelde, 2011; Verbecque, Coetzee, Ferguson, & Smits-Engelsman, 2021). The entire sample included 6720 children consisting of 3364 boys (50.1%) and 3356 girls (49.9%). The age of the participants ranged between 3-16 years and was distributed over the 3 ABs as follows: 32.4% (AB1), 43.0% (AB2) and 24.6% (AB3). The focus of this study was on the children with p-DCD (MABC-2 total score ≤P5, i.e. p-DCD) and the children with typical motor development (MABC-2 total score >P16) (Cairney, et al., 2005). The total scores of 982 children were in the 'at risk group' and therefore excluded.

2.2. MABC-2

The MABC-2 measures motor performance in children (age 3-16) and is divided into three ABs (AB1: 3–6 years, AB2: 7–10 years, AB3: 11–16 years) (Henderson, 2007). The items are different for each AB (**Table 1**). Despite this different item set, the items are intended to measure a comparable construct. Each item tested provides a raw score which is converted into an ISS (Henderson, 2007), with ISS ≤5 interpreted as the child having definite or severe movement difficulties. The ISS are summed to calculate the total standard scores (SS), which can also be converted into percentiles. We used the Dutch normative data in the entire sample as this constitutes the largest norm sample (n=3230) (B. Smits-Engelsman, 2010). Since many countries do not have norms, they apply one of the available norm samples. We chose to apply one normative dataset to make all data comparable.

2.3. Data and statistical analysis

All statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS 29.0). Descriptive statistics were used to characterize the sample and the distribution was reported per country (number of children and percentage) as well as the sexes across ABs. The magnitude of the SMD was estimated using Hedges' g (Marfo & Okyere, 2019), interpreted as substantial when absolute values at or above 1.4 (Alwahaibi, et al., 2020). The percentage of p-DCD/TD children obtaining an ISS≤5 was tabulated. At least 1 in 3 p-DCD children had to have ISS≤5 for it to be a "General characteristic" (≥33%) of p-DCD, and ≤5% of the TD group were expected to have ISS≤5.

The sensitivity and specificity of the raw item scores of the MABC-2 were determined with backward binary logistic regression analysis to predict p-DCD/TD (Portney, 2009). Due to the different countries included in the sample, the variable 'country' was added to the model. The data met the assumptions of independent observations and the linear relationships between the continuous predictors and the logit outcomes. The model was statistically significantly better than the baseline at explaining the MABC-2 outcome (p-DCD versus TD; omnibus test < 0.05) and the Hosmer and Lemeshow's test indicated the goodness of fit ($p>0.05$). Sensitivity and specificity were extracted from the expected and predicted p-DCD and TD group assignment. Odds ratios (Exp (B)) were extracted when applicable to provide insights into the relative contribution of each country in the group assignment compared to the Dutch sample, which was used for normative scores.

3. Results

3.1. Sample description

A total of 5738 children fulfilled the selection criteria and were included in the analysis. Of these children, 4916 children (85.7%) were identified as typically developing (49.6% boys) and 822 children (14.3%) were classified as p-DCD (52.9% boys). The overall prevalence of p-DCD was 14.3% and differed across countries (**Figure 1**). More boys were present in the p-DCD subgroups, except for AB3.

3.2. Age band 1

The children with p-DCD needed substantially more time to 'thread beads' (SMD: -1.71) and performed fewer correct jumps onto mats (SMD: 1.61) than their TD peers (**Figure 2**). The prevalence of the ISS≤5 varied between 15.5% and 50% for the p-DCD group and between 1.1% and 4.7% for the TD group (**Figure 3**). All items, except for 'Drawing trail 1', 'Throwing a bean bag onto the mat' and 'One-leg balance, non-preferred leg', exceeded the 33% threshold and were considered general characteristics (**Figure 3**). By combining all items together with the variable 'country', the sensitivity was 61.7% and the specificity 98.6% (**Figure 4, Appendix 2**). The odds of belonging to the p-DCD group, compared to the Dutch children, were 0.408 for Flemish children (p=0.013), 11.78 for South African children (p<0.001), 0.92 for Czech children (p=0.939), and 1.16 for children from the UK (p=0.611).

3.3. Age band 2

Children with p-DCD scored substantially poorer on ‘walking heel-to-toe forward’ (SMD: 1.65) compared with TD peers (**Figure 2**). The prevalence of the ISS $\leq$ 5 varied between 17.7% and 51.6% for the p-DCD group and between 2.4% and 9.7% for the TD group (**Figure 3**). All items, except for ‘placing pegs with the preferred hand’, ‘one-board balance with the non-preferred leg’ and ‘hopping on mats with the non-preferred leg’, were considered general characteristics of p-DCD. In the TD group, performances on ‘placing pegs with the non-preferred hand’, ‘threading lace’, ‘Drawing trail 2’, ‘catching a ball’, ‘throwing bean bag onto mat’ and ‘one-board balance with the preferred leg’ exceeded the 5% threshold (**Figure 3**). All items combined together with the variable ‘country’ provides a sensitivity of 79.0% and a specificity of 97.6% to identify children with p-DCD and TD, respectively (**Figure 4, Appendix 2**). The odds of belonging to the p-DCD group, compared to the Dutch children, were 2.45 for Flemish children (p=0.016), 7.22 for South African children (p<0.001), 16.99 for Ghanaian children (p<0.001), 1.73 for Czech children (p=0.183) and 4.28 for Nigerian children (p=0.005).

3.4. Age band 3

Throwing a ball with the preferred (SMD: 1.80) and non-preferred hand (1.61), tandem stance (1.45), and walking heel-to-toe backwards (SMD: 1.56) were significantly poorer in children with p-DCD compared to their TD peers (**Figure 2**). The following items can be considered general characteristics in p-DCD (**Figure 3**): ‘triangle with nuts and bolts’, ‘throwing and catching a ball with one hand’, ‘two-board balance’ and ‘walking heel-to-toe backwards’. For those same items, the TD group exceeded the 5% threshold. The variable country did not significantly contribute to the prediction model. All items combined result in a sensitivity of 94.4% and a specificity of 99.6% to predict p-DCD and TD (**Figure 4, Appendix 2**).

