JOURNAL OF SPORT & EXERCISE PSYCHOLOGY

Understanding School-Children's Perceived Movement Skill Competence: Evidence of Validity and Reliability According to the Type of Administration

Journal:	*Journal of Sport and Exercise Psychology*
Manuscript ID	JSEP.2023-0317.R3
Manuscript Type:	Original Research
Keywords:	self-perception, childhood, fundamental motor skills, motor development, physical education

Abstract:	The purpose of this study was to analyse and interpret evidence-based scores of validity and reliability of the Pictorial Movement Skill Competence (PMSC) according to the procedure of assessment (i.e., one-on-one interview and supervised-guided group administration) in a cohort of children. A sample of 635 children (49.1% girls, 8.0-10.9 years-old) participated. Confirmatory factor analyses (CFA) and Rasch models were conducted to examine evidence of different forms of validity and reliability. In the current sample, each CFA three factor PMSC model (locomotion, object control and stability skills) was an adequate fit. There was good-to-excellent agreement between both types of administration and higher internal consistency for the group administration. Both methods of administration are appropriate to assess PMC with the group method showing better reliability in this sample. Supervised group administration is more efficient than the one-on-one interview and could be considered in future assessments with children older than 8 years-old.

Understanding School-Children's Perceived Movement Skill Competence: Evidence of Validity and Reliability According to the Type of Administration

Abstract

The purpose of this study was to analyse and interpret evidence-based scores of validity and reliability of the Pictorial Movement Skill Competence (PMSC) according to the procedure of assessment (i.e., one-on-one interview and supervised-guided group administration) in a cohort of children. A sample of 635 children (49.1% girls, 8.0-10.9 years-old) participated. Confirmatory factor analyses (CFA) and Rasch models were conducted to examine evidence of different forms of validity and reliability. In the current sample, each CFA three factor PMSC model (locomotion, object control and stability skills) was an adequate fit. There was good-to-excellent agreement between both types of administration and higher internal consistency for the group administration. Both methods of administration are appropriate to assess PMC with the group method showing better reliability in this sample. Supervised group administration is more efficient than the one-on-one interview and could be considered in future assessments with children older than 8 years-old.

Key words: self-perception, childhood, fundamental motor skills, motor development, physical education.

Introduction

Perceived motor competence (MC) refers to a person's ability to recognise his/her proficiency in performing a wide range of fundamental motor skills (FMS) (Estevan & Barnett, 2018) and is an intrinsic factor of individuals' motor development influencing children's movement behaviour throughout their lifespan (Bardid et al., 2016; Barnett et al., 2022; Stodden et al., 2008). Perceived MC, as a branch of perceived physical competence and global self-concept (Fox & Corbin, 1989), is associated with enjoyment for physical activity (PA) participation (Babic et al., 2014; Menescardi et al., 2022).

Global data show the proportion of school children achieving PA is limited, with rates declining each year across childhood (Cooper et al., 2015; Farooq et al., 2018; Jago et al., 2017). As the conceptual model of motor development postulates (Stodden et al., 2008), perceived MC seems to mediate association between children's actual MC and PA (Barnett et al., 2022; De Meester et al., 2020). Those individuals with higher perceived MC tend to show higher motivation and enjoyment for PA (Cárcamo-Oyarzún et al., 2024), higher PA engagement (Babic et al., 2014) which in turn might foster actual MC (Estevan et al., 2023). Perceived MC might also stimulate self-esteem over time, especially in girls (Babic et al., 2014; Niemistö et al., 2019), generating a spiral of engagement towards positive healthy development such as physically active lifestyle or appropriate weight status (Stodden et al., 2008; Niemistö et al., 2019). Considering that children who underestimate their own skill might be less physically active (Estevan, Utesch, et al., 2019; Robinson et al., 2015), capturing children's perceived MC might be a preliminary step to conducting interventions to promote PA participation. Indeed, perceived MC is reported as a correlate of PA engagement (Estevan, Bardid, et al., 2021) as individuals with high perceived MC tend to persist and persevere in sports and motor challenges (Almeida et al., 2023; Estevan & Barnett, 2018).

Assessing perceived MC in childhood requires specific procedures to address issues such as 1) a construct structure that is qualitatively different across age periods and 2) children's cognitive development which may interfere with their capability to make true judgements about themselves (Estevan & Barnett, 2018; Harter & Pike, 1984). In this line, the assessment of perceived MC in childhood requires that instruments are configured according to 1) the structure and operationalization of the construct and 2) the individuals' cognitive development (Estevan & Barnett, 2018). With regard to the structure and configuration, to understand how well children's perceived MC are aligned with their actual MC, it is suggested that the constructs match or are identified in a similar manner (Estevan & Barnett, 2018); although this assumption is not confirmed yet (Barnett et al., 2022). A potential reason for this lack of confirmation is because during childhood, perceived MC development evolves in a non-linear evolution (Estevan & Barnett, 2018; Menescardi et al., 2022; van Veen et al., 2020) which might involve different understandings of perceived competence for children of similar age with consequences for their motor development. For instance, children initially think of 'effort' as the mastery of motor skills, but as they age, they seem to take a more abstract perspective in their descriptions (Dreiskaemper et al., 2018). In middle/late childhood an individual's ability to be realistic about his/her MC is purported to be more accurate (Dreiskämper et al., 2022; Estevan & Barnett, 2018). Regarding children's cognitive characteristics, Harter and Pike (1984) suggested that the instruments should be pictographic, offer an adequate number of responses, follow a procedure that allows the child to understand what is being asked and avoid socially desirable but unrealistic responses (discussed further below).

Two types of survey-based methods that are common in capturing perceived competence in childhood are the double dichotomous interview-based process and the self-report via Likert scale. Via pairwise preferences, the dichotomous process is preferred due to its ease of understanding and response for younger children, and its potential capability to eliminate the subjective notion of scaling (Barnett et al., 2016; Harter & Pike, 1984; Tietjens et al., 2018). On the other hand, via scaling, the rating or Likert scale (e.g., four-points) allows individuals to self-report their selected response from 1 to 4 in one response stage and this method is chosen for its consistency and sensitivity (Yannakakis & Hallam, 2011). However, despite the many instruments available for assessing actual MC, and the known methods to map perceived MC, there are few scales designed to assess perceived MC (Estevan, Bardid, et al., 2021).

The PMSC is highly used and was the first tool for assessing perceived MC in children in which pictorial items represent the same FMS that are commonly used to test actual MC (Barnett et al., 2016). FMS is typically categorised in three domains: stability, locomotion and object control (Gallahue et al., 2012), whereas the PMSC was originally composed of 12 - 13 FMS items only (encompassing locomotion and object control) as it was based on the second and third versions of the Test of Gross Motor Development (Ulrich, 2000; Webster & Ulrich, 2017). The PMSC test scores have shown evidence suggesting validity from many countries (Barnett et al., 2016; Chai et al., 2023; Diao et al., 2018; Maïano et al., 2022; Masci et al., 2018; Niemistö et al., 2019; Valentini et al., 2017; Venetsanou et al., 2018). Systematic review evidence shows that the association between children's actual and perceived MC depends on the type of the FMS domain under study (De Meester et al., 2020), therefore it is important scales cover all domains. Recently, an extended version of the PMSC including stability skills was created (Estevan, Menescardi, et al., 2021). The PMSC Stability is composed by seven items capturing different aspects and forms of stability (e.g., single-leg, unstable surface, dynamic balance, etc.). This scale also follows a pictorial dichotomous format to maximize children's ability to understand the constructs (Barnett et al., 2016; Harter & Pike, 1984). Evidence-based scores suggesting validity and reliability of the PMSC Stability in a sample of early adolescents (11-14 years-old) was reported. In this sample, test scores suggest construct validity (structural and convergent) and measurement invariance across gender, pointing to the PMSC in the two independent versions for boys and girls, being an appropriate scale to assess perceived MC in stability in adolescents

(Estevan, Menescardi, et al., 2021). But evidence for this scale version in primary/elementary school children is lacking.

