Journal Pre-proof

Psychometric proprieties of the Test of Gross Motor Development–Third Edition in a large sample of Italian children

Daniele Magistro, Giovanni Piumatti, Fabio Carlevaro, Lauren B. Sherar, Dale W. Esliger, Giulia Bardaglio, Francesca Magno, Massimiliano Zecca, Giovanni Musella

PII: S1440-2440(18)30377-3

DOI: https://doi.org/10.1016/j.jsams.2020.02.014

Reference: JSAMS 2268

To appear in: *Journal of Science and Medicine in Sport*

Received Date: 11 July 2018

Revised Date: 16 January 2020

Accepted Date: 23 February 2020

Title: ***Psychometric proprieties of the Test of Gross Motor Development–Third Edition in a large sample of Italian children.***

Running title: Psychometric proprieties of TGMD-3

Authors:

Daniele Magistro[a]; Giovanni Piumatti[b]; Fabio Carlevaro[c]; Lauren B Sherar[d,f]; Dale W Esliger[d,f]; Giulia Bardaglio[c,g]; Francesca Magno[c,g]; Massimiliano Zecca[e,f]; Giovanni Musella[c,g]

Institutions:

Nottingham Trent University

Clifton Lane, Nottingham, NG11 8NS, UK

[a]Department of Sport Science, School of Science and Technology

University of Geneva

24 rue du Général-Dufour, 1211 Genève 4, CH

[b]Unit of Development and Research in Medical Education (UDREM), Medical School (CMU),

[c]*Polo Universitario Asti Studi Superiori* (Uni-Astiss)

Piazzale Fabrizio de Andrè, Asti, 14100, IT

Loughborough University

Epinal Way, Loughborough, LE11 3TU, UK

[d]School of Sport, Exercise, and Health Sciences

[e]Wolfson School of Mechanical, Electrical and Manufacturing Engineering,

[f]*National Centre for Sport and Exercise Medicine* (NCSEM)

Epinal Way, Loughborough, LE11 3TU, UK

[g]*University of Torino*

Via Verdi, 8, Torino, 10124, IT

Corresponding author:

Magistro, Daniele, PhD

Department of Sport Science, School of Science and Technology, Nottingham Trent University

Clifton Lane, Nottingham, NG11 8NS, UK

+44 (0)115 941 8418

daniele.magistro@ntu.ac.uk

Declarations of interest: none.

Word count: 2986

Abstract word count: 247

Table: 2

Figure: 1

Supplementary material:

Table: 5

Figure: 2

Abstract

Objectives: The Test of Gross Motor Development-3 (TGMD-3) evaluates fundamental gross motor skills across two domains: locomotor and ball skills. This study aimed to perform a full psychometric assessment of this test in a large sample of Italian pre- and primary school children.

Design: Cross-sectional and test-retest study design.

Method: Children (N=5,210; Mean age (years)=8.38, SD=1.97; % females=48) completed three trials, including one practice a. Only the scores of the two latter 'formal' trials were recorded for the evaluation. Factorial validity and measurement invariance of TGMD-3 across age and gender groups and test-retest reliability for the overtime measure consistency were tested. Item response theory analysis further tested single items' performances.

Results: Explorative and confirmatory factor analyses confirmed the two-factor structure of the TGMD-3. Multi-group confirmatory factor analyses indicated that there were no significant reductions in model adjustments between the configural, metric and structural invariance solutions for gender and age groups. Test-retest results ranged between 0.967 and 0.990 for both skill sets across age groups. Item response theory analysis using a graded response model showed low standard error and high-test information levels covering a wide spectrum range of both locomotor and ball skills.

Conclusions: These results highlight the strong construct validity and reliability of the TGMD-3 to measure gross motor skills in children across gender and age groups. Item response theory analysis evidenced how the performance criteria included in this test cover a wide range of gross the motor skills spectrum. The use of TGMD-3 may inform motor development programs and support curricular decisions in schools.

Keywords: motor skills, multigroup analysis, item response theory, motor development

Introduction

Gross motor skills are goal-directed movement patterns[1] involving large whole-body movements, locomotion, and whole body stretches.[2] Fundamental motor skills represent a subset of gross motor skills and are the foundation of physical education curricula worldwide.[3] These skills can be broadly divided into two categories: locomotor skills and object control skills. The former include walking, running, hopping, galloping, jumping, sliding, and leaping; the latter include throwing, catching, striking, bouncing, and kicking.[4] An adequate proficiency in gross motor skills is a prerequisite for performing sport-specific skills and for successfully participating in organized and unorganized physical activities.[5–7] An inadequate motor competency or proficiency in gross motor skills is often associated to lower levels of perceived physical competence,[8] self-esteem and social acceptance in late childhood[9,10] and can negatively affect competences and proficiency in physical and motor activities later in life.[11,12]

The assessment of gross motor developmental status among children can provide valuable information to identify possible motor delays and deficits. The appropriate overtime assessment of proficiency and development of these skills depends on the use of reliable and valid instruments.[13,14] The first version of the Test of Gross Motor Development (TGMD)[15] and its second version (TGMD-2)[16] have been widely used to identify motor development delays in children. The TGMD measures 12 gross motor skills focused on both locomotor and object control skills (i.e., ball skills aspects). It is a norm- and criterion-referenced test administered and scored in a consistent manner against a fixed set of predetermined criteria in children from 3 to 10 years.

The TGMD has been recently revised into its current third edition (TGMD-3).[17,18] Although originally developed for populations of children aged 3 to 10/11.[19,20] Similar to the previous versions, this instrument is a process-orientated test of gross motor skills during childhood and has shown good validity and reliability.[17,21–24] However, to the best of the authors' knowledge, no study examined the validity and reliability of TGMD-3 using item response theory analysis (IRT). More specifically, when assessing the psychometric properties of a newly developed instrument in a particular context we are essentially addressing two main issues: to what extent assessments based on this instrument, given under the same conditions, yield the same results[25] (i.e. reliability, namely the consistency with previous studies in difference contents); and how accurately this instrument measures what it intended to measure[26] (i.e. validity). These evaluations are crucial for the understanding of any statistical finding based on multiple-item scales, since the single items that do not operate consistently fail to provide consistent outcomes.[27] Therefore, an examination of TGMD-3' items' performance by the means of IRT can not only better inform us about the validity and reliability of the TGMD-3 but also provide novel information to improve this instrument; for example, by evidencing the items that may provide more information across different levels (lower or higher) of the entire gross-motor skills spectrum among pre- and primary school children. The aim of the present study was thus to conduct a detailed psychometric assessment of the TGMD-3 in a large sample (N=5,210) of Italian pre- and primary school children. Specifically, factorial validity and measurement invariance of the TGMD-3 was tested across age and gender groups using multi-group factorial analysis, test-retest reliability and IRT. In sum, expanding our knowledge on the validity and reliability of this widely adopted instrument to measure gross motor skills among children may suggest best practices for practitioners and professionals for its adoption so to further promote cross-national comparisons on this topic, but also provide evidence for future test development in the motor domain.