4. Discussion

This study primarily investigated which raw item scores of the MABC-2 discriminate most between children with p-DCD and their TD peers. In addition, the sensitivity and specificity of the raw item scores were explored in both groups and the extent to which the raw item scores predicted group assignment (p-DCD versus TD). No key MABC-2 items were identified in any of the age bands. Only a few items revealed substantial differences between both groups and the type of items was different for each AB (**Figure 2**). Contrarily, the majority of the items exceeded the 33% threshold in the p-DCD group and could therefore be considered a general characteristic in the p-DCD group for AB1 (6/8 items) and AB2 (7/8 items) whereas this was only the case for half of the items in AB3 (**Figure 3**). Notwithstanding, the prevalence of detected motor difficulties in the TD group exceeded the 5% threshold in AB2 on 6 items and in AB3 on 5 items (**Figure 3**). All items combined provided the most sensitive and specific prediction of group assignment in each AB. The children's origin (i.e. their 'country') and the item performances together predicted group assignment in AB1 and AB2.

Literature reports inconsistencies in the core problems in DCD (Smits-Engelsman, et al., 2024), which may be due to variability in the type and complexity of the motor task as well as the measures used in experimental and field-based studies. However, we find similar diverging results in the different ABs despite this homogeneous set of standardized items used in all these children. Perhaps the age-specific character of the items also masks trends across ABs. In AB1 and AB2, almost all MABC-2 items can be considered general characteristics for p-DCD. Yet, in AB1 substantial differences were found for threading beads and jumping onto mats, whereas this was only the case for walking heel-to-toe forwards in AB2. The (pre)adolescents (AB3) with p-DCD mainly struggle with ball skills, two-board balance and hopping sideways onto mats. A major advantage of using age-specific criteria entails age-appropriate assessment, but with the disadvantage that seemingly similar tasks may not be measuring the same underlying construct. For example, MD2 (**Table 2**), threading beads in AB1, requires fine bimanual motor precision, eye-hand coordination and little motor planning skills, which becomes threading lace in AB2, requiring more preferred hand precision, eye-hand coordination and more motor planning (inserting the lace making it go back and forth without going around the board) and in AB3, (pre)adolescents need to construct a triangle, which mainly taps into insight in the construction (and screw turning movement) and preferred hand precision. Despite the clear link between the content of these items, the underlying systems required to perform these tasks, differ.

Notwithstanding the impact of age, another critical aspect of content validity is task-related sensitivity and specificity. Not all items are equally sensitive to pick up substantially different performances between p-DCD and TD children, and in their raw form, they are particularly useful to predict TD children (high specificity across all ABs), even though in AB2 and AB3 the prevalence of severe motor impairment at item level exceeded the threshold for many items. The sensitivity was only moderate in AB1 and AB2, the age at which children start learning culturally influenced motor skills (Flôres, Rodrigues, Copetti, Lopes, & Cordovil, 2019). We previously challenged the content validity of the MABC-2, since the actual prevalence of motor difficulties was much larger in African samples than expected and that systematic sex-related differences in performances exist, depending upon the sample's cultural background (B. Smits-Engelsman, et al., 2023; B. Smits-Engelsman, et al., 2022). Our results underscore the importance of culturally sensitive motor skills to identify p-DCD in (pre)school children, as the child's country of origin was a significant predictor for p-DCD in AB1 and AB2 and the odds of belonging to the p-DCD group increased significantly in African samples in both ABs. Yet, the combination of all items in AB3 resulted in very high sensitivity and specificity, suggesting (pre)adolescents should be able to perform these tasks, regardless their cultural background. Furthermore, it is striking to see that a much larger portion of boys is present in the p-DCD groups in AB1/AB2 compared to AB3. The question remains whether the prevalence of p-DCD in boys which decreases with increasing age, is caused by the content of the test items – also more difficult for TD boys compared to girls (B. Smits-Engelsman, et al., 2023) – or results from a slower development in boys that they may catch up in AB3.

Studies exploring performances at item level mainly applied factor analysis, to confirm the structure of the test's subscales (MD, AC, BAL) (Ellinoudis, et al., 2011; Hua, Gu, Meng, & Wu, 2013; Psotta & Abdollahipour, 2017; Serbetar, Loftesnes, & Mamen, 2019; Wagner, Kastner, Petermann, & Bös, 2011; Zoia, et al., 2019). Raw item scores for both groups, or SMDs at item level are hardly ever reported (Hill, et al., 1998; B. Smits-Engelsman, et al., 2023; B. Smits-Engelsman, et al., 2022). Hill and colleagues (Hill, et al., 1998) reported that clinically referred children with DCD needed much more time to place the pegs (item 1) than their controls (Hedges' g = -2.35, (non-)preferred hands collapsed) compared to our test results (Hedges' g = -1.1), which may be explained by the source population. Regardless their sparsity, particularly the raw item scores are crucial. By converting the raw scores to ISS, the normative data are smoothened and artificially handled as normally distributed data (mean ISS is set at 10 (percentile 50) and the SD at 3, creating a cut-point at ISS 7 (percentile 16) and ISS 5 (percentile 5)). Yet, the raw scores are far from normally distributed (French, Sycamore, McGlashan, Blanchard, & Holmes, 2018, 2021; B. Smits-Engelsman, et al., 2023) and seem to be less powerful to detect substantial differences between p-DCD and TD, but also in predicting group assignment, especially in younger children. The ISS exceeds the 5% cut-off that is inherent to a normal distribution in the TD group and although a total score ≤P5 (p-DCD group assignment) indicates that the majority of the item performances should also be in the severe motor impairment range (ISS ≤ 5), the 33% cut-off was not reached in a large number of items and reached approximately 60% at most. This underlines the heterogeneity within DCD. Even within one tool used worldwide, the probability of children performing at the lower end of the scale at item level is very variable. Our recent meta-analyses revealed that many studies show no significant differences between children with DCD and TD (B. Smits-Engelsman, et al., 2024).

