Measurement invariance of TGMD-3 in children with and without mental and behavioral disorders.

Daniele Magistro,

School of Sport, Exercise, and Health Sciences, Loughborough University

National Centre for Sport and Exercise Medicine (NCSEM) - Loughborough

Giovanni Piumatti

Unit of Development and Research in Medical Education (UDREM) Medical School(CMU), University of Geneva

Fabio Carlevaro

Polo Universitario Asti Studi Superiori (Uni-Astiss), Asti

Lauren B. Sherar

School of Sport, Exercise, and Health Sciences, Loughborough University

National Centre for Sport and Exercise Medicine (NCSEM) - Loughborough

Dale W. Esliger

School of Sport, Exercise, and Health Sciences, Loughborough University

National Centre for Sport and Exercise Medicine (NCSEM) - Loughborough

Giulia Bardaglio

Polo Universitario Asti Studi Superiori (Uni-Astiss)

University of Torino

Francesca Magno

Polo Universitario Asti Studi Superiori (Uni-Astiss)

University of Torino

Massimiliano Zecca

Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University

National Centre for Sport and Exercise Medicine (NCSEM) - Loughborough

Giovanni Musella

Polo Universitario Asti Studi Superiori (Uni-Astiss)

University of Torino

Author note

Daniele Magistro, School of Sport, Exercise, and Health Sciences, Loughborough University and National Centre for Sport and Exercise Medicine (NCSEM) – Loughborough; Giovanni Piumatti, Unit of Development and Research in Medical Education (UDREM) Medical School (CMU), University of Geneva; Fabio Carlevaro, Polo Universitario Asti Studi Superiori (Uni-Astiss), Asti; Lauren B. Sherar, School of Sport, Exercise, and Health Sciences, Loughborough University and National Centre for Sport and Exercise Medicine (NCSEM) – Loughborough; Dale W. Esliger, School of Sport, Exercise, and Health Sciences, Loughborough University and National Centre for Sport and Exercise Medicine (NCSEM) – Loughborough; Giulia Bardaglio, Polo Universitario Asti Studi Superiori (Uni-Astiss) and University of Torino; Francesca Magno, Polo Universitario Asti Studi Superiori (Uni-Astiss) and University of Torino; Massimiliano Zecca, Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University and National Centre for Sport and Exercise Medicine (NCSEM) – Loughborough; Giovanni Musella, Polo Universitario Asti Studi Superiori (Uni-Astiss) University of Torino

Acknowledgements. 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 Città di Asti. The authors declare no conflict of interest. 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 manuscript, 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. Correspondence concerning this article should be addressed to Daniele Magistro, School of Sport, Exercise, and Health Sciences, Loughborough University, Loughborough, Leicestershire, LE11 3TU, United Kingdom. E-mail: d.magistro@lboro.ac.uk

Abstract

This study evaluated whether the Test of Gross Motor Development 3 (TGMD-3) is a reliable tool to compare children with and without mental and behavioural disorders across gross motor skill domains. A total of 1075 children (aged 3-11 years), 98 with mental and behavioural disorders and 977 without (typically developing), were included in the analyses. The TGMD-3 evaluates fundamental gross motor skills of children across two domains: locomotor skills and ball skills. Two independent testers simultaneously observed children's performances (agreement over 95%). Each child completed one practice and then two formal trials. Scores were recorded only during the two formal trials. Multigroup Confirmatory Factor Analysis tested the assumption of TGMD-3 measurement invariance across disability groups. According to the magnitude of changes in Root Mean Square Error of Approximation and Comparative Fit Index between nested models, the assumption of measurement invariance across groups was valid. Loadings of the manifest indicators on locomotor and ball skills were significant ($p < .001$) in both groups. Item Response Theory analysis showed good reliability results across locomotor and the ball skills full latent traits. The present study confirmed the factorial structure of TGMD-3 and demonstrated its feasibility across normally developing children and children with mental and behavioural disorders. These findings

provide new opportunities for understanding the effect of specific intervention strategies on this population.