Methods

The study involved 37 schools (14 pre-schools and 23 primary schools) located in the north west of Italy. A total of 5,956 children from 275 classes were included in the current analysis. Within the Italian educational system, primary education lasts five years and is preceded by three years of non-compulsory nursery school (or kindergarten). Children can start primary school between the age of 5 (if born between January and March) and 6. Accordingly, in the current sample we had 745 children from kindergarten, 858 from the first grade of primary school, 941 from grade two, 970 from grade three, 997 from grade four and 1,445 from grade five. A written informed consent was obtained from the parents/guardians and a verbal assent from the children. The ethical committee of the University of Turin approved the study (ID 100949).

The TGMD-3 is divided into two sub-scales, the *locomotor skill sub-scale* composed of six skills: run, gallop, hop, horizontal jump, slide (judged on four performance criteria) and skip (judged on three criteria); and the *ball skill sub-scale* (previously named *object control skill* in the TGMD-2) composed of seven skills: one hand forehand strike of self-bounced, kick a stationary ball, overhand throw, underhand throw (judged on four criteria), two-hand strike of a stationary ball (judged on five criteria), one hand stationary dribble and two hand catch (judged on three criteria). Compared with its previous versions, three skills were added to the TGMD-3 (i.e., skip, underhand throw and one-handed forehand strike) and two were removed (i.e., leap and underhand roll).

In order to develop an Italian translation and edited score sheet for the TGMD-3 an independent English specialist translated the original version into Italian while another independent specialist completed a blinded back translation to English. The translation included all the elements of the TGMD-3: skills' name, performance criteria along with the instructions used for a standardized administration. Subsequently, the Italian version was checked again and compared with the original English version by two post-doctoral researchers and two independent professionals (physical activity teacher with MSc) within the field of sport science and education.

The TGMD-3 tests were administered in each school's gym by 16 professionals: 4 sport science researchers, 2 psychologists and 10 physical education professionals. All the examiners participated in a 2-hour session to be instructed about the test protocol. The test was administered during school hours and with the presence of physical education teachers. Children's socio-demographic information were reported by parents through a questionnaire (see Table S2 in the Supplementary Material). To test-retest reliability 450 children were randomly chosen (50 for each age year from 3 to 11 years old) and re-tested again between 7 to 10 days from the first assessment by the same evaluators. The administration of the TGMD-3 followed original authors' recommendations[17] and required approximately 20 minutes per child. Two independent testers, randomly coupled each time (, simultaneously observed and scored in real time each child's performance. The proportion of agreement between the scores recorded by the two testers was over 95%. Cohen's k[28] was used to determine if there was agreement between the two observers for all TGMD-3's criteria (values from 0.8 to 1 indicating excellent, almost perfect agreement). Moreover, the inter-rater reliability for the final score of the TGMD-3 was determined using a two-way random intraclass correlation (absolute agreement) coefficients (values above 0.90 = excellent*)*. At the beginning of each test section an accurate verbal description and demonstration of each skill was carried out by one examiner. Each child completed three trials, one for practice and then two formal. Only the scores of the two formal trials were recorded for the evaluation. Performances were observed and evaluated following the qualitative performance criteria for each TGMD-3 assessment skill: every criterion was scored based on whether it was fulfilled (score awarded = 1) or not (score awarded = 0).[17] Skill – henceforth defined as *items* – were evaluated twice. The total score for each item is given by the sum of both trials. Items' sums were used to calculate the score for the locomotor and ball control skills sub-scales as well as for the overall TGMD-3 scores.[3,17] Accordingly, the maximum score a participant could obtain for the overall gross motor performance was 100,46 for the locomotor sub-scale and 54 the for the ball skills sub-scale.

As a preliminary analytical step, data were examined for accuracy of data entry, missing values and fit between their distributions and the assumption of normality. Little's test was used to determine whether missing data were randomly distributed throughout the dataset rather than revealing a systematic pattern. Non-significant results of this test indicate that the missing values were randomly distributed. Absolute skewness and kurtosis values lower than 3 and 8 respectively were used to indicate when assumption of normal distribution was not met.[29] Finally, univariate outliers were detected computing standardized scores. Cases with standardized scores in excess of 3.29 ($p < 0.001$, two-tailed test) were flagged as potential outliers.[30]

To test the construct validity of the TGMD-3 we used explorative factor analysis (EFA) and confirmatory factor analysis (CFA) techniques. Initially, the EFA was conducted using principal component analysis with promax rotation. Subsequently, a confirmatory factor analysis (CFA) with maximum likelihood (ML) estimation method was conducted. The following indexes were adopted to assess overall model fit: χ^2 goodness-of-fit, root mean square error of approximation (RMSEA; values ≤. 08 considered acceptable), and comparative fit index (CFI; values ≥ 0.90 considered acceptable). In addition, by the means of a multi-group confirmatory factor analysis (MGCFA) with ML estimation, we tested the assumption of TGMD-3 measurement invariance across gender and age groups (3 to 7, 8 to 9 and 10 to 11 years). As pointed out by Brown,[31] although it is possible to test for MGCFA with unequal sample sizes, it may significantly undermine the power to detect violation of invariance as the ratio of the sample size between the compared groups increases. Accordingly, we opted for splitting the sample into age groups of similar sizes. TGMD-3 factorial invariance was tested following a sequence of three hierarchically nested models, in a similar manner to Valentini and colleagues.[23] First, an initial confirmatory analysis tested the factorial model we obtained from the previous EFA and CFA steps. Then, the invariance of the model structure across gender was tested by examining whether the same parameters existed for both groups (*configural invariance*). Next, additional constraints were investigated through factor loadings (*metric invariance*) and covariance between factors (*structural invariance*). Model fit was evaluated using χ^2 goodness-of-fit, RMSEA and CFI. A change of ≥ -0.01 in CFI in addition to a change of ≥ 0.015 in RMSEA indicated non-invariance.[32]

Item response theory analysis (IRT) using a graded response model (GRM) was finally used to test at which levels of each underlying gross motor skill (i.e. locomotor and ball skills) children were most likely to be located when receiving a specific point along a performance criterion or item. Specifically, recognises that a participant's score on a given item is determined by the participant's overall level on the underlying measured trait and by specific item properties. GRM was chosen given the polytomous ordered-response nature of the TGMD-3's items.[33] As an extension of the two-parameter logistic (2-PL) model, GRM is better suited for these types of items where increasing scores represent an increment in the underlying measured trait in comparison with 1-PL or 3-PL models since it does not include an adjustment for guessing, which is inappropriate for what the TGMD-3 is measuring.[34] Applying GRM we were thus able to estimate the discrimination item parameters (DIs), representing the extent to which an item discriminates between different trait levels (higher values indicating a stronger association with the measured construct), and the threshold parameters of each scoring level, representing the latent trait score needed to have a 0.5 probability of scoring on a particular level or higher. Items information functions and test information functions along with the conditional standard error were plotted to graphically assess the fit of the model to the data.

To understand the consistency of the test over time (test-retest reliability), a two-way mixed intraclass correlation (absolute agreement) coefficients (ICC(3,2)) were used.[35] ICC vary between 0 and 1 (below 0.50 = poor; between 0.50 and 0.75 = moderate; between 0.75 and 0.90 = good; above 0.90 = excellent).[35] A coefficient close to 1 indicates high similarity between values from the same group (good reliability) and a coefficient close to zero means that values from the same group are not similar (poor reliability).