We fulfilled criterion A with the MABC-2, however, this diagnostic criterion comprises more than just the execution of coordinated motor skills below what can be expected based on the individual's chronological age. It also encompasses the acquisition of coordinated motor skills and the opportunity for skill learning and use (Association., 2022). These last elements are not assessed with the MABC-2 (van der Linde, et al., 2015). Difficulties experienced by children with DCD are often described as clumsiness (e.g., dropping or bumping into objects) as well as slowness and inaccuracies in performance of motor skills (e.g., catching an object, using scissors or cutlery or participating in sports). Most of these activities are not tested using the MABC-2. For AB3, this does not seem to matter, whereas for the younger children the choice of the tasks may interfere with the diagnostic accuracy of identifying a child with DCD. However, task-specificity seems crucial in AB1 and AB2. The MABC-2 assesses activities outside a natural context; for example, throwing a bean bag onto a target on a mat on the floor or catching a bean bag in a standing position is not similar to learning or performing a ball skill. For younger ages, ball and locomotor skills which require more precision are lacking, as well as skill-related fitness needed for participation in leisure and sports (King-Dowling, et al., 2019).

4.1. Study limitations

Since many countries do not have MABC-2 norms, they apply one of the available norm samples. We therefore determined all ISS and total scores using the Dutch normative data, which may have caused some over- or underestimations of the performances. However, by using the same normative data, all children, regardless their background are compared to the same reference. Furthermore, country was only a significant predictor for AB1 and AB2 and higher odds of p-DCD were mainly seen in African subsamples, confirming the cross-continental differences.

In a clinical setting, children with limitations in their (physical) daily activities will most likely also perform poorly on the MABC-2. Nevertheless, the children included in this sample as p-DCD do not fulfill all criteria, thereby potentially influencing the sensitivity of raw items. Furthermore, in our sample, AB1 consisted of children aged 3 to 6 years old. The international guideline for DCD (Blank, et al., 2019) recommends the confirmation of criterion A in young children using more than one motor assessment with at least three months in between, which was not done here. These recommendations stem from a delayed motor development in young children with a spontaneous catch-up (late bloomers) and the variable cooperation and motivation of young children for motor assessments. Our Czech and South African subsamples in AB1 consisted of only 6-year-olds, whereas all ages were covered in the other countries. Appendix 2 provides further evidence for the challenges we face with accurately identifying motor difficulties in very young children.

We present a secondary analysis on merged data sets, and specific information and as such some background information, e.g. the socio-economic status of the children, is lacking, but may have provided additional insights.

4.2. Recommendations for clinical practice and future research

Despite its very frequent use in both clinical practice and research, the MABC-2 needs to be applied with some caution. Users should be aware that the ISS artificially create a large difference between TD and p-DCD at item level, causing children to be labelled as DCD, although their raw item scores do not support this, especially in very young children. The emphasis in DCD research gradually shifted to early identification facilitating early treatment. Even though a low total MABC-2 score may be used to stimulate motor skills and recreational sports in preschoolers, MABC-2 raw item scores do not support the identification of p-DCD at a young age. This warrants for a different test approach for criterion A in preschoolers and maybe even in the primary school-age children. Since no key items were indicative of DCD, it may even be questioned whether a test for DCD exists. This matches with modern motor control theories of task specificity and intervention recommendations for task-oriented training (Smits-Engelsman, et al., 2018). Individualized approaches based on the request for help, considering the individual’s task context by collecting information about the child, task, and environment during the anamnesis and performance analysis are therefore needed. Finally, this large dataset should be utilized to gain a deeper understanding of the structural validity of the MABC-2. Previous studies using raw scores (Psotta & Abdollahipour, 2017; Serbetar, et al., 2019; Zoia, et al., 2019) or z-scored values based on the sample’s own mean and standard deviation

(Nazario, Ferreira, Both, & Vieira, 2021) highlight that, although the MABC-2 was developed with the assumption that its items are culturally neutral, the theoretical three-domain structure is often not confirmed in other populations (e.g., Italian, Czech, Croatian and Brazilian children). Previous confirmatory factor analyses have already demonstrated that different structures emerge in various cultural contexts, particularly for AB1 (Hirata, et al., 2018; Hua, et al., 2013; Kokštejn, Musálek, & Tufano, 2018; Nazario, et al., 2021; Okuda, Pangelinan, Capellini, & Cogo-Moreira, 2019; Serbetar, et al., 2019). Some evidence also suggests that item 3 (drawing trail) does not fit the model (Nazario, et al., 2021; Zoia, et al., 2019). For older children, the hypothesized structure seems to hold better, although the loadings of some items are weak (Zoia, et al., 2019), indirectly implicating the structure. Therefore, previous research has already indicated that the MABC-2 structure varies depending on the study population. Conducting a multi-group confirmatory factor analysis on this large dataset could further clarify the necessary adjustments based on cultural context and aid in the development of future MABC versions and the way it is used.

5. Conclusion

Despite its popularity in clinical and research settings to identify DCD, not all MABC-2 items are equally sensitive to pick up different performances in p-DCD and TD children, and in their raw form, they are particularly useful to identify TD children (i.e. high specificity). Their sensitivity was only moderate in AB1/AB2, the age at which children start learning culturally influenced motor skills, highlighting the urgent need for culturally sensitive motor assessments for DCD detection across diverse cultures.

6. References

Abdollahipour, R., Valtr, L., Banátová, K., Bizovská, L., Klein, T., Svoboda, Z., Steenbergen, B., & Wilson, P. H. (2023). Motor-cognitive coupling is impaired in children with mild or severe forms of developmental coordination disorder. *Front Hum Neurosci, 17*, 1252852.

Aertssen, W., Bonney, E., Ferguson, G., & Smits-Engelsman, B. (2018). Subtyping children with developmental coordination disorder based on physical fitness outcomes. *Hum Mov Sci, 60*, 87-97.

Alwahaibi, I., Alhidabi, D., & Alkharusi, H. (2020). Cohen’s criteria for interpreting practical significance indicators: A critical study. *Cypriot Journal of Educational Sciences, 15*, 246-258.

Asonitou, K., & Koutsouki, D. (2016). Cognitive process-based subtypes of developmental coordination disorder (DCD). *Hum Mov Sci, 47*, 121-134.

Association., A. P. (2022). Diagnostic and statistical manual of mental disorders, text revision (5th ed.).

Blank, R., Barnett, A. L., Cairney, J., Green, D., Kirby, A., Polatajko, H., Rosenblum, S., Smits-Engelsman, B., Sugden, D., Wilson, P., & Vinçon, S. (2019). International clinical practice recommendations on the definition, diagnosis, assessment, intervention, and psychosocial aspects of developmental coordination disorder. *Dev Med Child Neurol, 61*, 242-285.