To assess younger children's (8 years and less) perceived MC, individual interviews are recommended with support of the pictorial description (Barnett, Ridgers, et al., 2015). These procedures adapted to the characteristics of the children, have allowed reliable results providing an insight into the field of motor development and Physical Education (PE) (Barnett et al., 2016). However, in the educational setting the time devoted to PE is limited overall as the session duration is typically 45-60 min (with an actual teaching session time of around 40 minutes). Moreover, the student-teacher ratio in class groups (one teacher) ranges from 20-27 pupils. Therefore, conducting one-on-one individual interviews can be very time consuming which in turn may be a barrier for PE teachers wishing to assess perceived MC. Considering that a key aspect to check what students learn is to assess the content taught, not having appropriate instruments and procedures to assess perceived MC might result in students limited psychological development. In this line, López-Pastor et al (2013) argued that developing efficient, easy to use approaches to assess and compile data on students, is a challenge to be addressed in PE.

Unlike one-on-one individual interviews, supervised group administration can 1) be time efficient making dissemination potentially cost-effective and 2) be a suitable approach for older children (8-12 years-old and beyond) (Sport Australia, 2021, p. 8). Therefore, the purpose of the current study was to analyse and interpret evidence-based scores of validity and reliability of the PMSC (locomotion, object control and stability) according to the type of administration (i.e., one-on-one individual interview and supervised-guided group administration) in a sample of 3rd-5th grade primary school children. Based on previous studies (Estevan et al., 2018; Estevan, Menescardi, et al., 2021; Estevan, Molina-García, et al., 2019), it is expected that the construct of perceived MC in this age group (8-11 years-old) will exhibit a good structure using both types of administration.

Furthermore, construct validity in terms of 'relations with other variables' (McCoach et al., 2013) will be assessed based on the association between perceived MC and PA participation. This investigation will be seeking to assess whether the scores from the instrument using both administration types demonstrates associations that we would expect based on theoretical expectations, i.e., that higher perception is associated with higher PA (Carcamo-Oyarzun et al., 2023; De Meester et al., 2016).

Additionally, the present study aims to use two different test theoretical approaches (classical and probabilistic test theory) to analyse and interpret the scores of reliability and validity of the scale in a cohort of children using the two types of administration. Whereas the classical test theory describes reliability as a point measure, probabilistic test theory understands reliability as a probability function across the level of perceived MC for each dimension of perceived MC which can lead us to understand reliability in terms of different levels of difficulty by participants.

Materials and Methods

Participants

Six primary schools (four public and two private-subsidized) in Valencia and the metropolitan area (Spain) were recruited randomly based on a publicly available list. All the school principals were contacted and agreed to participate in the study. A total sample of 635 children (49.1% girls), 8 to 11 years of age (M = 9.37, SD = 0.94) and a mean body mass index of 19.17 $kg \cdot m^{-2}$ (SD = 3.49) participated with parent consent (90.6% consent rate). The response rate was 88.2%. Three inclusion criteria were established for data analysis: (1) were between 8 and 11 years old; (2) had Spanish as a main language at school; (3) provided verbal assent to be involved.

Procedure

Data collection was conducted by trained research assistants (all specialists in teaching PE) in the respective schools between February and May 2021 during school hours. Approval was obtained from the Institutional Review Board of the University of the first author (1564606).

To assess children's perceived MC, a counterbalanced-based two-session procedure was conducted as follows. Before any assessment, children were provided with an explanation of the study and any doubts were clarified prior to questionnaire administration. In session A (one-on-one interview), the PMSC (Barnett et al., 2016; Estevan, Molina-García, et al., 2019), was administered in a one-on-one individual interview using a double-dichotomized process in a calm and private room. Research assistants read out loud each item and guided each participant individually by providing a verbal description and demonstration of the skill if needed. Children were instructed to point at which picture looked like them (first dichotomized choice); then for their marked picture they were asked again to point at their perceived MC (second dichotomized choice). The time required for assessing children' perceived MC (including the PMSC and PMSC Stability) was approximately 10-15 minutes per child. In session B (group administration), in the regular class each student completed an inventory composed of the PMSC (Estevan et al., 2018), the PMSC Stability (Estevan, Menescardi, et al., 2021) both adapted to be self-reported and the Physical Activity Questionnaire for Children (PAQ-C; Kowalski et al., 2004). The administration of these questionnaires was verbally guided by one research assistant to guarantee that participants understood each item. A second research assistant was available for any specific student query. The time required for assessing participants' perceived MC and PA participation was approximately 25 minutes per group (around 22-27 children).

In the first session, regardless of the type of administration, all children were initially asked if they had tried the skill previously. Children who had not experience or who did not recognize the skill, received a demonstration by the research assistant but were not allowed to perform it. The majority of children recognized the depicted skills; there were only few cases where galloping (locomotion) or superwoman/superman (stability) required additional demonstration. The time between sessions A and B was 10-14 days (M = 12.83 days, SD = 1.60).

Instruments

Perceived motor competence. Children's perceived MC was assessed with the Spanish version of the PMSC (Barnett, Ridgers, et al., 2015; Estevan, Molina-García, et al., 2019) and the PMSC Stability (Estevan, Menescardi, et al., 2021). The PMSC consisted of 13 items, six of them representing locomotor and seven representing object control skills (Estevan, Molina-García, et al., 2019); whereas the PMSC Stability consisted of seven items representing stability skills (Estevan, Menescardi, et al., 2021). To assist children's rating, each item in both scales is accompanied by two pictures depicting an appropriate task performance (wherein a girl/boy performs the skill correctly) and a lower performance (wherein a girl/boy performs the skill poorly). Each scale item is rated on a 4-point scale: from one being 'not so good', two being 'sort of good', three being 'pretty good', and four being 'really good'. As a result, scores in each scale range from 13 to 52 for the PMSC total scale and from 7 to 28 for the PMSC stability.

Physical Activity participation. To assess students' PA, the Spanish version of the self-reported PAQ-C (Kowalski et al., 2004; Martínez-Gómez et al., 2009) was used. The

PAQ-C is a 7-day recall instrument that consists of 9 items providing PA information during a school week. Each of these 9 items is rated on a Likert style from 1 (e.g., none) to 5 (e.g., 7 or more times), with a lower score indicating low PA participation. Finally, a total relative score ranging from 1 to 5 is reported. A tenth item is used to capture potential unusual weak PA levels due to illness. Those participants reporting illness repeated the PAQ-C in the second session. The internal consistency (α = .78) derived from the test scores, supports the interpretation of the PAQ-C being considered good for the sample under study. PA was categorised as either meeting or not meeting PA guidelines based on a cut-off score of 2.75 (Tapia-Serrano et al., 2022).

Statistical analysis

All analyses were conducted in R (R Core Team, 2018) using primarily the *tidyverse* packages for data preparation, the *lavaan* package for structural equation modelling and the *mirt* package (Chalmers, 2012) for item response theory models. To answer our research questions, we used two different test theoretical approaches to evaluate the reliability and the validity of the PMSC scale. That is, we aimed at using the strength of both classical and probabilistic approaches. Classical approaches have their strength in widespread usage, the potential to easily describe the data, and to identify factor structures under relatively weak assumptions. However, probabilistic approaches can add to the comprehensive testing of an instrument as not only the structure is tested but also the usage of the scale itself can be illustrated. The probabilistic approach is especially suited to test whether a 4-point scale is interval scaled or only ordinally scaled. That is, if the thresholds of the scale usage are ordered, the scale is proven to be interval scaled.

The analysis plan proceeded with the following steps: First, we checked differences in means between the types of administration and describe standard errors using paired t-tests. Second, for construct validity we tested evidence of structural validity using confirmatory factor analyses (CFA) for both the PMSC and PMSC Stability in each type of administration (i.e., one-on-one interview and group administration). Models were fit using WLSMV estimators and fit indices of CFAs were interpreted as being good for $\chi^2/df < 2.00$, $p < .05$, TLI > .90, CFI > .90, and RMSEA < .08 (Kelava et al., 2011).

Third, to test evidence of construct validity in terms of relations with other variables, the correlation of PMSC and PMSC Stability with self-reported PA was calculated. Moreover, to test evidence of concurrent validity, the association between PMSC locomotion and object control with PMSC Stability was calculated. Significance levels were set to $p = .05$ for all models.