Amos (version 20; SPSS, Inc., Chicago, IL) software was used for factorial analyses. Descriptive analyses and IRT were carried out using Stata software (version 15; StataCorp LP, College Station, TX). SPSS (version 20; SPSS, Inc., Chicago, IL) software was used for the inter-rater, intra-rater and test-retest reliability.

Results

Out of the 5,956 participants, data were excluded from 630 because of missing values on any TGMD-3's item. Little's test for data missing completely at random (MCAR) applied to the entire set of TGMD-3's items was not significant (χ^2=248.40, *df*=248, *p*=0.499), indicating that these data were MCAR and supporting listwise deletion. In addition, 106 participants were also deleted because they did not provide their age, while 10 participants were not included in the current analyses because they reported a severe physical disorder that prevented independent mobility or a severe cognitive disorder that prevented full participation in the test (Table S1), leaving a total of 5,210 cases for the analyses (Mean age=8.40, SD=1.98, *Range*=3–11; females=48%) (Table S2). Absolute values of skewness and kurtosis ranged respectively from 0.36 to 1.47 and from 1.87 to 4.38 suggesting that scores on all 13 TGMD-3's items were reasonably normally distributed. No univariate outliers with absolute standardized scores higher than 3.29 were detected. The intra-rater reliability analysis showed results with high *k* magnitude, indeed the Cohen's *k* was always over 0.8 for each criteria of each skills (Table S3). The inter-rater reliability showed strong positive significant results ($p < 0.001$) within couples of examiners for the TGMD-3 total scores (0.973; 95% Confidence Interval: Lower Bound = 0.969 and Upper Bound = 0.977).

An initial EFA using principal component analysis without rotation yielded two factors with eigenvalues greater than 1 (4.97 and 1.24). The first factor explained 38.29% of variance while the second factor accounted for 9.59%. With the exception of the item pertaining to the "Kick a stationary ball" skill, all items' standardized factor loadings ranged from 0.323 to 0.454. Results from a CFA with ML estimation method based on these exploratory analyses confirmed the two-factor structure of

TGMD-3: χ^2 = 916.284, df = 64, p < 0.001, RMSEA = 0.050 (90% Confidence Intervals: 0.048, 0.053), CFI = 0.955. Factor loadings were all significant at p < 0.001 and ranged between 0.583 to 0.671. Reading from Brown,[31] explicit or widely accepted guidelines to define factor loadings as salient or non-salient do not exist. Accordingly, in light of the CFA results and given the very large sample size, we retained the item pertaining to the "Kick a stationary ball" skill as acceptable despite the low factor loading exhibited in the EFA. Furthermore, this item had a standardized factor loading below 0.2 indicating that it was not related to the locomotor skills factor. Table 1 reports descriptive statistics for all TGMD-3 tasks as well as the standardized factor loadings for each skill from the EFA and CFA solutions.

Results from MGCFA indicate that there was no significant reduction in model adjustments between the configural, metric and structural invariance solutions for both gender (Table S4) and age (Table S5) groups. Results from the metric invariance model suggested that all factor loadings were invariant for boys and girls (ΔRMSEA = -0.001), as well as across age groups (ΔRMSEA = 0). Finally, results from the structural invariance model, demonstrated that the two-factor had similar correlations for gender (ΔRMSEA = 0.002) and age groups (ΔRMSEA = 0.006).

Table S6 reports the parameter estimates of the GRM for all the items of the two TGMD-3's sub-scales. The item information functions for the locomotor and ball skills sub-scales are shown respectively in Figures S1 and S3. "Hop" (DI=1.714), "Run" (DI=1.632) and "Horizontal jump" (DI=1.502) best discriminate between children with low and high levels of the locomotor skills, thus providing the highest amount of information. For the ball skills sub-scale, the highest DI were observed for the items measuring "Forehand strike of self-bounced ball" (DI=1.650), "Kick a stationary ball" (DI=1.430) and "Two- hand catch" (DI=1.418).

Figure 1 reports test information functions along with their conditional standard errors for locomotor and ball skills sub-scales. Both sets of items exhibit low standard error and high-test information levels covering a wide range in the spectrum of locomotor and ball skills. These results support the strong discriminative power and reliability of the TGMD-3 as a tool to measure gross motor skills among children. Nevertheless, it is important to underline how standard error tends to increase at higher skills' levels. Thus, the TGMD-3 can better discriminate between performers with low or moderate scores rather than between performers with high or very high scores.

The test-retest reliability showed strong positive significant results between the test and retest for both the sub-scales score (locomotor sub-scale 0.996; ball sub-scale: 0.997) and the TGMD-3 total scores (0.996). These results were also significant for each age group from 3 to 11 years of age showing a range between 0.967 to 0.990 (Table 2).

Discussion

The aim of the present study was to test the validity and reliability of the TGMD-3 in a sample of Italian children aged 3 to 11. Our results highlight the strong construct validity and reliability of the TGMD-3 to measure gross motor skills across gender and age groups. These results expand previous validations by adopting state-of-the-art psychometric techniques such as IRT that evidenced how the performance criteria included in this test cover a wide range of the gross motor skills spectrum. Results of explorative and confirmatory factor analyses confirm the two-factor model structure proposed by Ulrich since the first version of this instrument[15] and are aligned with previous validity studies in different countries and children samples.[14,36,37,23,24,38,39] In particular, skip, strike with one hand, and underhand throw, the new three skills included in this third version of the TGMD, were strongly correlated to their respective sub-scale. Unlike the previous validity studies, it is important to highlight one main difference of our results with regarding to 'the kick a stationary ball' skill. The EFA showed a lower factor loading for this skill in comparison to other ball skills. The reason of this could be related to the specific criteria used to evaluate this skill: 1) Rapid, continuous approach to the ball; 2) Child takes an elongated stride or leap just prior to ball contact; 3) Non- kicking foot placed close to the ball; and 4) Kicks ball with instep or inside of preferred foot (not the toes). The first two criteria of this skill covers the preparation phase to kick the ball that can also be considered 'locomotion' rather than object control. Indeed, all other ball skills are evaluated in a stationary situation where locomotion is not involved. Hence, based on our results, it is possible to infer that it might be better to reconsider the criteria used to evaluate the kick a stationary ball skill. Nevertheless, it is relevant to highlight that the CFA analysis showed an acceptable factor loading for the kick a stationary ball skill.

The high values of ICCs confirmed the consistency of the TGMD-3 measurements, highlighting that the differences observed in the test were mainly due to differences between children. Such reliability results were similar to previous TGMD-3's validation studies in Germany,[24] Finland,[22] Spain[21] and Brazil.[23] In addition, multi-group factorial analyses showed that the proposed TGMD-3's theoretical structure is equally valid across gender and age groups. Only one other study[23] tested this same assumption and showed similar results, albeit in a smaller sample of children. These findings have important implications for professionals and practitioners, such as school teachers who are keen to design inclusive assessments of physical abilities in schools. The identification of delayed gross motor skill development in children might suggest a need for adaptive interventions to mitigate the limitations which may prevent lifelong involvement in physical activities.