Cairney, J., Hay, J. A., Faught, B. E., Wade, T. J., Corna, L., & Flouris, A. (2005). Developmental coordination disorder, generalized self-efficacy toward physical activity, and participation in organized and free play activities. *J Pediatr, 147*, 515-520.

Eddy, L. H., Preston, N., Boom, S., Davison, J., Brooks, R., Bingham, D. D., Mon-Williams, M., & Hill, L. J. B. (2024). The validity and reliability of school-based fundamental movement skills screening to identify children with motor difficulties. *PLoS One, 19*, e0297412.

Ellinoudis, T., Evaggelinou, C., Kourtessis, T., Konstantinidou, Z., Venetsanou, F., & Kambas, A. (2011). Reliability and validity of age band 1 of the Movement Assessment Battery for Children--second edition. *Res Dev Disabil, 32*, 1046-1051.

Flôres, F. S., Rodrigues, L. P., Copetti, F., Lopes, F., & Cordovil, R. (2019). Affordances for Motor Skill Development in Home, School, and Sport Environments: A Narrative Review. *Percept Mot Skills, 126*, 366-388.

French, B., Sycamore, N. J., McGlashan, H. L., Blanchard, C. C. V., & Holmes, N. P. (2018). Ceiling effects in the Movement Assessment Battery for Children-2 (MABC-2) suggest that non-parametric scoring methods are required. *PLoS One, 13*, e0198426.

French, B., Sycamore, N. J., McGlashan, H. L., Blanchard, C. C. V., & Holmes, N. P. (2021). Correction: Ceiling effects in the Movement Assessment Battery for Children-2 (MABC-2) suggest that non-parametric scoring methods are required. *PLoS One, 16*, e0245674.

Geuze, R. H., Jongmans, M. J., Schoemaker, M. M., & Smits-Engelsman, B. C. (2001). Clinical and research diagnostic criteria for developmental coordination disorder: a review and discussion. *Hum Mov Sci, 20*, 7-47.

Gras, D., Ploix Maes, E., Doulazmi, M., Huron, C., Galléa, C., Boespflug Tanguy, O., Germanaud, D., & Roze, E. (2023). Developmental coordination disorder subtypes in children: An unsupervised clustering. *Dev Med Child Neurol, 65*, 1332-1342.

Green, D., Chambers, M. E., & Sugden, D. A. (2008). Does subtype of developmental coordination disorder count: is there a differential effect on outcome following intervention? *Hum Mov Sci, 27*, 363-382.

Henderson, S. E. S., D.A.; Barnett, A.L. . (2007). *Movement Assessment Battery for Children-2 (MABC-2), 2nd ed.* London, UK: The Psychological Corporation.

Hill, E. L., Bishop, D. V. M., & Nimmo-Smith, I. (1998). Representational gestures in Developmental Coordination Disorder and specific language impairment: Error-types and the reliability of ratings. *Human Movement Science, 17*, 655-678.

Hirata, S., Kita, Y., Yasunaga, M., Suzuki, K., Okumura, Y., Okuzumi, H., Hosobuchi, T., Kokubun, M., Inagaki, M., & Nakai, A. (2018). Applicability of the Movement Assessment Battery for Children-Second Edition (MABC-2) for Japanese Children Aged 3-6 Years: A Preliminary Investigation Emphasizing Internal Consistency and Factorial Validity. *Front Psychol, 9*, 1452.

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. *Res Dev Disabil, 34*, 801-808.

King-Dowling, S., Kwan, M. Y. W., Rodriguez, C., Missiuna, C., Timmons, B. W., & Cairney, J. (2019). Physical activity in young children at risk for developmental coordination disorder. *Dev Med Child Neurol, 61*, 1302-1308.

Kokštejn, J., Musálek, M., & Tufano, J. J. (2018). Construct Validity of the Movement Assessment Battery for Children-Second Edition Test in Preschool Children with Respect to Age and Gender. *Front Pediatr, 6*, 12.

Lust, J. M., Steenbergen, B., Diepstraten, J., Wilson, P. H., Schoemaker, M. M., & Poelma, M. J. (2022). The subtypes of developmental coordination disorder. *Dev Med Child Neurol*.

Macnab, J. J., Miller, L. T., & Polatajko, H. J. (2001). The search for subtypes of DCD: is cluster analysis the answer? *Hum Mov Sci, 20*, 49-72.

Marfo, P., & Okyere, G. A. (2019). The accuracy of effect-size estimates under normals and contaminated normals in meta-analysis. *Heliyon, 5*, e01838.

Nazario, P. F., Ferreira, L., Both, J., & Vieira, J. L. L. (2021). MOVEMENT ASSESSMENT BATTERY FOR CHILDREN-SECOND EDITION: THEORETICAL ADEQUACY OF A MOTOR ASSESSMENT INSTRUMENT. *Rev Paul Pediatr, 40*, e2020205.

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

Okuda, P. M. M., Pangelinan, M., Capellini, S. A., & Cogo-Moreira, H. (2019). Motor skills assessments: support for a general motor factor for the Movement Assessment Battery for Children-2 and the Bruininks-Oseretsky Test of Motor Proficiency-2. *Trends Psychiatry Psychother, 41*, 51-59.

Pieters, S., Roeyers, H., Rosseel, Y., Van Waelvelde, H., & Desoete, A. (2015). Identifying Subtypes Among Children With Developmental Coordination Disorder and Mathematical Learning Disabilities, Using Model-Based Clustering. *Journal of Learning Disabilities, 48*, 83-95.

Portney, L. G., Watkins, MP. (2009). *Foundations of Clinical Research Applications to Practice* (Vol. 3rd ed.). Upper Saddle River, New Jersey: Pearson Prentice Hall.

Psotta, R., & Abdollahipour, R. (2017). Factorial Validity of the Movement Assessment Battery for Children-2nd Edition (MABC-2) in 7-16-Year-Olds. *Percept Mot Skills, 124*, 1051-1068.

Serbetar, I., Loftesnes, J. M., & Mamen, A. (2019). Reliability and Structural Validity of the Movement Assessment Battery for Children-2 in Croatian Preschool Children. *Sports (Basel), 7*.