Fourth, reliability in terms of evidence of internal consistency (i.e., Cronbach's alpha) from classical test theory as well as reliability (rxx) from probabilistic test theory were calculated. The difference between both test theoretical approaches is that classical test theory can describe reliability only as a point measure for average trait levels while the probabilistic estimation gives information as an information curve that reflects reliability across all trait levels. Reliability was interpreted acceptable for .60 and above because participants are children and the self-concept is usually inflated, i.e., (Estevan, Menescardi, et al., 2021; Estevan & Barnett, 2018). Using probabilistic test theory, we can describe reliability as a point measure and, additionally, as function across different levels of the PMSC scales. As an example illustration, Figure 1 (A, B) represents two reliability functions of two scales or domains of self-perception. Considering the average difficulty (zero value), in the example A (Figure 1A), reliability function is high across all self-perception levels whereas in example B (Figure 1B) reliability function is low for the top half of self-perception levels, which would indicate low reliability for the top half.

***** Figure 1 *****

Fifth, we used probabilistic test theory underpinning the validation process and to compare the two different types of administrations. We used Generalized Partial Credit Models (GPCM) to investigate the scale usages or perceptions and whether the scale fulfils the assumptions of an interval scale. To do so, we followed several steps: (1) A probability function was calculated for each item and for each category that refers to the four options of response in each item. In this process, three thresholds (due to the points where two functions crossed each other: b1, b2 and b3) between item categories were calculated for each item, and the lengths between the thresholds were also computed. It is necessary for the interpretation of outputs as acceptable validity of the scale in the sample under study that the thresholds are ordered (i.e., b1 < b2 < b3). The scale can then be identified as being interval scaled as we can count how many thresholds have been achieved. The thresholds were computed on the same (logit) scale as the self-perception (the assessed trait) of the participants. Hereby, a value of zero means that the threshold is equivalent to an average trait. For illustration purposes, (2) we illustrated the category probability functions, the thresholds and the length between thresholds for the two previous items in Figure 1 (C, D). In Figure 1C, the threshold $b2_c$ is approximately zero, which means that for those individuals who rated their perceived competence in the lower half, it was most probable to score 1 or 2, whereas for those with scores in the top half, self-perception is most probable as a score of 3 or 4. However, in Figure 1D the threshold $b3_d$ is approximately zero, which shows that category four is most probable for the top half of self-perception in this item. Additionally, the thresholds 1 ($b1_c > b1_d$) and 2 ($b2_c > b2_d$) are higher in Figure 1C compared to the example in Figure 1D which indicates that the item in Figure 1C is different in terms of difficulty for individuals compared to the items in Figure 1D. Further, the lengths between thresholds are different in the example ($y_c > y_d$ and $z_c < z_d$), which means that the scale perceptions are different, too. In Item C, all categories are used equally distanced and in item D the perceived distances between the categories are different (i.e., category 2 in pink is only most probable for a small part of the assessed trait on the x axis, whereas category 3 in green is most probable for a larger part of trait level). The showcase between those two items illustrates two very different items that correspond to the reliability functions (A and C, B and D). As a last step, (3) the item fits, which are a statistical test of the fit of an item to the scale dimension, are modelled for both types of administration using signed chi-squared tests. Items fit to the dimension in those cases wherein they do not significantly deviate. Significance level was adjusted to $\alpha = .001$ due to multiple testing (Kang & Chen, 2007).

An a priori power analysis was not conducted for the following reasons. For applying CFA and Partial Credit Models, at least 20 participants per item is recommended. With 400 needed participants, we oversampled to cover a potential participation rate of 50% per class. With 625 participants and 20 items, this number is covered. Regarding the correlations and tests difference used in this study, the number of 625 participants far exceeds the required number even for very small effect sizes.

Results

First, differences in means between the types of administration were analysed (see Table 1.). Descriptively, locomotion showed the highest levels followed by object control and stability. Overall, means were significantly smaller for the group administration type compared to the one-on-one interview for all three categories (all *p*s < .001).

Second, CFAs were conducted to examine construct validity in terms of factorial validity in both types of administration. The children self-report modelling exhibited that the obtained test scores can be considered as excellent goodness of fit statistics for both types of administration (see Table 1). Thus, factorial validity is given for both scales based on

CFA. Further, all estimates show the expected fit to the factors as all paths are higher than .40 (see supplement Table S1).

***** Table 1 *****

Third, relation with other variables was also tested. Here, Table 2 shows the results of the associations which were run between groups with lower or higher PA levels ($p < .006$). The results show significant differences for each variable and both types of administration in that higher PA levels were associated with higher perceived MC levels.

***** Table 2 *****

Fourth, there were significant small to medium relationships between PMSC and PMSC Stability with self-reported PA. Also, in both types of administration there was also a significant medium to large association between PMSC locomotion and object control with PMSC Stability (Table 3).

Fifth, reliability scores (Table 3) showed that reliability point estimates are good for the group administration. However, they are too low for the one-on-one interview (with the exception of object control) considering the .60 threshold. In comparison, reliability functions show that reliability is shifted towards lower levels of perceived MC across the measures (Figure 2, first column). It can be derived in detail from the functions that the high levels of perceived MC scores indicate less reliability especially in the one-on-one interview for locomotion and stability (Figure 2, D and F first column), which can be explained by the scale usage.

***** Figure 2 and Table 3 *****

Sixth, scale usage and perception were analysed. GPCM threshold analysis shows that all thresholds were ordered for all the items, i.e., $b_1 < b_2 < b_3$ (Table 4) as it is illustrated in Figure 2 (columns 2 to end). The fact that the thresholds are ordered indicates that scale perception is generally pointing towards the validity of the scales, because the length between thresholds is not a criterion in terms of validity. However, we found a general pattern indicating a skewness in the scale usage between the two types of administration: b_2 is generally lower for the one-on-one interview throughout the scales. Practically, this means that the score 2 is less probable for children to choose given the same level of perceived MC while it is more probable that the scores 3 and 4 are chosen for lower levels of perceived MC. Further, the length between b1 and b2 is very small (or even tiny) for example in locomotion (Figure 2, B). This means that category 2 has a very tiny margin of the trait where it is most probable, which is not desired when developing such a scale. Looking at the single items, it is visible that the items superman/woman and move sideways deviate from the pattern of all other items and it can be interpreted that the perception of these items is not as originally intended by the test.

***** Table 4 *****

Discussion

For the assessment of children's perceived MC, a one-on-one interview has been suggested because some children may be unfamiliar with some movement skills and consequently, they would require further description beyond the visual picture provided (Barnett, Robinson, et al., 2015). This is true for young children, but uncommon for 8–12-year-old children who have usually been exposed to all the movement skills included in the PMSC (Estevan, Molina-García, et al., 2019), or if not, can still recognize them after seeing them performed by others. However, this method of administering (one-on-one interview) is time consuming in the school setting. Therefore, the purpose of the current study was to analyse and interpret evidence-based scores of validity and reliability of the pictorial scale PMSC (locomotion, object control and stability) according to the type of administration (i.e., one-on-one interview and supervised group administration) in a sample of primary school children aged 8-11 years-old. Further, scale perceptions and the consequences for its usage between the two types of administration were analysed and compared.

Based on classical test theory, the scores demonstrated evidence of structural and construct validity for the PMSC in locomotion, object control and stability for the two types of administration in the sample under study. In line with a previous study using the extended version of the PMSC in early adolescents (Estevan, Menescardi, et al., 2021), the results of the CFA in the current study showed an excellent fit. That is, in both the one-on-one interview and the supervised group administration, structural validity was confirmed by conducting CFA which supports the view that the pictograms help children to understand the content of the scale properly (Tietjens et al., 2018). Although the fit statistics were slightly better for the one-on-one interview, both goodness of fit tests were excellent. In addition, comparative analyses to assess the relationships between the scores from the PMSC and PA in a manner that was expected (association with PA levels) showed that those who practiced lower PA also exhibited lower perceived MC (and the converse), provided further validity evidence.