IRT analyses revealed that hop, run and horizontal jump best discriminated between children with lower and higher levels of locomotor skills. Regarding ball skills, strike of self-bounced ball, kick of a stationary ball, and two- hand catch were the best discriminators. Despite the fact that both TGMD-3's sub-scales showed strong discriminative power to measure gross motor skills among children, IRT results can still help us to evidence where TGMD-3 may be improved and how future assessments should be carried out. Specifically, we noticed how single skills included in the test were most helpful in discriminating between children with different levels of motor skills proficiency in particular among children with average or lower levels of both locomotor and ball skills. This means that when using TGMD-3's scores, it becomes more difficult to differentiate two children who are both highly proficient or skilled. Indeed, since its original version the TGMD was developed for identifying developmental delays during childhood,[15] therefore including tasks to specifically differentiate low skilled performers. On a related note, some skills may not add substantial information over and above other skills. This was particularly evident when looking at the performed tasks for the ball skills sub-scale where different skills overlap providing the same type of information. Future elaboration of this instrument may thus consider reducing the number of single skills or substitute some of them with more difficult tasks. This may in turn improve the ability for the test to detect children with real developmental delays across a wider range of locomotor and ball skills and not just at the lower end of gross motor skills' spectrum. For example, the six most discriminative performance tasks observed in the present analyses (i.e. hop, run, horizontal jump, strike of self-bounced ball, kick of a stationary ball and two- hand catch) may be adopted to develop a shorter screening instrument encompassing both locomotor and ball skills, similarly to what Valentini et al. showed for the TGMD-2.[40] Further research is nevertheless needed to specifically assess these measurement aspects by the means of further psychometric analyses such as IRT; for example, focusing on the performance at single criteria composing each single skill.

There are a number of limitations than should be discussed. First, concurrent validity and longitudinal construct stability were not investigated. Future research should examine both these aspects. Moreover, to enhance the generalizability of these findings and create normative data for the TGMD-3, future studies should recruit a larger sample of children from each age group, ethnicity and geographical areas. Indeed, in the current study the number of 3-year-old children was relative to small compare to the other age groups. Moreover, the use of more objective measures/approaches such as video analysis and wearable sensors, in combination with the direct observation assessment of the TGMD-3, might improve our ability to evaluate these fundamental motor skills in routine assessment. Another limitation is that the testers of this study were all experts in sport science and motor development, while practitioners, such as school teachers and coaches, were not involved. Considering that the varied application of TGMD-3 we should evaluate the validity and reliability of the TGMD-3 when administered by an experts and non-experts. Nevertheless, this work confirms that the TGMD-3 is a valid and reliable measure to appropriately assess gross motor skills in children and that its assessment scores are statistically related to performance data.

Conclusion

The TGMD-3 is a reliable instrument to assess gross motor skills in pre- and primary school children. This instrument can be simultaneously adopted across boys and girls with age range from 3 to 11 years. The use of TGMD-3 may inform motor development programs and support curricular decisions in schools. In sum, our results have significant implications in educational settings where it is important to verify in a reliable way motor development among children. This is crucial if we consider how important childhood is for motor development.[1] The TGMD-3 may thus be a very useful instrument for researchers and teachers to identify children's motor delays and motor competence in gross motor skill development.

Practical Implications

- The TGMD-3 is a reliable instrument to assess gross motor skills in pre- and primary school children.
- This instrument can be simultaneously adopted across boys and girls with age ranging from 3 to 11 years old.
- Some physical ability tasks (hop, run, horizontal jump, strike of self-bounced ball, kick of a stationary ball and two- hand catch) in the TGMD-3 are more useful than others to differentiate between children with different degree of ability in motor skills.

Acknowledgements

Blinded for review

Acknowledgements

This study is part of a three-year longitudinal project called "Benessere in Gioco". This work was supported by the Fondo Assistenza e Benessere S.M.S (FAB), Fondazione Cassa di Risparmio di Asti, Polo Universitario Asti Studi Superiori (UNI-Astiss) and Citta` di Asti. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the article, and in the decision to publish the results. The authors would like to express their gratitude to Professor Giuseppe Alloatti of Torino University for the support. The authors would like to express their gratitude to all the children, teachers, and schools for their participation.

Competing interests: The authors declared that no competing interests exist.

References

1 Payne VG, Isaacs LD. *Human motor development: A lifespan approach*. Routledge, 2017.

2 Woodfield L. *Physical development in the early years*. Bloomsbury Publishing, 2004.

3 Allen KA, Bredero B, Van Damme T, et al. Test of Gross Motor Development-3 (TGMD-3) with the Use of Visual Supports for Children with Autism Spectrum Disorder: Validity and Reliability. *J Autism Dev Disord* 2017; 47(3):813–833. Doi: 10.1007/s10803-016-3005-0.

4 Burton AW, Miller DE. *Movement skill assessment*. Human Kinetics, 1998.

5 Karabourniotis D, Evaggelinou C, Tzetzis G, et al. Curriculum Enrichment with Self-Testing Activities in Development of Fundamental Movement Skills of First-Grade Children in Greece. *Percept Mot Skills* 2002; 94(3_suppl):1259–1270. Doi: 10.2466/pms.2002.94.3c.1259.

6 Magistro D, Bardaglio G, Rabaglietti E. Gross Motor Skills and Academic Achievement in Typically Developing Children: The Mediating Effect of Adhd Related Behaviours. *Cognitie, Creier, Comportament/Cognition, Brain, Behavior* 2015; 19(2).

7 Piek JP, Dawson L, Smith LM, et al. The role of early fine and gross motor development on later motor and cognitive ability. *Human Movement Science* 2008; 27(5):668–681.

8 Robinson LE, Rudisill ME, Goodway JD. Instructional climates in preschool children who are at-risk. Part II: Perceived physical competence. *Research Quarterly for Exercise and Sport* 2009; 80(3):543–551.

9 Skinner RA, Piek JP. Psychosocial implications of poor motor coordination in children and adolescents. *Human Movement Science* 2001; 20(1):73–94.

10 Valentini NC, Rudisill ME. Motivational climate, motor-skill development, and perceived competence: Two studies of developmentally delayed kindergarten children. *Journal of Teaching in Physical Education* 2004; 23(3):216–234.

11 Stodden DF, Goodway JD, Langendorfer SJ, et al. A developmental perspective on the role of motor skill competence in physical activity: An emergent relationship. *Quest* 2008; 60(2):290–306.

12 Hulteen RM, Morgan PJ, Barnett LM, et al. Development of foundational movement skills: A conceptual model for physical activity across the lifespan. *Sports Medicine* 2018; 48(7):1533–1540.

13 Netelenbos JB. Teachers' ratings of gross motor skills suffer from low concurrent validity. *Human Movement Science* 2005; 24(1):116–137.

14 Valentini NC. Validity and reliability of the TGMD-2 for Brazilian children. *Journal of Motor Behavior* 2012; 44(4):275–280.

15 Ulrich DA, Sanford CB. *Test of gross motor development*. Pro-ed Austin, TX, 1985.

16 Ulrich DA. TGMD-2. *Test of Gross Motor Development Examiner's Manual* 2000.

17 Webster EK, Ulrich DA. Evaluation of the Psychometric Properties of the Test of Gross Motor Development. Third Edition. *Journal of Motor Learning and Development* 2017; 5(1):45–58. Doi: 10.1123/jmld.2016-0003.