Smits-Engelsman, B. (2010). *Movement ABC-2-NL Dutch Manual, 2nd ed*. Amsterdam, The Netherlands: Pearson Assessment and Information.

Smits-Engelsman, B., Coetzee, D., Valtr, L., & Verbecque, E. (2023). Do Girls Have an Advantage Compared to Boys When Their Motor Skills Are Tested Using the Movement Assessment Battery for Children, 2nd Edition? *Children (Basel), 10*.

Smits-Engelsman, B., Denysschen, M., Lust, J. M., Coetzee, D., Valtr, L., Schoemaker, M., & Verbecque, E. (2024). Which outcomes are key to the assessment profile of a child with developmental coordination disorder? A systematic review and meta-analysis. *Biomedical Journal, unpublished, under review*.

Smits-Engelsman, B., Schoemaker, M., Delabastita, T., Hoskens, J., & Geuze, R. (2015). Diagnostic criteria for DCD: Past and future. *Hum Mov Sci, 42*, 293-306.

Smits-Engelsman, B., Verbecque, E., Denysschen, M., & Coetzee, D. (2022). Exploring Cultural Bias in Two Different Motor Competence Test Batteries When Used in African Children. *Int J Environ Res Public Health, 19*.

Smits-Engelsman, B., Vinçon, S., Blank, R., Quadrado, V. H., Polatajko, H., & Wilson, P. H. (2018). Evaluating the evidence for motor-based interventions in developmental coordination disorder: A systematic review and meta-analysis. *Res Dev Disabil, 74*, 72-102.

Smits-Engelsman, B. C., 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? *Res Dev Disabil, 32*, 1370-1377.

Steenbergen, B., Valtr, L., Dunford, C., Prunty, M., Bekhuis, H., Temlali, T. Y., van Abswoude, F., Lust, J. M., Warlop, G., Augustijn, M., Smits-Engelsman, B. C. M., & Wilson, P. H. (2024). Awareness about developmental coordination disorder. *Front Public Health, 12*, 1345257.

Vaivre-Douret, L., Lalanne, C., Cabrol, D., Golse, B., & Falissard, B. (2012). IMPORTANCE OF MULTIDIMENSIONAL ASSESSMENT TO REFINE SUBTYPES OF DEVELOPMENTAL COORDINATION DISORDER. *European Psychiatry, 27*.

Vaivre-Douret, L., Lalanne, C., Cabrol, D., Ingster-Moati, I., Falissard, B., & Golse, B. (2011). Identification of diagnostis criteria of developmental coordination disorder (DCD) or developmental dyspraxia subtypes. *Neuropsychiatrie de l'Enfance et de l'Adolescence, 59*, 443-453.

Vaivre-Douret, L., Lalanne, C., Ingster-Moati, I., Boddaert, N., Cabrol, D., Dufier, J. L., Golse, B., & Falissard, B. (2011). Subtypes of developmental coordination disorder: research on their nature and etiology. *Dev Neuropsychol, 36*, 614-643.

van der Linde, B. W., van Netten, J. J., Otten, E., Postema, K., Geuze, R. H., & Schoemaker, M. M. (2015). A systematic review of instruments for assessment of capacity in activities of daily living in children with developmental co-ordination disorder. *Child Care Health Dev, 41*, 23-34.

Van Dyck, D., Baijot, S., Aeby, A., De Tiège, X., & Deconinck, N. (2022). Cognitive, perceptual, and motor profiles of school-aged children with developmental coordination disorder. *Front Psychol, 13*, 860766.

Venetsanou, F., Kambas, A., Ellinoudis, T., Fatouros, I., Giannakidou, D., & Kourtessis, T. (2011). Can the movement assessment battery for children-test be the "gold standard" for the motor assessment of children with Developmental Coordination Disorder? *Res Dev Disabil, 32*, 1-10.

Verbecque, E., Coetzee, D., Ferguson, G., & Smits-Engelsman, B. (2021). High BMI and Low Muscular Fitness Predict Low Motor Competence in School-Aged Children Living in Low-Resourced Areas. *Int J Environ Res Public Health, 18*.

Visser, J. (2003). Developmental coordination disorder: A review of research on subtypes and comorbidities. *Human Movement Science, 22*, 479-493.

Wagner, M. O., Kastner, J., Petermann, F., & Bös, K. (2011). Factorial validity of the Movement Assessment Battery for Children-2 (age band 2). *Res Dev Disabil, 32*, 674-680.

Zoia, S., Biancotto, M., Guicciardi, M., Lecis, R., Lucidi, F., Pelamatti, G. M., Carrozzi, M., Skabar, A., Sugden, D. A., Barnett, A. L., & Henderson, S. E. (2019). An evaluation of the Movement ABC-2 Test for use in Italy: A comparison of data from Italy and the UK. *Res Dev Disabil, 84*, 43-56.

7. Tables

Table 1. Description of the MABC-2 items

MABC-2 Items	Age Band 1 (3–6 years)	Age Band 2 (7–10 years)	Age Band 3 (11-16 years)
		Manual Dexterity (MD)	
Item 1	Posting coins Preferred (a) /Non-preferred hand (b)	Placing pegs Preferred (a) /Non-preferred hand (b)	Turning Pegs Preferred (a) /Non-preferred hand (b)
Item 2	Threading beads	Threading lace	Triangle with nuts and bolts
Item 3	Drawing trail 1	Drawing trail 2	Drawing trail 3
		Aiming and Catching (AC)	
Item 4	Catching bean bag	Catching ball	Throw and catching ball with 1 hand Preferred (a) / Non-preferred (b)
Item 5	Throwing bean bag onto mat	Throwing bean bag onto mat	Throwing at wall-mounted target
		Balance (BAL)	
Item 6	One-leg balance Preferred (a) / Non-preferred (b)	One-board balance Preferred (a) / Non-preferred (b)	Two-board balance
Item 7	Walking heels raised	Walking heel-to-toe forwards	Walking heel-to-toe backwards
Item 8	Jumping on mats	Hopping on mats Preferred (a) / Non-preferred (b)	Hopping sideways on mats Preferred (a) / Non-preferred (b)

8. Figures

Figure 1: Description of the sample.