Despite the appropriate pictorial description, the procedure of administration seems to impact the reliability of children's self-reports. Whereas a supervised-guided group administration exhibited acceptable score of reliability, the one-on-one interview showed low levels of reliability estimates for two of the three scales. Heterogeneous levels in the means of PMSC according to the type of FMS have previously been reported – although these are mainly acceptable values. See for example, in 5-10 years-old Australian (Barnett, Ridgers, et al., 2015; Johnson et al., 2016), 4-8 years-old Brazilian (Valentini et al., 2017), 4-9 years-old Chinese (Diao et al., 2018), 5-9 years-old Greek (Venetsanou et al., 2018), around 7 years-old Italian (Masci et al., 2018) and 4-11 years-old Spanish children (Estevan et al., 2018; Estevan, Molina-García, et al., 2019). Different values might be due to the children's cognitive development, as in older children and early adolescents the internal consistency findings in one study were slightly higher (Estevan, Menescardi, et al., 2021) than in the current study. In addition, when the descriptives were compared according to the procedure of administration, individuals reported higher perceived MC when using the one-on-one interview than what they responded by using a supervised-guided group administration. This might indicate that the group session approach minimised children's socially desirable responses when capturing their perceived competence (Dreiskämper et al., 2022; Estevan & Barnett, 2018; Harter & Pike, 1984). This was an unexpected result because the one-on-one interview process of administration is suggested to avoid this socially desirable responses; however, in line with an approach recommended by the Australian Sports Commission (Sport Australia, 2021) for physical literacy, it seems that for children 8-12 years-old, a supervised-guided group administration is an appropriate method to map children's self-reported MC. As such, the class setting with a supportive style may avoid social desirability and therefore improve evidence of reliability because individuals are confident enough of the importance of being sincere in their responses (Ortega-Benavent et al., 2024). Therefore, the instrument reporting procedure is an important consideration for the validity and reliability of the instrument (Yannakakis & Hallam, 2011) overall during childhood.

When using probabilistic test theory to analyse reliability function, the results showed a general pattern wherein the supervised group administration reliability was good across low to high levels of perceived MC. In addition, when using the supervised group administration, in the majority of the items the distance between thresholds were similar [see for instance Figure 2, lines A (*gallop* and *jump forward*), C (*overhand* and *underhand throw*) and E (*unipodal* and *walking backwards*)] which reflects the equivalent probability to choose scores 1 or 2 for those in the lower half of self-perception and scores

3 or 4 for those in the top half. In the case of the one-on-one interview, the results showed that the reliability function skewed to the left (Figure 2, lines B, D and F) which means reliability is low for those reporting high perceived MC, whereas it is good in those children who considered themselves to have low perceived MC (Table 5). Specifically in locomotion [see for instance Figure 2, line B (run, gallop, hop, skip, jump forward and slide)] but also in stability skills [see Figure 2, line F (unipodal and move sideways)], it is more probable that children at the lower half of self-perception reported a score of 1 but not a score of 2; on the other hand, the reliability for those at the top half was low. So, for assessing perceived MC in typically developing children 8-11 years-old, PE teachers and researchers may wish to use the supervised-guided group administration.

In addition, the generalized partial credit models showed a consistent pattern of scale perception within the group administration and within the one-on-one interview. Comparing both, the length between the thresholds 1 and 2 is generally larger in the group administration compared to the one-on-one interview situation and vice versa between thresholds 2 and 3. We interpret this as an indicator for the level of perceived MC where scales can better differentiate. With larger space between thresholds, a larger part of the x axis (representing perceived MC level) is covered by the same item answer and, thus, it cannot differentiate that well in this space. That is, with larger length for the lower perceived MC levels and lower length between thresholds for higher perceived MC levels, the group administration can differentiate better for higher perceived MC levels compared to the one-on-one individual interview.

In this study more than 600 primary school children aged 8-11 years-old participated which is an excellent sample size to analyse evidence-based scores of validity and reliability of the extended version of the PMSC. In addition, the novelty of the study design resides in using two different counterbalanced procedures of administration which can be helpful for PE teachers and researchers looking to maximise feasibility of assessing perceived MC in the school setting. Nonetheless, this study is not absent of limitations. Whilst thorough data analysis was conducted by applying classical and probabilistic test theories, actual MC was not assessed so it was not possible to clarify criterion validity in terms of whether a one-on-one interview or a supervised-guided group administration matches or how accurately it reflects actual MC. In addition, when required during the assessments, research assistants performed the depicted skill to help individuals to understand the exact meaning of the item and this might impact children's self-perception. Some research has suggested that animated skilled performances by another child might be more relatable to children (Copetti et al., 2022) but we do not have evidence of it yet. So, future studies should replicate the current study according to the process of administration and using appropriate animated-based video performances to clarify the exact meaning of movement skills. In addition, it is also recommended to assess perceived MC in combination with actual MC to analyse veridicability (i.e., whether those who consider themselves highly competent are indeed motor competent and vice versa) (Almeida et al., 2023) because such congruency might lead to higher PA participation during childhood (Utesch et al., 2018) and adolescence (Philpott et al., 2021).

Practical implications

Generally, assessing individuals' MC can be time consuming. The current study shows that for children aged 8-11 years-old, a 4-point scale is an appropriate number of choices when accompanied by an appropriate guide, supervision and pictorial descriptions and allows PE teachers and researchers to capture perceived MC in a valid and reliable manner. Thus, considering that supervised-guided group administration 1) is time efficient (as it is possible to assess a group of 25 in 15 minutes), 2) avoids inflated selfperceptions and 3) is reliable enough and provides evidence of construct validity in children, the extended version of the PMSC (locomotion, object control and stability) could be applied by using an appropriate supervised-guided group administration process which is characterized by providing comprehensive instruction and supportive accompaniment to the students.

In addition, the results of the current study indicated that the procedure of administration to map students' perceived MC might impact their responses in terms of validity and reliability and consequently our capability to capture appropriate levels of self-perception. In this line, for those students with low levels of perceived MC, an alternative could be following a one-on-one interview because it can be an effective method of administration. On the other hand, a supervised-guided group administration could be a suitable choice by a PE teacher who would like to capture students' perceived MC in a typically developing group because the reliability function and the results from the generalized partial credit models were good for the whole range of answers.

Conclusions

The extended version of the PMSC exhibited evidence-based scores of validity and reliability in a sample of Spanish children aged 8-11 years-old. This version of the PMSC includes pictograms about locomotion, object control and stability skills accompanying the descriptive text which can help PE teachers and researchers to achieve a proper assessment of perceived MC. By following the two processes of administration tested, and thanks to the analyses based on the probabilistic test theory, the supervised-guided group administration is an appropriate method to map these older primary school children's perceived MC.

References

Almeida, G., Luz, C., Rodrigues, L. P., Lopes, V., & Cordovil, R. (2023). “Profiles of motor competence and its perception accuracy among children: Association with physical fitness and body fat". *Psychology of Sport and Exercise, 68*, 102458. https://doi.org/10.1016/j.psychsport.2023.102458

Babic, M. J., Morgan, P. J., Plotnikoff, R. C., Lonsdale, C., White, R. L., & Lubans, D. R. (2014). Physical activity and physical self-concept in youth: Systematic review and meta-analysis. *Sports Medicine*, *44*(11), 1589–1601. https://doi.org/10.1007/s40279-014-0229-z

Bardid, F., De Meester, A., Tallir, I., Cardon, G., Lenoir, M., & Haerens, L. (2016). Configurations of actual and perceived motor competence among children: Associations with motivation for sports and global self-worth. *Human Movement Science*, *50*, 1–9. https://doi.org/10.1016/j.humov.2016.09.001

Barnett, L. M., Ridgers, N. D., Zask, A., & Salmon, J. (2015). Face validity and reliability of a pictorial instrument for assessing fundamental movement skill perceived competence in young children. *Journal of Science and Medicine in Sport*, *18*(1), 98–102. https://doi.org/10.1016/j.jsams.2013.12.004

Barnett, L. M., Robinson, L. E., Webster, E. K., & Ridgers, N. D. (2015). Reliability of the Pictorial Scale of Perceived Movement Skill Competence in 2 Diverse Samples of Young Children. *Journal of Physical Activity & Health*, *12*(8), 1045–1051. https://doi.org/10.1123/jpah.2014-0141

Barnett, L. M., Vazou, S., Abbott, G., Bowe, S. J., Robinson, L. E., Ridgers, N. D., & Salmon, J. (2016). Construct validity of the pictorial scale of Perceived Movement Skill Competence. *Psychology of Sport and Exercise*, *22*, 294–302. https://doi.org/10.1016/j.psychsport.2015.09.002