18 Ulrich DA. Introduction to the Special Section: Evaluation of the Psychometric Properties of the TGMD-3. *Journal of Motor Learning and Development* 2017; 5(1):1–4. Doi: 10.1123/jmld.2017-0020.

19 Brian A, Taunton S, Lieberman LJ, et al. Psychometric Properties of the Test of Gross Motor Development-3 for Children With Visual Impairments. *Adapted Physical Activity Quarterly* 2018; 35(2):145–158.

20 Burns R, Brusseau T, Hannon J. Multivariate Associations Among Health-Related Fitness, Physical Activity, and TGMD-3 Test Items in Disadvantaged Children From Low-Income Families. *Percept Mot Skills* 2017; 124(1):86–104. Doi: 10.1177/0031512516672118.

21 Estevan I, Molina-García J, Queralt A, et al. Validity and Reliability of the Spanish Version of the Test of Gross Motor Development-3. *Journal of Motor Learning and Development* 2017:1–21.

22 Rintala PO, Sääkslahti AK, Iivonen S. Reliability Assessment of Scores From Video-Recorded TGMD-3 Performances. *Journal of Motor Learning and Development* 2017; 5(1):59–68. Doi: 10.1123/jmld.2016-0007.

23 Valentini NC, Zanella LW, Webster EK. Test of Gross Motor Development—Third Edition: Establishing Content and Construct Validity for Brazilian Children. *Journal of Motor Learning and Development* 2016; 5(1):15–28. Doi: 10.1123/jmld.2016-0002.

24 Wagner MO, Webster EK, Ulrich DA. Psychometric Properties of the Test of Gross Motor Development 3rd Edition (German translation)–Results of a Pilot-Study. *Journal of Motor Learning and Development* 2016:1–27.

25 Huck SW, Cormier WH, Bounds WG. *Reading statistics and research*. Harper & Row New York, 1974.

26 Tavakol M, Dennick R. Making sense of Cronbach's alpha. *International Journal of Medical Education* 2011; 2:53.

27 Pickett AC, Valdez D, Barry AE. Psychometrics matter in health behavior: a long-term reliability generalization study. *American Journal of Health Behavior* 2017; 41(5):544–552.

28 Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005; 85(3):257–268.

29 Kline RB. *Principles and practice of structural equation modeling*. London, Guilford publications, 2015.

30 Tabachnick BG, Fidell LS. *Using multivariate statistics*. Allyn & Bacon/Pearson Education, 2007.

31 Brown TA. *Confirmatory factor analysis for applied research - Second edition*. New York, NY, Guilford Publications, 2015.

32 Davidov E, Schmidt P, Billiet J. *Cross-cultural analysis: Methods and applications*. Routledge, 2011.

33 Samejima F. Graded Response Model. *Encyclopedia of Social Measurement*. New York, 2004.

34 Chang C-C, Lin C-Y, Gronholm PC, et al. Cross-validation of two commonly used self-stigma measures, Taiwan versions of the Internalized Stigma Mental Illness Scale and Self-Stigma Scale–Short, for people with mental illness. *Assessment* 2016:1073191116658547.

35 Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016; 15(2):155–163. Doi: 10.1016/j.jcm.2016.02.012.

36 Farrokhi A, Zadeh Z, Kazemnejad A, et al. Reliability and validity of test of gross motor development-2 (Ulrich, 2000) among 3-10 aged children of Tehran City. *Journal of Physical Education and Sport Management* 2014; 5(2):18–28.

37 Kim S, Kim MJ, Valentini NC, et al. Validity and reliability of the TGMD-2 for South Korean children. *Journal of Motor Behavior* 2014; 46(5):351–356.

38 Capio CM, Eguia KF, Simons J. Test of gross motor development-2 for Filipino children with intellectual disability: validity and reliability. *Journal of Sports Sciences* 2016; 34(1):10–17.

39 Magistro D, Piumatti G, Carlevaro F, et al. Measurement invariance of TGMD-3 in children with and without mental and behavioral disorders. *Psychological Assessment* 2018; 30(11):1421–1429. Doi: 10.1037/pas0000587.

40 Valentini NC, Rudisill ME, Bandeira PFR, et al. The development of a short form of the Test of Gross Motor Development-2 in Brazilian children: Validity and reliability. *Child: Care, Health and Development* 2018; 44(5):759–765. Doi: 10.1111/cch.12598.

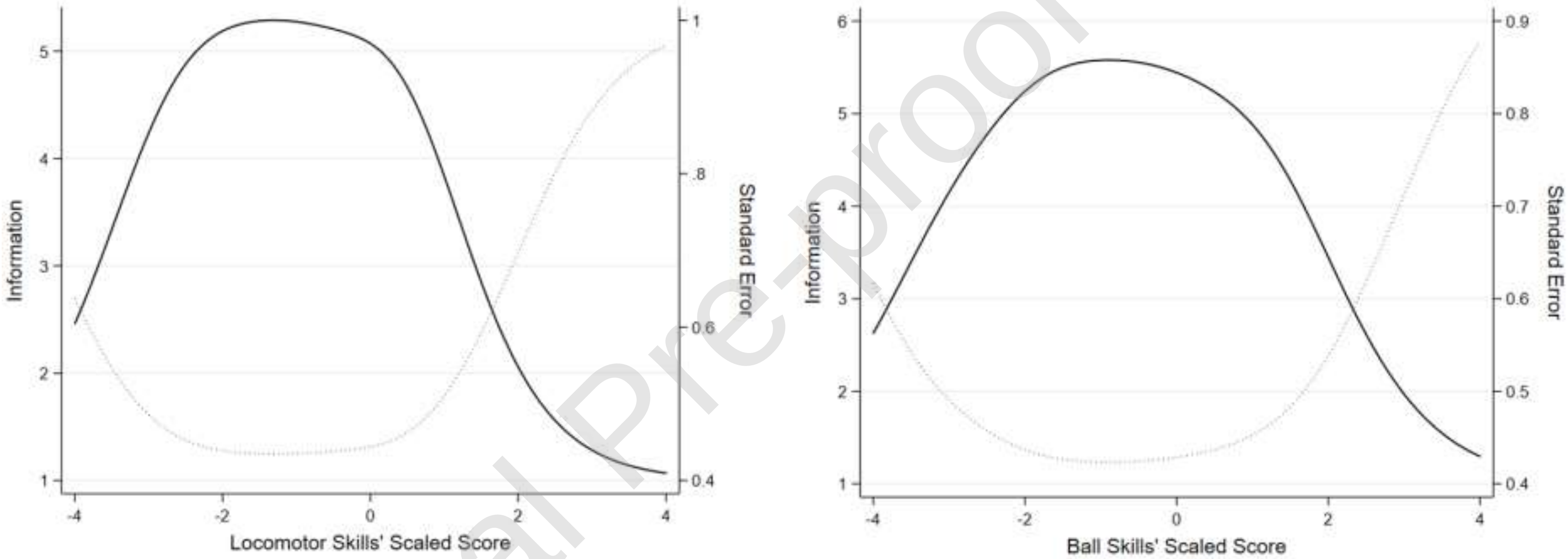
Table 1. Descriptive statistics and standardized results of explorative and confirmatory factor analyses for all TGMD-3 tasks (N = 5,210)