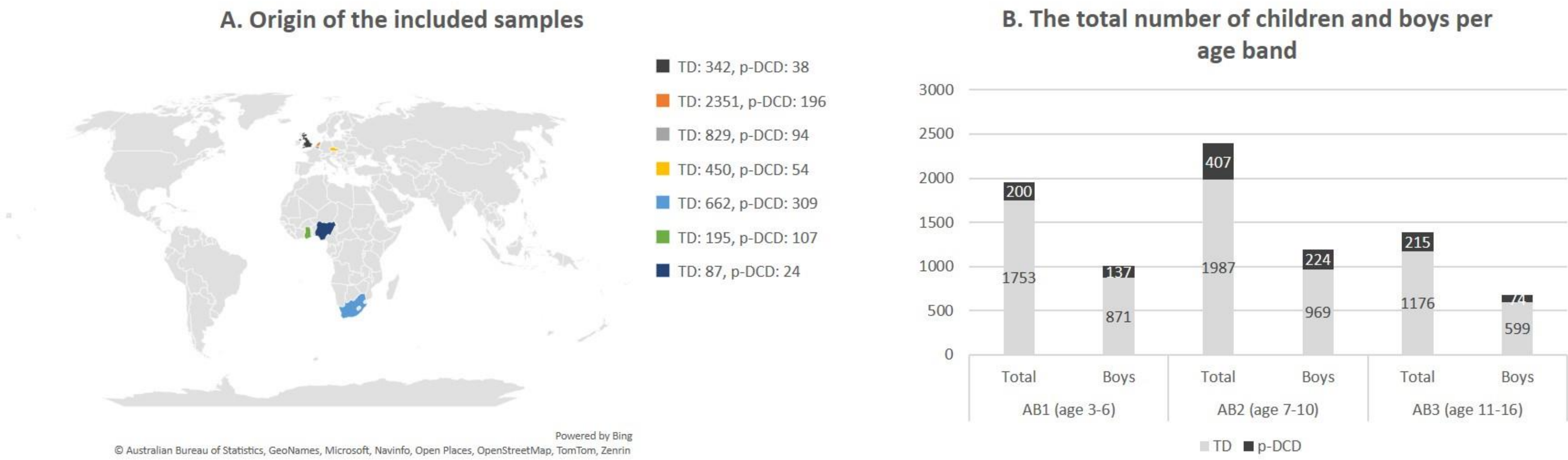
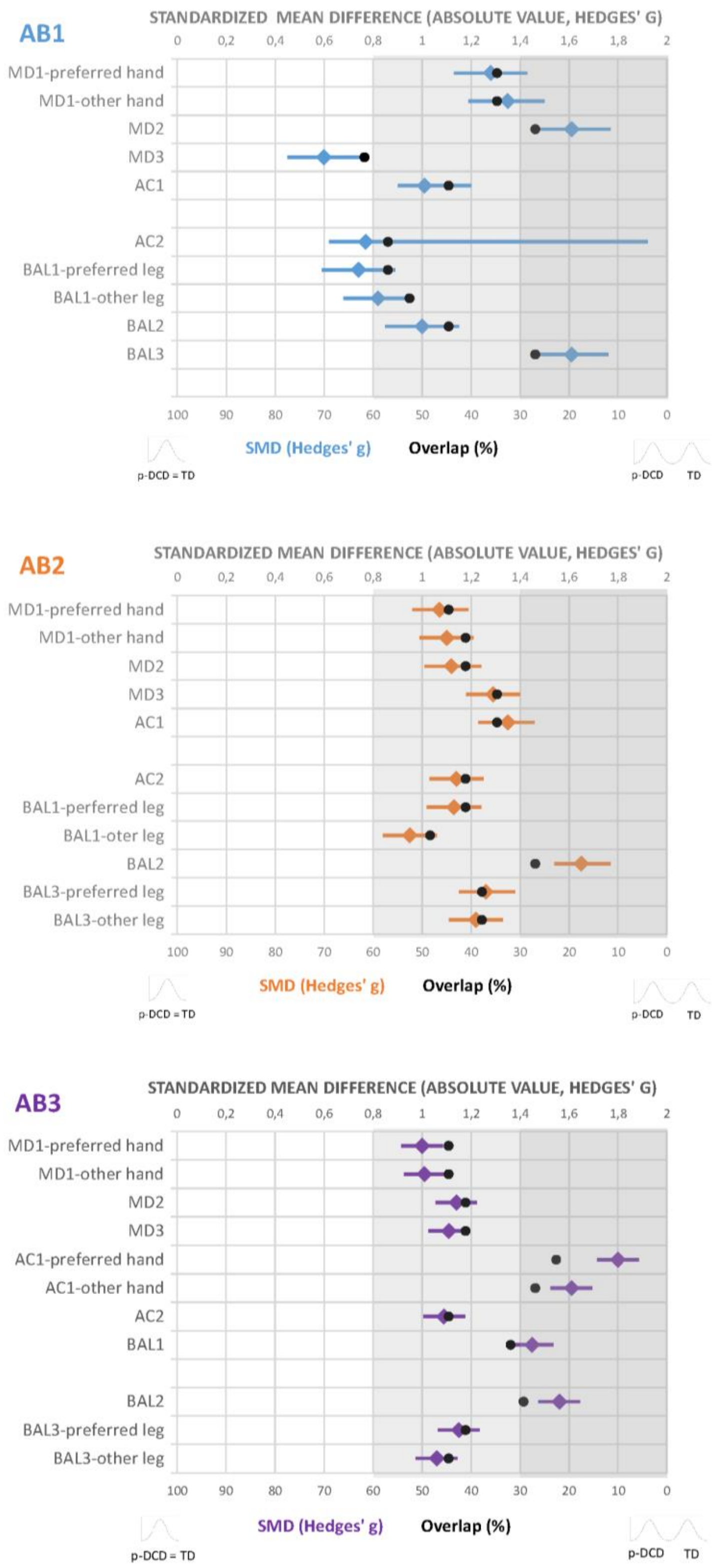


Figure 2: Standardized mean differences (SMD) between p-DCD and TD groups per item and age band and the percentage of overlap based on the mean SMD between these groups



Legend: Hedges' g is presented as the mean with its 95% confidence interval. Dark shaded area represents a standardized mean difference (SMD) of 1.4 or lager. Light shaded area represents an SMD between 0.8 and 1.4. Black dots are the percentages overlap between p-DCD and TD that is based on the mean SMD. Lower percentages of overlap indicate better sensitivity, which will correspond to higher SMD values.