Barnett, L. M., Webster, E. K., Hulteen, R. M., De Meester, A., Valentini, N. C., Lenoir, M., Pesce, C., Getchell, N., Lopes, V. P., Robinson, L. E., Brian, A., & Rodrigues, L. P. (2022). Through the Looking Glass: A Systematic Review of Longitudinal Evidence, Providing New Insight for Motor Competence and Health. *Sports Medicine*, *52*, 875–920. https://doi.org/10.1007/s40279-021-01516-8

Carcamo-Oyarzun, J., Herrmann, C., Gerlach, E., Salvo-Garrido, S., & Estevan, I. (2023). Motor competence, motivation and enjoyment in physical education to profile children in relation to physical activity behaviors. *Physical Education and Sport Pedagogy*, 1–16. https://doi.org/10.1080/17408989.2023.2265399

Chai, H., Xue, R., Yao, L., Miao, M., & Han, B. (2023). Configurations of actual and perceived motor competence among elementary school children in China: Differences in physical activity. *Frontiers in Public Health*, *11*. https://doi.org/10.3389/fpubh.2023.1280643

Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, *48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Cooper, A. R., Goodman, A., Page, A. S., Sherar, L. B., Esliger, D. W., van Sluijs, E. M., Andersen, L. B., Anderssen, S., Cardon, G., Davey, R., Froberg, K., Hallal, P., Janz, K. F., Kordas, K., Kreimler, S., Pate, R. R., Puder, J. J., Reilly, J. J., Salmon, J., … Ekelund, U. (2015). Objectively measured physical activity and sedentary time in youth: The International children's accelerometry database (ICAD). *The International Journal of Behavioral Nutrition and Physical Activity*, *12*, 113. https://doi.org/10.1186/s12966-015-0274-5

Copetti, F., Valentini, N. C., Deslandes, A. C., & Webster, E. K. (2022). Pedagogical support for the Test of Gross Motor Development – 3 for children with neurotypical development and with Autism Spectrum Disorder: Validity for an animated mobile application. *Physical Education and Sport Pedagogy*, *27*(5), 483–501. https://doi.org/10.1080/17408989.2021.1906218

De Meester, A., Barnett, L. M., Brian, A., Bowe, S. J., Jiménez-Díaz, J., Van Duyse, F., Irwin, J. M., Stodden, D. F., D'Hondt, E., Lenoir, M., & Haerens, L. (2020). The Relationship Between Actual and Perceived Motor Competence in Children, Adolescents and Young Adults: A Systematic Review and Meta-analysis. *Sports Medicine*, *50*(11), 2001–2049. https://doi.org/10.1007/s40279-020-01336-2

De Meester, A., Maes, J., Stodden, D., Cardon, G., Goodway, J., Lenoir, M., & Haerens, L. (2016). Identifying profiles of actual and perceived motor competence among adolescents: Associations with motivation, physical activity, and sports participation. *Journal of Sports Sciences*, *34*(21), 2027–2037. https://doi.org/10.1080/02640414.2016.1149608

Diao, Y., Barnett, L., Estevan, I., Dong, C., & Li, J. (2018). Validity and Reliability of a Pictorial Instrument for Assessing Fundamental Movement Skill Perceived Competence in Chinese Children. *Journal of Motor Learning and Development*, *6*(S2), S223–S238. https://doi.org/10.1123/jmld.2016-0082

Dreiskaemper, D., Tietjens, M., & Schott, N. (2022). The physical self-concept across childhood: Measurement development and meaning for physical activity. *Psychology of Sport and Exercise*, 102187. https://doi.org/10.1016/j.psychsport.2022.102187

Dreiskaemper, D., Utesch, T., & Tietjens, M. (2018). The Perceived Motor Competence Questionnaire in Childhood (PMC-C). *Journal of Motor Learning and Development*, *6*(s2), S264–S280. https://doi.org/10.1123/jmld.2016-0080

Estevan, I., Bardid, F., Utesch, T., Menescardi, C., Barnett, L. M., & Castillo, I. (2021). Examining early adolescents' motivation for physical education: Associations with actual and perceived motor competence. *Physical Education and Sport Pedagogy*, *26*(4), 359–374. https://doi.org/10.1080/17408989.2020.1806995

Estevan, I., & Barnett, L. M. (2018). Considerations Related to the Definition, Measurement and Analysis of Perceived Motor Competence. *Sports Medicine*, *48*(12), 2685–2694. https://doi.org/10.1007/s40279-018-0940-2

Estevan, I., Meester, A. D., & Barnett, L. M. (2023). Advancing the understanding in the association between actual/perceived motor competence and health-related factors among children and adolescents. *Psychology of Sport and Exercise*, *67*, 102420. https://doi.org/10.1016/j.psychsport.2023.102420

Estevan, I., Menescardi, C., Castillo, I., Molina-García, J., García-Massó, X., & Barnett, L. M. (2021). Perceived Movement Skill Competence in Stability: Validity and Reliability of a Pictorial Scale in Early Adolescents. *Scandinavian Journal of Medicine & Science in Sports*, *31*(5), 1135–1143. https://doi.org/10.1111/sms.13928

Estevan, I., Molina-García, J., Abbott, G., Bowe, S., Castillo, I., & Barnett, L. (2018). Evidence of Reliability and Validity for the Pictorial Scale of Perceived Movement Skill Competence in Spanish Children. *Journal of Motor Learning and Development*, *6*(S2), S205–S222. https://doi.org/10.1123/jmld.2016-0065

Estevan, I., Molina-García, J., Queralt, A., Bowe, S. J., Abbott, G., & Barnett, L. M. (2019). The new version of the pictorial scale of Perceived Movement Skill Competence in Spanish children: Evidence of validity and reliability. [La nueva versión de la escala pictográfica de Percepción de Competencia de Habilidades Motrices in niños y niñas españoles: Evidencias de validez y fiabilidad]. *RICYDE. Revista Internacional de Ciencias Del Deporte*, *15*(55), 35–54. https://doi.org/10.5232/ricyde2019.05503

Estevan, I., Utesch, T., Dreiskämper, D., Tietjens, M., Barnett, L. M., & Castillo, I.

(2019). Validity and reliability of a pictorial scale of physical self-concept in spanish children. [Validez y fiabilidad de la escala pictográfica de autoconcepto físico en niños y niñas españoles]. *RICYDE. Revista Internacional de Ciencias Del Deporte*, *15*(55), 102–118. https://doi.org/10.5232/ricyde2019.05507

Farooq, M. A., Parkinson, K. N., Adamson, A. J., Pearce, M. S., Reilly, J. K., Hughes, A. R., Janssen, X., Basterfield, L., & Reilly, J. J. (2018). Timing of the decline in physical activity in childhood and adolescence: Gateshead Millennium Cohort Study. *British Journal of Sports Medicine*, *52*(15), 1002–1006. https://doi.org/10.1136/bjsports-2016-096933

Fox, K. R., & Corbin, C. B. (1989). The Physical Self-Perception Profile: Devlopment and Preliminary Validation. *Journal of Sport and Exercise Psychology*, *11*(4), 408–430. https://doi.org/10.1123/jsep.11.4.408

Gallahue, D. L., Ozmun, J. C., & Goodway, J. D. (2012). *Understanding Motor Development: Infants, Children, Adolescents, Adults* (Edición: 7). McGraw-Hill Education.

Harter, S., & Pike, R. (1984). The pictorial scale of perceived competence and social acceptance for young children. *Child Development*, *55*(6), 1969–1982.

Jago, R., Solomon-Moore, E., Macdonald-Wallis, C., Sebire, S. J., Thompson, J. L., & Lawlor, D. A. (2017). Change in children's physical activity and sedentary time between Year 1 and Year 4 of primary school in the B-PROACT1V cohort. *International Journal of Behavioral Nutrition and Physical Activity*, *14*(1), 33. https://doi.org/10.1186/s12966-017-0492-0

Johnson, T. M., Ridgers, N. D., Hulteen, R. M., Mellecker, R. R., & Barnett, L. M. (2016). Does playing a sports active video game improve young children's ball skill competence? *Journal Of Science And Medicine In Sport / Sports Medicine Australia*, *19*(5), 432–436. https://doi.org/10.1016/j.jsams.2015.05.002

Kang, T., & Chen, T. T. (2007). An Investigation of the Performance of the Generalized SX^2 Item-Fit Index for Polytomous IRT Models. *ACT*, 1–25. JSTOR.