Task	Number of performance criteria	Range score	M	SD	Factor 1 Locomotor skills		Factor 2 Ball skills	
					EFA β	CFA β	EFA β	CFA β
Locomotor skills		0–47	34.48	9.13				
Run	4	0–8	6.35	2.00	0.323	0.671	-	-
Gallop	4	0–8	5.82	2.22	0.363	0.615	-	-
Hop	4	0–8	5.88	2.21	0.405	0.675	-	-
Skip	3	0–6	4.10	2.10	0.426	0.584	-	-
Horizontal jump	4	0–8	5.98	2.32	0.426	0.622	-	-
Slide	4	0–8	6.35	2.17	0.454	0.585	-	-
Ball skills		0–56	34.99	10.55				
Forehand strike of self-bounced ball	4	0–8	4.51	2.54	-	-	0.387	0.565
One-hand stationary dribble	3	0–6	3.66	2.15	-	-	0.433	0.656
Two- hand catch	3	0–6	4.87	1.61	-	-	0.374	0.604
Kick a stationary ball	4	0–8	5.32	2.31	-	-	0.244	0.629

Overhand throw	4	0–8	4.95	2.41	-	-	0.421	0.603
Underhand throw	4	0–8	5.48	2.05	-	-	0.353	0.589
Two-hand strike of a stationary ball	5	0–10	6.17	2.49	-	-	0.376	0.597

Notes. TGMD-3: Test of Gross Motor Development – 3rd edition. EFA = exploratory factor analysis. CFA = confirmatory factor analysis. The explorative factor solution adopted principal component analysis with promax rotation. The confirmatory solution adopted maximum likelihood estimation method. Factor loadings < 0.20 were omitted.

Figure 1.



Notes. TGMD-3: Test of Gross Motor Development – 3rd edition.

Table 2. Intraclass correlation coefficients (ICC) and confidence intervals (CI)

		ICC	95% CI	
			Lower Bound	Upper Bound
Total sample (*n* = 450)	Locomotor skills sub-scale	0.996	0.996	0.997
	Ball skills sub-scale	0.997	0.996	0.997
	TGMD-3 total	0.996	0.996	0.997
3 years old group (*n* = 50)	Locomotor skills sub-scale	0.992	0.986	0.996
	Ball skills sub-scale	0.990	0.983	0.994
	TGMD-3 total	0.991	0.984	0.995
4 years old group (*n* = 50)	Locomotor skills sub-scale	0.994	0.990	0.997
	Ball skills sub-scale	0.991	0.984	0.995
	TGMD-3 total	0.993	0.998	0.996
5 years old group (*n* = 50)	Locomotor skills sub-scale	0.994	0.989	0.996
	Ball skills sub-scale	0.994	0.990	0.997
	TGMD-3 total	0.992	0.985	0.995
6 years old group (*n* = 50)	Locomotor skills sub-scale	0.993	0.987	0.996
	Ball skills sub-scale	0.992	0.986	0.995
	TGMD-3 total	0.991	0.983	0.995
7 years old group (*n* = 50)	Locomotor skills sub-scale	0.983	0.971	0.990
	Ball skills sub-scale	0.989	0.981	0.984
	TGMD-3 total	0.979	0.964	0.988
8 years old group (*n* = 50)	Locomotor skills sub-scale	0.985	0.974	0.992
	Ball skills sub-scale	0.993	0.987	0.996
	TGMD-3 total	0.981	0.967	0.989
9 years old group (*n* = 50)	Locomotor skills sub-scale	0.991	0.985	0.995
	Ball skills sub-scale	0.995	0.991	0.997
	TGMD-3 total	0.989	0.980	0.993

10 years old group (*n* = 50)	Locomotor skills sub-scale	0.990	0.983	0.994
	Ball skills sub-scale	0.996	0.993	0.998
	TGMD-3 total	0.993	0.987	0.996
11 years old group (*n* = 50)	Locomotor skills sub-scale	0.982	0.968	0.990
	Ball skills sub-scale	0.994	0.989	0.996
	TGMD-3 total	0.984	0.972	0.991

Figure Legend

Figure 1. Information graphs showing the TGMD-3’s locomotor and ball skills sub-scales: Test information function (solid line) and conditional standard error curve (dotted line) (N = 5,210)

Psychometric proprieties of the Test of Gross Motor Development–Third Edition in a large sample of Italian children

SUPPLEMENTARY MATERIAL

List of contents:

Table S1. Excluded physical and severe cognitive disorders

Category	Frequency
Specific language impairment and Motor Impairments	1
DX leg injury	1
SX foot malformation	1
Dwarfism	1
Cognitive and Motor Impairments	1
Down syndrome	2
Hemiplegia and deafness	1
Cornelia de Lange syndrome	1
Möbius syndrome	1

Table S2. Socio-demographic descriptive statistics for the final sample (N = 5,210). Values are frequencies (percentages) unless stated otherwise

Variable	
Gender	
Females	2,496 (48%)
Males	2,714 (52%)
Age *Mean years*	8.37 (SD 1.97)
Age groups	
3 years	50 (1%)
4 years	139 (3%)
5 years	270 (5%)
6 years	495 (10%)
7 years	770 (15%)
8 years	799 (15%)
9 years	863 (16%)
10 years	1,001 (19%)
11 years	823 (16%)
Nationality	
Italian	4,988 (96%)
Other	213 (4%)
Residence	
Urban	3,910 (80%)
Rural	987 (20%)
Weight *Mean kg*	32.22 (SD 10.25)
Height *Mean cm*	131.17 (SD 12.39)
Body mass index, *Mean* kg/m^2	19.38 (SD 3.97)

Table S3. Cohen's *k* for all skills' criteria of the TGMD-3

Sub-scale	Skill	Criterion	*Cohen's k*
Locomotor	Run	1	0.877
		2	0.883
		3	0.867
		4	0.869
	Gallop	1	0.886
		2	0.875
		3	0.837
		4	0.849
	Hop	1	0.837
		2	0.843
		3	0.838
		4	0.818
	Skip	1	0.849
		2	0.856
		3	0.846
	Horizontal jump	1	0.813
		2	0.816
		3	0.814
		4	0.826
	Slide	1	0.883
		2	0.837
		3	0.823
		4	0.831
Ball	Two-hand strike of a stationary ball	1	0.853
		2	0.884

		3	0.813
		4	0.845
		5	0.813
	Forehand strike of self-bounced ball	1	0.811
		2	0.883
		3	0.846
		4	0.811
	One-hand stationary dribble	1	0.807
		2	0.812
		3	0.814
	Two- hand catch	1	0.802
		2	0.829
		3	0.807
	Kick a stationary ball	1	0.811
		2	0.800
		3	0.802
		4	0.815
	Overhand throw	1	0.830
		2	0.864
		3	0.848
		4	0.860
	Underhand throw	1	0.877
		2	0.869
		3	0.853
		4	0.834