Figure 3: Percentage of children (p-DCD versus TD) in the range of 'severe motor impairment' (item standard score 0-5) at item level

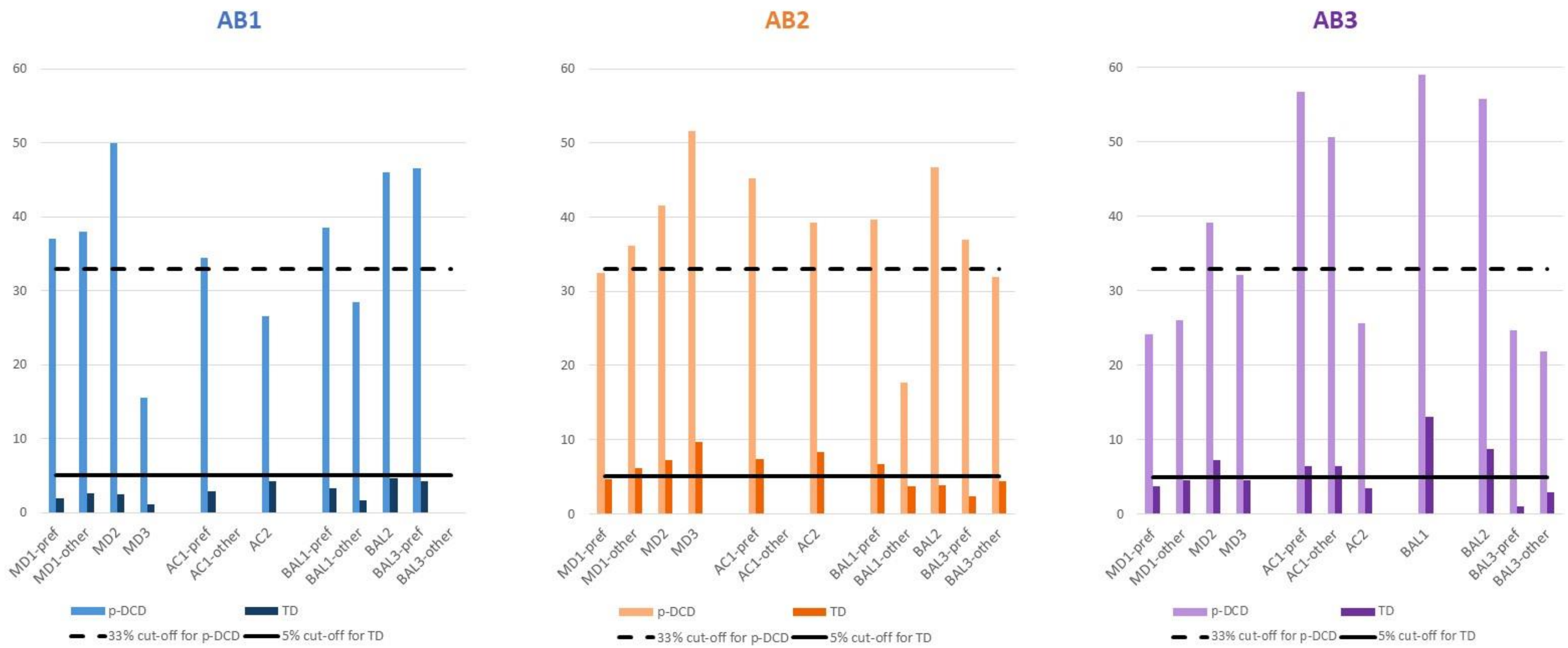
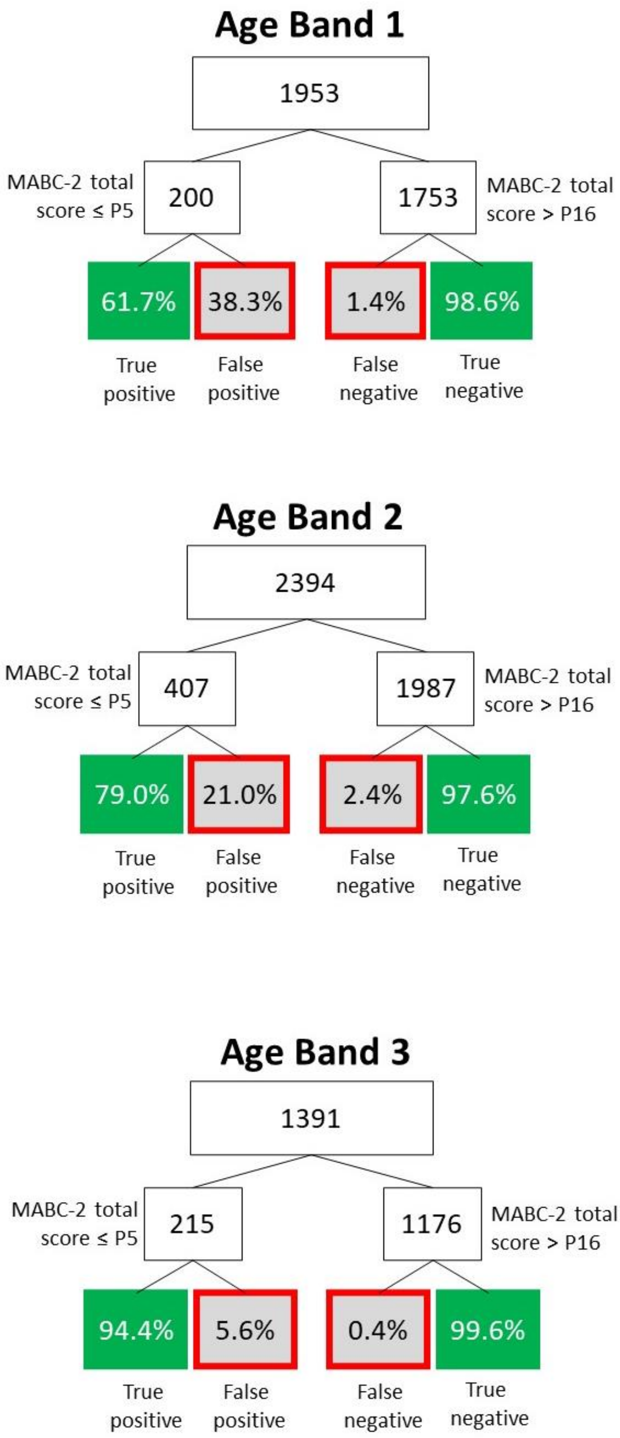