Kelava, A., Werner, C. S., Schermelleh-Engel, K., Moosbrugger, H., Zapf, D., Ma, Y., Cham, H., Aiken, L. S., & West, S. G. (2011). Advanced Nonlinear Latent Variable Modeling: Distribution Analytic LMS and QML Estimators of Interaction and Quadratic Effects. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(3), 465–491. https://doi.org/10.1080/10705511.2011.582408

Kowalski, K. C., Crocker, P. R., & Donen, R. (2004). *The physical activity questionnaire for older children (PAQ-C) and adolescents (PAQ-A) manual.* University of Saskatchewan.

López-Pastor, V. M., Pintor, P., Muros, B., & Webb, G. (2013). Formative assessment strategies and their effect on student performance and on student and tutor workload: The results of research projects undertaken in preparation for greater convergence of universities in Spain within the European Higher Education Area (EHEA). *Journal of Further and Higher Education*, *37*(2), 163–180. https://doi.org/10.1080/0309877X.2011.644780

Maïano, C., Morin, A. J. S., April, J., Barnett, L. M., St-Jean, C., Gagnon, C., & Aimé, A. (2022). Psychometric Properties of the French Version of the Pictorial Scale of Perceived Movement Skill Competence for Young Children (PMSC). *European Review of Applied Psychology*, *72*(2), 100700. https://doi.org/10.1016/j.erap.2021.100700

Martínez-Gómez, D., Haro, V. M. de, Pozo, T., Welk, G. J., Villagra, H. A., Calle, M. E., Marcos, A., & Veiga, O. L. (2009). Fiabilidad y validez del cuestionario de actividad física PAQ-A en adolescentes españoles. *Revista Española de Salud Pública*, *83*(3), 427–439.

Masci, I., Schmidt, M., Marchetti, R., Vannozzi, G., & Pesce, C. (2018). When Children's Perceived and Actual Motor Competence Mismatch: Sport Participation and Gender Differences. *Journal of Motor Learning and Development*, *6*(S2), S440–S460. https://doi.org/10.1123/jmld.2016-0081

McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). Evidence Based on Relations to Other Variables: Bolstering the Empirical Validity Arguments for Constructs. In D. B. McCoach, R. K. Gable, & J. P. Madura (Eds.), *Instrument Development in the Affective Domain: School and Corporate Applications* (pp. 209–248). Springer New York. https://doi.org/10.1007/978-1-4614-7135-6_6

Menescardi, C., Meester, A. D., Morbée, S., Haerens, L., & Estevan, I. (2022). The role of motivation into the conceptual model of motor development in childhood. *Psychology of Sport and Exercise*, *61*, 102188. https://doi.org/10.1016/j.psychsport.2022.102188

Niemistö, D., Barnett, L. M., Cantell, M., Finni, T., Korhonen, E., & Saakslahti, A. (2019). Socioecological correlates of perceived motor competence in 5-to 7-year-old Finnish children. *Scandinavian Journal of Medicine & Science in Sports*, *29*(5), 753–765. https://doi.org/10.1111/sms.13389

Ortega-Benavent, N., Menescardi, C., Romero-Martínez, J., Barnett, L. M., & Estevan, I. (2024). Perceived physical literacy in children and early adolescents: Two valid and reliable versions of the PL-C Quest. *Psychology of Sport and Exercise*, 102694. https://doi.org/10.1016/j.psychsport.2024.102694

Robinson, L. E., Stodden, D. F., Barnett, L. M., Lopes, V. P., Logan, S. W., Rodrigues, L. P., & D'Hondt, E. (2015). Motor Competence and its Effect on Positive Developmental Trajectories of Health. *Sports Medicine (Auckland, N.Z.)*, *45*(9), 1273–1284. https://doi.org/10.1007/s40279-015-0351-6

Sport Australia. (2021). *Physical literacy in children questionnaire user guide*. https://www.sportaus.gov.au/__data/assets/pdf_file/0009/994248/Physical-Literacy-in-Children-Questionnaire-User-Guide.pdf

Stodden, D. F., Goodway, J. D., Langendorfer, S. J., Roberton, M. A., Rudisill, M. E., Garcia, C., & Garcia, L. E. (2008). A Developmental Perspective on the Role of Motor Skill Competence in Physical Activity: An Emergent Relationship. *Quest*, *60*(2), 290–306. https://doi.org/10.1080/00336297.2008.10483582

Tapia-Serrano, M. A., García-Hermoso, A., Sevil-Serrano, J., Sánchez-Oliva, D., & Sánchez-Miguel, P. A. (2022). Is adherence to 24-Hour Movement Guidelines associated with a higher academic achievement among adolescent males and females? *Journal of Science and Medicine in Sport*, *25*(2), 155–161. https://doi.org/10.1016/j.jsams.2021.09.005

Tietjens, M., Dreiskaemper, D., Utesch, T., Schott, N., Barnett, L., & Hinkley, T. (2018). Pictorial Scale of Physical Self-Concept for younger children (P-PSC-C): A Feasibility-Study. *Journal of Motor Learning and Development*, 1–21. https://doi.org/10.1123/jmld.2016-0088

Ulrich, D. A. (2000). *Test of Gross Motor Development-2*. Prod-Ed.

Valentini, N., Barnett, L., Bandeira, P. F., Nobre, G. C., Zanell, L. W., & Sartori, R. F. (2017). The Pictorial Scale of Perceived Movement Skill Competence: Determining Content and Construct Validity for Brazilian Children. *Journal of Motor Learning and Development*, 1–26. https://doi.org/10.1123/jmld.2016-0043

van Veen, C., Schott, N., Lander, N., Tietjens, M., Hinkley, T., Dreiskämper, D., Holfelder, B., Utesch, T., & Barnett, L. M. (2020). The Stability of Perceived Motor Competence of Primary School Children from Two Countries over One Year. *Measurement in Physical Education and Exercise Science*, *24*(1), 74–80. https://doi.org/10.1080/1091367X.2019.1675665

Venetsanou, F., Kossyva, I., Valentini, N., Afthentopoulou, A.-E., & Barnett, L. (2018). Validity and Reliability of the Pictorial Scale of Perceived Movement Skill Competence for Young Greek Children. *Journal of Motor Learning and Development*, *6*(s2), s239–s251. https://doi.org/10.1123/jmld.2017-0028

Webster, E. K., & Ulrich, D. A. (2017). Evaluation of the Psychometric Properties of the Test of Gross Motor Development—Third Edition. *Journal of Motor Learning and Development*, *5*(1), 45–58. https://doi.org/10.1123/jmld.2016-0003

Yannakakis, G. N., & Hallam, J. (2011). Ranking vs. Preference: A Comparative Study of Self-reporting. In S. D'Mello, A. Graesser, B. Schuller, & J.-C. Martin (Eds.), *Affective Computing and Intelligent Interaction* (pp. 437–446). Springer Berlin Heidelberg.