Table S4. Results of measurement invariance testing across gender groups

Models	χ^2	$\Delta\chi^2$	*df*	Δdf	RMSEA (90% CIs)	ΔRMSEA	CFI	ΔCFI	Comparison
Gender single-group solutions									
Males (n = 2,714)	484.604	-	64	-	0.049 (0.045, 0.053)	-	0.960	-	-
Females (n = 2,496)	422.551	-	64	-	0.047 (0.043, 0.052)	-	0.957	-	-
Model 1. Configural invariance	907.154	-	128	-	0.034 (0.032, 0.036)	-	0.959	-	-
Model 2. Metric invariance	941.736	34.582^{***}	139	11	0.033 (0.031, 0.035)	-0.001	0.958	-0.001	Model 2 vs. Model 1
Model 3. Structural invariance	953.810	12.074^{***}	140	1	0.033 (0.031, 0.035)	0	0.957	-0.001	Model 3 vs. Model 2

Notes. χ^2: Chi-square goodness of fit; *df*: degrees of freedom; RMSEA: Root Mean Square Error of Approximation; 90% CIs: 90% Confidence Intervals for RMSEA; CFI: Comparative Fit Index; $\Delta\chi^2$: Chi-square goodness of fit difference; Δdf: degrees of freedom difference; ΔCFI: CFI difference; ΔRMSEA: RMSEA difference.

$^{**}p < 0.01$; $^{***}p < 0.001$.

Table S5. Results of measurement invariance testing across age groups

Models	χ^2	$\Delta\chi^2$	*df*	Δdf	RMSEA (90% CIs)	ΔRMSEA	CFI	ΔCFI	Comparison
Age single-group solutions									
3 to 7 years old (n = 1724)	365.447	-	64	-	0.052 (0.047, 0.058)	-	0.955	-	-
8 to 9 years old (n = 1662)	254.080	-	64	-	0.042 (0.037, 0.048)	-	0.932	-	-
10 to 11 years old (n = 1824)	475.831	-	64	-	0.059 (0.054, 0.064)	-	0.851	-	-
Model 1. Configural invariance	1095.353	-	192	-	0.030 (0.028, 0.032)	-	0.926	-	-
Model 2. Metric invariance	1346.423	251.07***	214	22	0.032 (0.030, 0.034)	0.002	0.907	-0.019	Model 2 vs. Model 1
Model 3. Structural invariance	1865.779	519.356***	216	2	0.038 (0.037, 0.040)	0.006	0.865	-0.042	Model 3 vs. Model 2

Notes. χ^2: Chi-square goodness of fit; *df*: degrees of freedom; RMSEA: Root Mean Square Error of Approximation; 90% CIs: 90% Confidence Intervals for RMSEA; CFI: Comparative Fit Index; $\Delta\chi^2$: Chi-square goodness of fi difference; Δdf: degrees of freedom difference; ΔCFI: CFI difference; ΔRMSEA: RMSEA difference.

$^{**}p < 0.01$; $^{***}p < 0.001$.

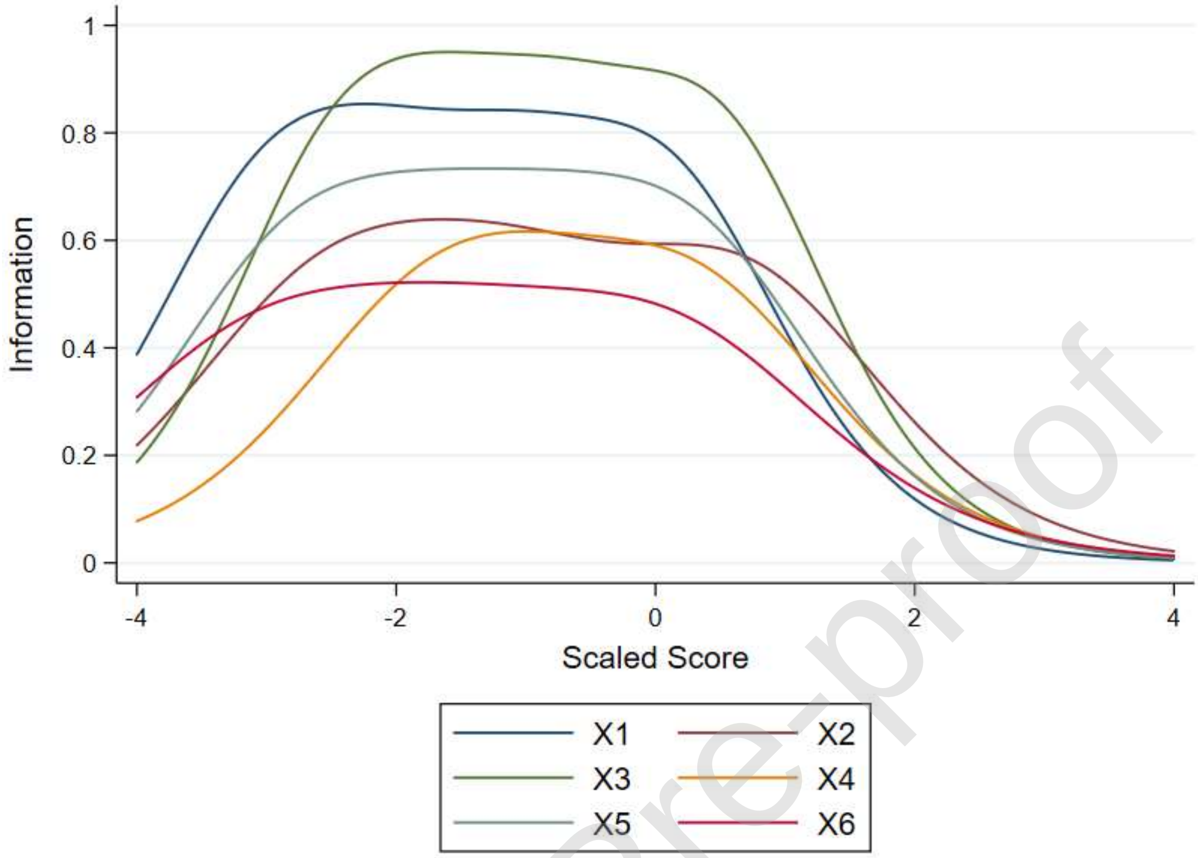
Table S6. Estimated GRM parameters of TGMD-3's sub-scales tasks (N = 5,210; standard errors in parentheses)