Key words: motor skills, TGMD-3, multigroup analysis, item response theory, motor development

Public Significance Statements

The findings from this study show that the TGMD-3 assessment technique for gross motor skills provides the same degree of accuracy in children with or without intellectual or behavioural disorders.

The TGMD-3 has potential to inform the design and evaluation of motor development programs targeting/including children with mental and behavioural disorders.

Introduction

Gross motor skills refer to goal-directed movement patterns (Burton & Miller, 1998) involving large whole body movements, locomotion, and whole body stretches (Woodfield, 2004). These skills can be broadly divided into: *locomotor skills*, such as walking, running, hopping, galloping, jumping, sliding, and leaping; and *object control skills*, such as throwing, catching, striking, bouncing, kicking, pulling, and pushing (Burton & Miller, 1998). Gross motor skills proficiency is a prerequisite for performing sport-specific skills and for successfully participating in organized and unorganized physical activities (Magistro, Bardaglio, & Rabaglietti, 2015; Piek, Dawson, Smith, & Gasson, 2008; Karabourniotis, Evaggelinou, Tzetzis, & Kourtessis, 2002). In general, individuals with higher levels of gross motor skills are more physically active than those with lower levels in childhood (Estevan et al., 2017; C.-I. Kim, Han, & Park, 2014; Lopes, Stodden, Bianchi, Maia, & Rodrigues, 2012) also in later life (Smith, Fisher, & Hamer, 2015). Children with inadequate motor competency or proficiency in gross motor skills often report lower levels of perceived physical competence (Robinson, Rudisill, & Goodway, 2009), self-esteem, and social acceptance (Skinner & Piek, 2001; Nadia C. Valentini, Zanella, & Webster, 2016). Moreover, a general motor delay at young age can negatively affect competences in physical and motor activities later in life (Jürimäe & Jürimäe, 2000).

In children with mental and behavioural disorders (e.g. intellectual disability, specific developmental disorder, autism and Attention Deficit Hyperactivity Disorder – ADHD) competence in gross motor skills may be limited and their development less predictable (Goodway & Branta, 2003; Goodway, Crowe, & Ward, 2003). These children usually show a delay in both gross and fine motor skills (Goodway & Branta, 2003; Goodway et al., 2003; Krebs & Winnick, 2005; Westendorp, Hartman, Houwen, Smith, & Visscher, 2011; Woodward & Surbrug, 2001), and report lower scores than typically developing children on measures of strength, endurance, balance, running, speed, coordination, and reaction time (Krebs & Winnick, 2005; Simons et al., 2008).

Adequate gross motor skills are important in children with and without mental and behavioural disorders, as it can foster future achievements in motor and cognitive performances (Son & Meisels, 2006; Viholainen et al., 2006). Consequently, motor skills assessment in children with disabilities is particularly essential during the preschool and school years. It is thus fundamental to monitor changes and identify possible deficits in motor skills so to support the design of specific programmes/interventions aimed at improving motor capability and competency (Burton & Miller, 1998). Indeed, there is a strong consensus that motor skills are positively associated with multiple aspects of health-related fitness, including cardiorespiratory fitness, as well as muscular strength, endurance and flexibility (Robinson et al., 2015). Furthermore, previous studies suggested a relationship between early gross motor and later school aged cognitive development (i.e. processing speed and working memory) (Piek et al., 2008; Son & Meisels, 2006; H. Kim, Duran, Cameron, & Grissmer, 2017) and a relationship between reading and locomotor skills, and mathematics and object-control skills in children with learning disorder (Westendorp et al., 2011).