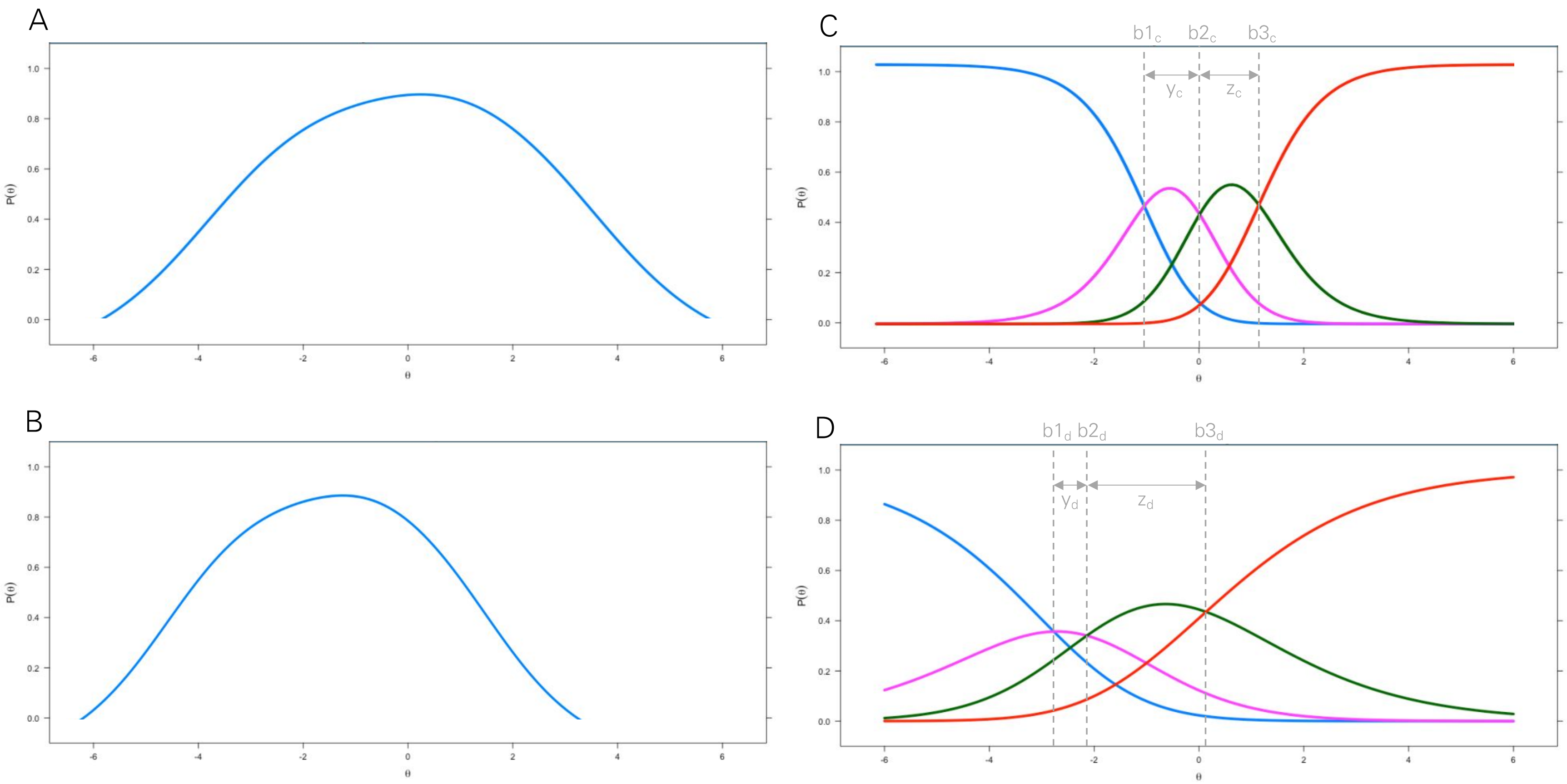

Figure 1. Comparison of two reliability functions with (A) being reliable from low to high values of the scale and (B) being reliable only for the bottom half of the scale. C showcases an ICC with equally distanced thresholds ($y_c = z_c$) and the item is average difficult for $b2_c = 0$. D illustrates that $y_d < z_d$ and the average difficulty is approximately at $b3_d = 0$.

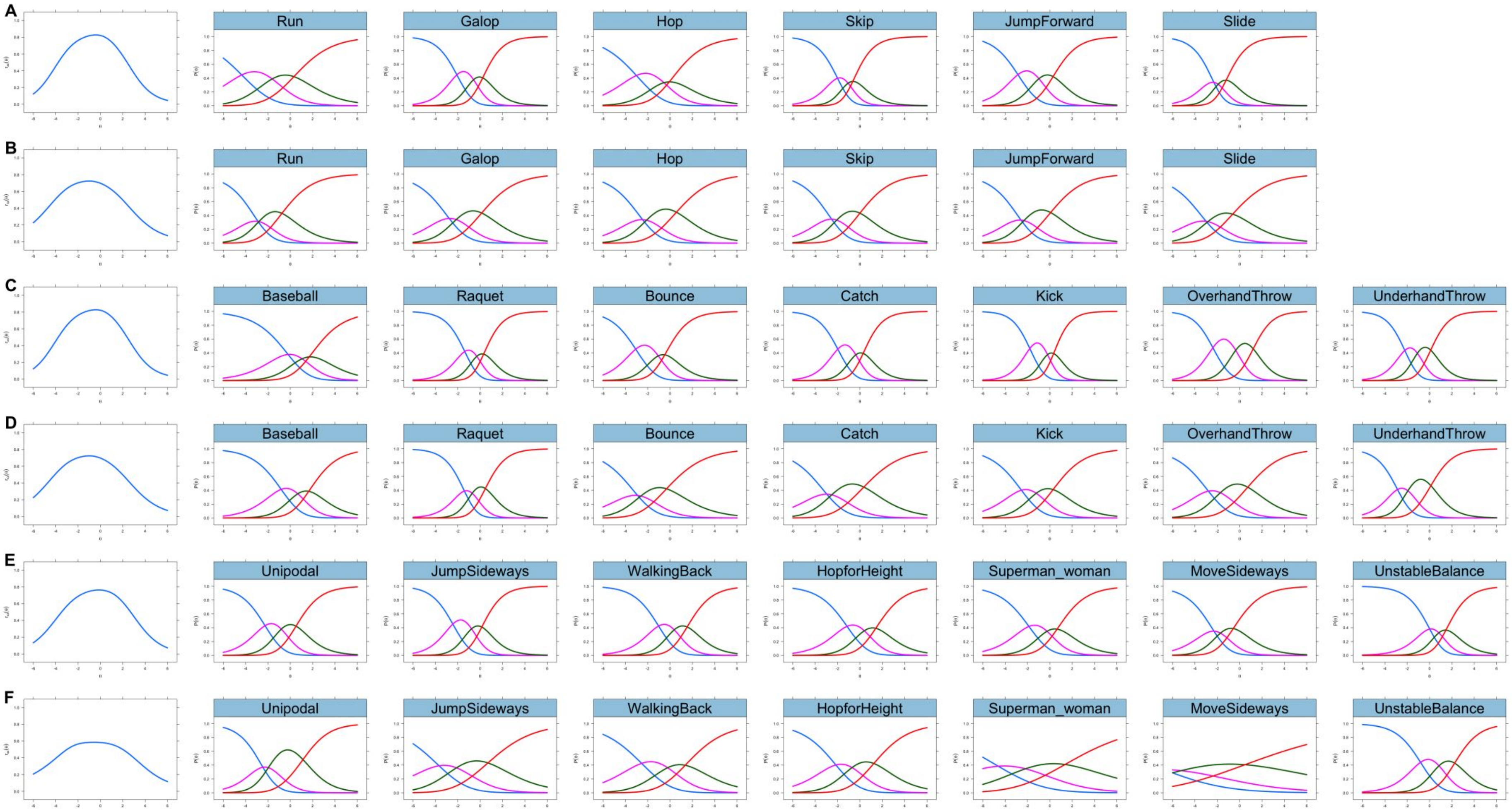

Figure 2. Reliability curve (column 1) and item information curves (ICC) for the single items. (A) refers to the locomotion in supervised group administration while (B) refers to locomotion in one-on-one interviews. (C) refers to the object control in supervised group administration while (D) refers to object control in one-on-one interviews. (E) refers to the stability in supervised group administration while (F) refers to stability in one-on-one interviews.

Table 1. Descriptive (Means and standard errors) and fit statistics of the pictorial scale of Perceived Movement Skill Competence.

	Group administration		One-on-one interview			
Dimension	*M*	*se*	*M*	*se*	*t*	*p*
Locomotion	3.15	0.02	3.25	0.02	3.65	< .001
Object Control	2.82	0.02	2.99	0.02	5.30	< .001
Stability	2.67	0.02	2.81	0.02	4.53	< .001
Fit statistics*						
chi square	400.680		205.814			
df	167		167			
p value	< .001		.022			
CFI	.967		.986			
RMSEA	.047		.019			
SRMR	.057		.042			

Note. * fit statistics of the correlated three factor structure in the confirmatory factor analysis.

Table 2. Evidence of association between perceived motor competence and physical activity according to the level of engagement.