Sub-scale	Task	Range score	DI	≥1	≥2	≥3	≥4	≥5	≥6	≥7	≥8	≥9	=10
Locomotor	Run	0–8	1.635 (0.052)	-3.05 (0.08)	-2.93 (0.08)	-2.40 (0.06)	-2.06 (0.05)	-1.17 (0.04)	-0.96 (0.03)	-0.05 (0.02)	0.15 (0.02)	-	-
	Gallop	0–8	1.404 (0.045)	-2.62 (0.07)	-2.42 (0.07)	-1.97 (0.06)	-1.70 (0.05)	-1.17 (0.04)	-0.90 (0.03)	0.44 (0.03)	0.80 (0.03)	-	-
	Hop	0–8	1.717 (0.053)	-2.47 (0.06)	-2.34 (0.06)	-1.91 (0.05)	-1.60 (0.04)	-0.90 (0.03)	-0.60 (0.03)	0.28 (0.02)	0.57 (0.03)	-	-
	Skip	0–6	1.400 (0.047)	-1.75 (0.05)	-1.58 (0.05)	-1.18 (0.04)	-0.98 (0.04)	0.14 (0.03)	0.36 (0.03)	-	-	-	-
	Horizontal jump	0–8	1.503 (0.047)	-2.82 (0.08)	-2.60 (0.07)	-1.91 (0.05)	-1.56 (0.04)	-0.90 (0.03)	-0.70 (0.03)	-0.01 (0.03)	0.35 (0.03)	-	-
	Slide	0–8	1.266 (0.043)	-3.13 (0.10)	-2.91 (0.09)	-2.21 (0.07)	-2.00 (0.06)	-1.53 (0.05)	-1.27 (0.04)	-0.30 (0.03)	0.23 (0.03)	-	-
Ball	Two-hand strike of a stationary ball	0–10	1.242 (0.039)	-3.29 (0.10)	-2.90 (0.09)	-2.31 (0.07)	-1.83 (0.06)	-1.16 (0.04)	-0.59 (0.03)	0.24 (0.03)	0.74 (0.03)	1.63 (0.05)	2.13 (0.06)

	Forehand strike of self-bounced ball	0–8	1.652 (0.049)	-1.75 (0.05)	-1.44 (0.04)	-0.97 (0.03)	-0.64 (0.03)	-0.10 (0.02)	0.30 (0.02)	1.04 (0.03)	1.48 (0.04)	-	-
	One-hand stationary dribble	0–6	1.417 (0.045)	-1.64 (0.05)	-1.32 (0.04)	-0.72 (0.03)	-0.36 (0.03)	0.42 (0.03)	0.72 (0.03)	-	-	-	-
	Two- hand catch	0–6	1.416 (0.047)	-2.98 (0.09)	-2.72 (0.08)	-2.00 (0.06)	-1.45 (0.05)	-0.63 (0.03)	-0.18 (0.03)	-	-	-	-
	Kick a stationary ball	0–8	1.430 (0.043)	-3.09 (0.09)	-2.55 (0.07)	-1.58 (0.05)	-1.19 (0.04)	-0.50 (0.03)	-0.17 (0.03)	0.65 (0.03)	0.93 (0.03)	-	-
	Overhand throw	0–8	1.376 (0.043)	-2.45 (0.07)	-2.04 (0.06)	-1.36 (0.04)	-1.06 (0.04)	-0.21 (0.03)	0.05 (0.03)	0.97 (0.04)	1.25 (0.04)	-	-
	Underhand throw	0–8	1.408 (0.043)	-3.02 (0.09)	-2.74 (0.08)	-2.09 (0.06)	-1.63 (0.05)	-0.75 (0.03)	-0.37 (0.03)	0.90 (0.03)	1.25 (0.04)	-	-

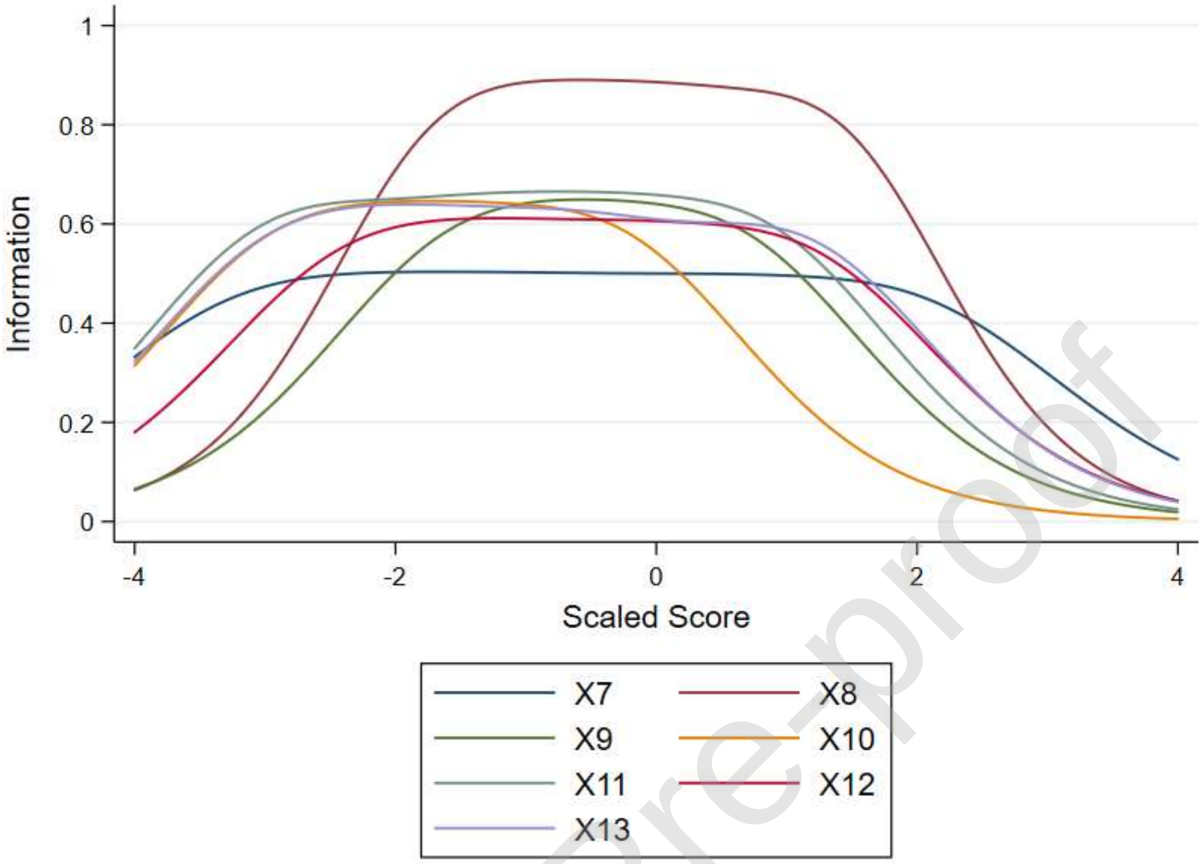
Notes. GRM: Graded Response Model. TGMD-3: Test of Gross Motor Development – 3rd edition. DIF: Discrimination Item parameter.

Figure S1. Item information graph for graded response IRT analysis of the performance criteria in the TGMD-3 pertaining to locomotor skills (N = 5,210)



Notes. IRT: Item Response Analysis. TGMD-3: Test of Gross Motor Development – 3rd edition. X1: Run; X2: Gallop; X3: Hop; X4: Skip; X5: Horizontal jump; X6: Slide.

Figure S2. Item information graph for graded response IRT analysis of the performance criteria in the TGMD-3 pertaining to ball skills (N = 5,210)



Notes. IRT: Item Response Analysis. TGMD-3: Test of Gross Motor Development – 3rd edition. X7: Two- hand strike of a stationary ball; X8: One- hand forehand strike of self-bounced ball; X9: One-hand stationary dribble; X10: Two- hand catch; X11: Kick a stationary ball; X12: Overhand throw; X13: Underhand throw.