An appropriate assessment of children's motor skills and the monitoring of the development of these overtime depends on the use of reliable instruments (Netelenbos, 2005; Nadia Cristina Valentini, 2012). The first and second editions of the Test of Gross Motor Development (TGMD) (Dale Allen Ulrich & Sanford, 1985) have been shown to be valid and reliable tools to assess gross motor skills in children without (Nadia Cristina Valentini, 2012; Wong & Yin Cheung, 2010; J.-T. Kim, 2001; Woodard & Surburg, 1997) and with (Allen, Bredero, Van Damme, Ulrich, & Simons, 2017; Benedict et al., 2011; Simons et al., 2008) mild intellectual disability. The current slightly modified third edition (TGMD-3), as the previous ones, is a process-oriented test of gross motor skills in young children, aged 3–10 years (D.A. Ulrich, 2016).

The TGMD-3 assessment focuses on both locomotor and object control (ball skills aspects). Such skills represent the fundamental motor skills that are commonly taught in primary school (from 5/6 to 10/11 years old) physical education curricula worldwide (Allen et al., 2017). An important function of the TGMD-3 is to help identifying possible delays and deficits in gross motor skills development among children, as well as helping to understand and compare their development among typically developing children and children with atypical movement function (Krebs & Winnick, 2005; Simons et al., 2008; Allen et al., 2017; D.A. Ulrich, 2016; Webster & Ulrich, 2017).

To the best of our knowledge, no study has yet examined the measurement invariance properties of the TGMD-3 across children with and without mental and behavioural disorders. One study by Simons and Eyitayo (2016) showed an acceptable level of inter and intra-rater reliability and some content validity aspects of the TGMD-3 score in children with intellectual disability, but the sample size was small (n=19). Therefore, the aim of this study was to assess whether the TGMD-3 is a reliable tool to directly compare typically developing children and children with mental and behavioural disorders across gross motor skills domains. Simultaneously assessing children with and without mental and behavioural disorders in the same context may indeed facilitate the inclusion of non-normally developing children during research and school-based interventions possibly reducing in turn feelings of social exclusion.

Method

Participants

This study is part of a three-year longitudinal project “Benessere in Gioco” (BiG, roughly translated in English as “Wellness at Play”) focusing on understanding physical and behavioural development among preschool and primary school children in the north west of Italy. The project involved 32 schools (14 nursery schools and 18 primary schools), including a total of 4035 children. The present analyses focused on preschools and primary children with mental and behavioural disorders and their typically developing classmates. Children were considered to have mental and behavioural disorders if they had an official certification issued by the Italian health service. For the certification the Italian health service system is using the International Statistical Classification of Diseases and Related Health Problems 10th Revision - ICD-10 (World Health Organization, 2004). Consequently, the same classification has been used in this study. The certification is composed by: a functional diagnosis, drawn up by the local health authority; and a dynamic-functional profile, draw up by the local health authority together with curricular and support teachers. Therefore, 8 nursery schools and 3 primary schools were excluded because they did not have any children with mental and behavioural disorders attending their schools. Thus, the sample for the present analyses were recruited from 6 nursery schools (137 children, aged 3-5 years old) and 15 primary schools (938 children, aged 6-11 years old).

The Consolidated Standards of Reporting Trials diagram (Figure S1) summarizes the recruitment attendance and exclusion criteria information of our study sample. According to the school records, 134 (3% of the total recruited sample) children were classified as having a diagnosis of mental and behavioural disorders. Children were excluded from the study if they had a severe physical disorder that prevented independent mobility or a severe cognitive disorder that prevented participation in the tests. Regarding the typically developing children group, 1083 (27% of the total recruited sample) children who attended the same class as the children with mental and behavioural disorders were included in the present analyses. The final sample size for the present analysis was 1075 (977 typically developing children and 98 children with mental and behavioural disorders). Table S1 reports the classification of the children with mental and behavioural disorders based on ICD-10 (Version 2016). This selection method allowed the comparison of children with mental and/or behavioural disorders to their typically developing peers, who were exposed to the same daily school context. Informed consent was obtained from both children's parents/guardians and the children. The ethical committee of University of Torino approved the study (ID 100949).