	Low PA		High PA			
	M	*se*	*M*	*se*	*t*	*p*
Locomotion [SGA]	2.91	0.05	3.20	0.02	5.118	< .001
Object Control [SGA]	2.42	0.06	2.89	0.03	7.203	< .001
Stability [SGA]	2.39	0.06	2.73	0.03	5.326	< .001
Locomotion [interview]	3.13	0.04	3.28	0.02	2.975	.003
Object Control [interview]	2.76	0.05	3.03	0.02	5.045	< .001
Stability [interview]	2.70	0.04	2.83	0.02	2.739	.006

Note. SGA means supervised-guided group administration. Interview refers to one-on-one interview. PA means physical activity. Physical activity was categorised based on a cut-off score of 2.75 out of 5. $p < .006$

Table 3. Correlations indicating association between perceived motor competence and physical activity, and concurrent validity.

	Cronbach's alpha	Marginal_rel	Physical Activity		Stability (within administration)	
	α	*rxx*	*r*	*p*	*r*	*p*
Locomotion [SGA]	.67	.66	.310	< .001	.678	< .001
Object Control [SGA]	.79	.78	.378	< .001	.498	< .001
Stability [SGA]	.74	.74	.319	< .001	1	< .001
Locomotion [interview]	.54	.53	.202	< .001	.557	< .001
Object Control [interview]	.66	.67	.302	< .001	.480	< .001
Stability [interview]	.52	.57	.248	< .001	1	< .001

Note. SGA means supervised group administration. Interview refers to one-on-one individual interview. Marginal_rel (*rxx*) refers to a reliability function estimate. Estimates in the column of perceived MC Stability are calculated within the same type of administration.

Table 4. Estimate and threshold parameters modelled within the Generalized Partial Credit Model.

	Group administration				One-on-one interview				Group administration		One-on-one interview	
item	est	b1	b2	b3	est	b1	b2	b3	y	z	y	z
Run	0.54	-4.34	-1.62	0.32	0.65	-2.88	-2.81	-0.75	2.72	1.94	0.07	2.06
Gallop	1.01	-2.08	-0.59	0.23	0.6	-2.77	-2.14	0.13	1.49	0.82	0.63	2.27
Hop	0.57	-3.01	-0.55	-0.07	0.59	-2.46	-2.1	0.57	2.46	0.48	0.36	2.67
Skip	0.95	-2	-1.06	-0.64	0.65	-2.53	-2.08	0	0.94	0.42	0.45	2.08
Jump Forward	0.83	-2.85	-0.97	0.29	0.63	-2.57	-2.31	0.07	1.88	1.26	0.26	2.38
Slide	0.91	-2.31	-1.93	-1.22	0.54	-3	-2.84	-0.63	0.38	0.71	0.16	2.21
Two-hand strike	0.57	-0.2	1.06	1.69	0.69	-0.81	0.67	1.65	1.26	0.63	1.48	0.98
One-hand forehand	1.04	-1.38	-0.3	0.31	1	-1.41	-0.71	0.47	1.08	0.61	0.70	1.18
Catch	105	-1.98	-0.37	0.29	0.55	-2.96	-2.52	0.35	1.61	0.66	0.44	2.87
Bounce	0.83	-3.07	-1.05	-0.47	0.53	-2.9	-2.61	-0.31	2.02	0.58	0.29	2.3
Kick	1.2	-1.79	-0.2	0.34	0.65	-2.54	-1.22	0.3	1.59	0.54	1.32	1.52
Overhand	1.1	-2.4	-0.36	1.12	0.61	-2.8	-1.73	0.72	2.04	1.48	1.07	2.45
Underhand	1.14	-2.22	-1.09	0.1	0.97	-2.87	-1.94	0.12	1.13	1.19	0.93	2.06
Unipodal Balance	0.83	-2.26	-0.8	0.51	0.81	-2.45	-1.96	1.13	1.46	1.31	0.49	3.09
Jump Sideways	0.97	-2.44	-0.77	0.15	0.45	-3.68	-2.12	0.7	1.67	0.92	1.56	2.82
Walking Backward	0.82	-1.03	0.37	1.53	0.51	-2.54	-0.22	1.36	1.40	1.16	2.32	1.58
Hop for Height	0.72	-1.15	0.39	1.42	0.59	-2.15	-0.71	1.24	1.54	1.03	1.44	1.95
Superman_woman	0.68	-1.87	-0.25	0.65	0.28	-4.61	-2.17	1.36	1.62	0.9	2.44	3.53
Move Sideways	0.7	-2.3	-1.72	-0.57	0.18	-6.76	-5.24	0.48	0.58	1.15	1.52	5.72
Unstable Balance	0.87	0.02	0.83	1.46	0.84	-0.81	0.84	2.24	0.81	0.63	1.65	1.40

Note. Est is the estimation of the within item difficulty in the partial credit model. b1, b2 and b3 refers to the first, second and third threshold between the four categories. (y) is the length between the thresholds b1 and b2 while (z) is the length between the thresholds b2 and b3. The positive values of the lengths indicate that all thresholds are ordered.

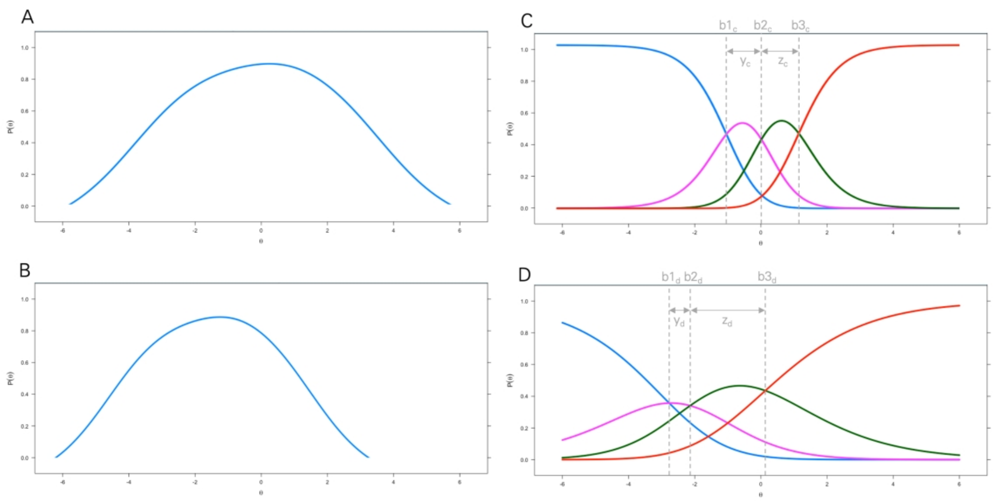

Figure 1. Comparison of two reliability functions with (A) being reliable from low to high values of the scale and (B) being reliable only for the bottom half of the scale. C showcases an ICC with equally distanced thresholds (yc = zc) and the item is average difficult for b2c = 0. D illustrates that yd < zd and the average difficulty is approximately at b3d = 0.

338x190mm (225 x 225 DPI)

Figure 2. Reliability curve (column 1) and item information curves (ICC) for the single items. (A) refers to the locomotion in supervised group administration while (B) refers to locomotion in one-on-one interviews. (C) refers to the object control in supervised group administration while (D) refers to object control in one-on-one interviews. (E) refers to the stability in supervised group administration while (F) refers to stability in one-on-one interviews.

1411x934mm (72 x 72 DPI)

Supplement

Table S1. Estimates of items fit from confirmatory factor analysis

	Group Administration		Individual Interview	
item	X^2	*p*	X2	*p*
Run	41.777	.035	21.921	.188
Galop	33.040	.080	15.726	.675
Hop	15.915	.938	15.323	.640
Skip	12.842	.955	26.068	.128
Jump Forward	27.982	.261	16.417	.563
Slide	19.687	.661	15.934	.597
Baseball	33.159	.362	22.646	.482
Raquet	46.005	.052	26.607	.430
Bounce	42.090	.133	21.974	.821
Catch	36.109	.282	21.605	.868
Kick	34.676	.297	28.939	.416
Overhand Throw	32.607	.340	22.796	.696
Underhand Throw	23.386	.759	18.041	.755
Unipodal	32.189	.457	26.637	.322
Jump Sideways	3.702	.430	21.091	.782
Walking Back	53.979	.009	38.435	.071
Hop for Height	38.427	.237	29.330	.296
Superman / Woman	41.395	.179	18.208	.955
Move Sideways	43.560	.067	32.181	.312
Unstable Balance	26.219	.711	25.931	.255

Note. Items are modeled using signed chi-squared tests. p values < .001 indicate that items do not fit into the scale and would violate model assumptions.