Figure 4: Sensitivity and specificity of MABC-2 items to predict p-DCD or TD

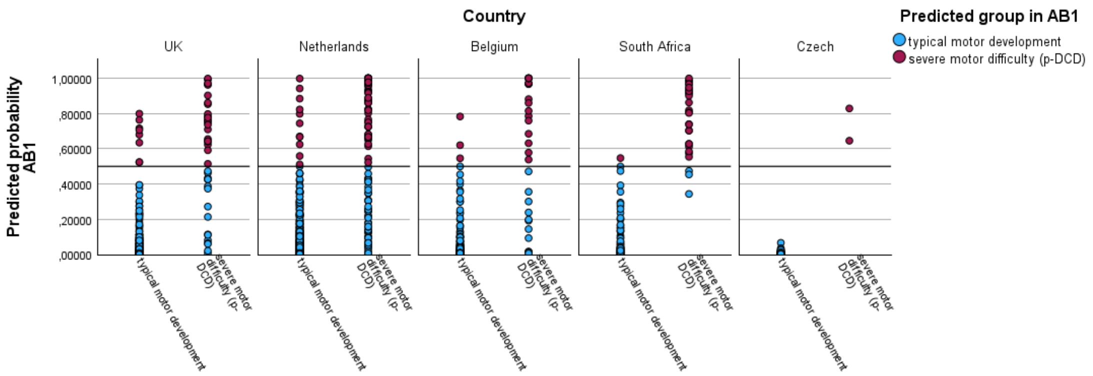


9. Appendix 1

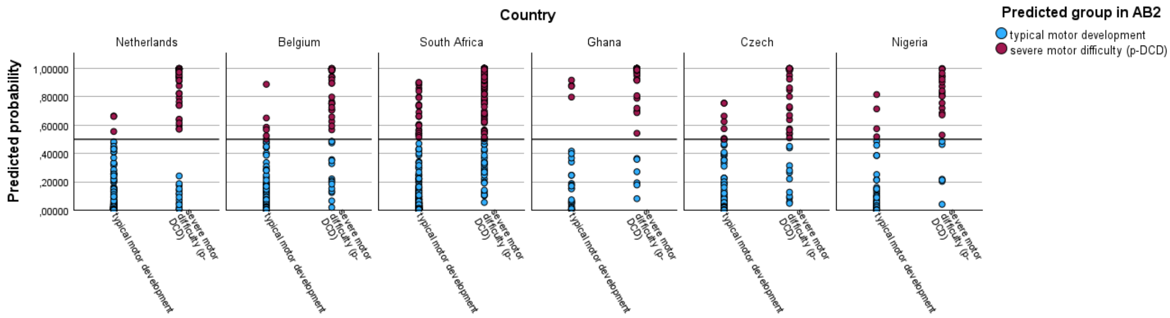
The diagnostic criteria for DCD as defined by the DSM-5-TR (Association., 2022)	
(A)	The acquisition and execution of coordinated motor skills is substantially below that expected given the individual's chronological age and opportunity for skill learning and use. Difficulties are manifested as clumsiness (e.g., dropping or bumping into objects) as well as slowness and inaccuracy of performance of motor skills (e.g., catching an object, using scissors or cutlery, handwriting, riding a bike, or participating in sports)
(B)	The motor skills deficit in Criterion A significantly and persistently interferes with activities of daily living appropriate to chronological age (e.g., self-care and self-maintenance) and impacts academic/school productivity, prevocational and vocational activities, leisure, and play.
(C)	Onset of symptoms is in the early developmental period.
(D)	The motor skills deficits are not better explained by intellectual disability or visual impairment and are not attributable to a neurological condition affecting movement.

10. Appendix 2

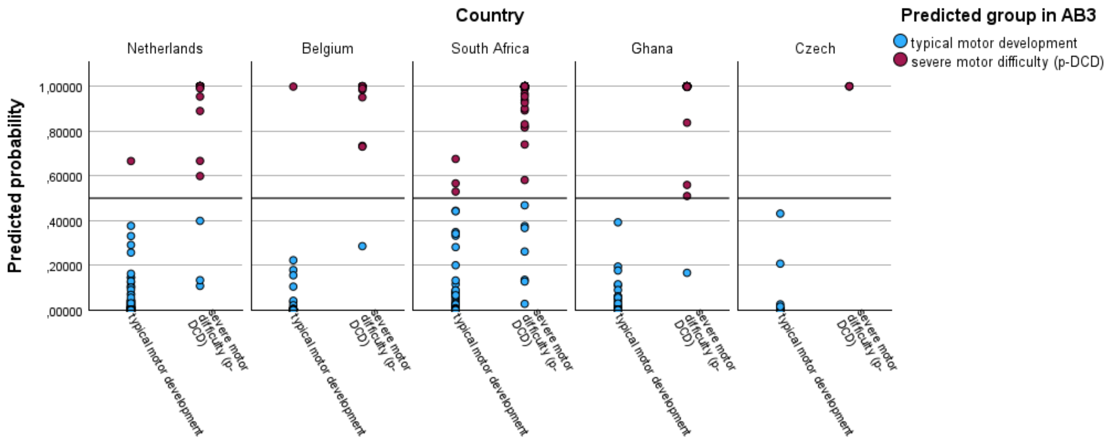
1) Age band 1



2) Age band 2



3) Age band 3



Note: The predicted probability (y-axis) based on the logistic regression model is presented as a function of the group assignment based on the MABC-2 total score (typical motor development (0) versus severe motor difficulty (p-DCD; 1)). Colored dots indicate whether the predicted probability would result in predicted typical development (0) or severe motor difficulty (p-DCD; 1). The horizontal line corresponds to a predicted probability of 0.5, which is the cut-off between the two categories. Red dots in each 'typical motor development' column indicate false positive results; blue dots in the 'severe motor difficulty' column represent false negative results. The results are presented separately per country.