Instruments

The TGMD-3 evaluates fundamental gross motor skills of typically developing children between 3 and 10 years of age (D.A. Ulrich, 2016; Webster & Ulrich, 2017). As in the previous version, the test is divided in two subscales: the locomotor skill subscale and the ball skill (named object control skill in TGMD-2) subscale. The locomotor skill subscale consists of six skills: run, gallop, hop, skip, horizontal jump and slide. There are four performance criteria for every skill, while skip is judged on three criteria. The object control subscale is composed of seven skills: forehand strike of self-bounced ball, kick a stationary ball, overhand throw, underhand throw that are 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. The underhand roll and leap skills have been removed and a skip, one-handed forehand strike, and underhand throw have been added. Thus, the total number of observed skills in the TGMD-3 is thirteen (compared with twelve in the TGMD-2). The performance were observed and evaluated following the qualitative performance criteria that represent each TGMD-3 assessment item: each respective criterion was scored based on whether the criterion was fulfilled (awarded score $= 1$) or not (awarded score $= 0$) (D.A. Ulrich, 2016). Two trials are performed for each item. The total score for each item is established by the sum of all performance criteria scores in both trials. To determine the total locomotor and the object control skills subtest scores and the overall TGMD-3 were considered the sum of performance criteria scores of both trials (Allen et al., 2017; D.A. Ulrich, 2016; Webster & Ulrich, 2017). The maximum scores a participant can obtain for the locomotor subtest is 46, for ball skills subtest is 54, and for overall gross motor performance is 100.

Procedure

The total sample of 1075 children (98 children with mental and behavioural disorders and 977 typically developing children) were assessed to investigate the measurement invariance. A total of 16 professionals were involved in the test administration: four sport science researchers, two psychologists and ten physical education professionals. The tests were administered in the gym in each school. All the children were tested during the school hours in the presence of their physical education teachers. The support teacher assisted the administration of the tests in children with mental and behavioural disorders. The demographic data (age, gender, nationality) were reported by the children's parents. Each child's height, weight, body mass was assessed at baseline. Stretch stature was assessed using a wall mounted stadiometer (International Standard ISO/TR 7250-2, 2010) and weight using a Tanita Body Composition Analyzer BF-350. BMI was calculated using the following calculation: weight(kg)/height(m)^2. Children were classified as underweight, normal weight, overweight or obese using Italian growth charts (Cacciari et al., 2006).

The administration of the TGMD-3 test required approximately 20 minutes for each child. Two independent testers simultaneously observed each child's performance. The proportion of agreement between the scores recorded by the two testers was over 95%. The administration of the TGMD-3 was conducted following the authors' recommendations (D.A.

Ulrich, 2016; Webster & Ulrich, 2017): all equipment organized prior to the test; and an accurate demonstration and verbal description of each skill were carried out at the beginning of each test section. Each child completed one practice and then two formal trials and the scores were recorded for only the two formal trials.

Data analyses

Descriptive statistics

Descriptive statistics and analyses were run to check for accuracy of data entry and underlying assumption of normality. To test the latter assumption, the rules-of-thumb by Kline (2015) were followed : absolute skewness and kurtosis values lower than 3.0 and 8.0, respectively. Subsequently, Chi-square (χ^2) and t-test for independent samples tested for significant differences between typically developing children and children with mental and behavioural disorders socio-demographic variables. A series of t-tests were conducted to test for significant differences in single tasks of aggregated scales between children with and without mental and behavioural disorders to verify and describe potential differences between the two groups.

Confirmatory Factor Analysis

A Multigroup Confirmatory Factor Analysis (MGCFA), using the maximum likelihood (ML) estimation procedure, was used to test the assumption of TGMD-3 measurement invariance across typically developing children and children with mental and behavioural disorders. Factorial invariance testing followed a series of hierarchical steps, each comprising consecutive constraints across groups. An initial confirmatory analysis tested the proposed model (see Figure 1) in each group separately. Then, it was tested whether the same parameters existed for both groups (*configural invariance*), namely if the model structure is invariant across groups. Following this first step, additional constraints were investigated through factor loadings (*metric invariance*), item intercepts (*scalar invariance*), and residual variances (*strict invariance*). As recommended by Kenny (2015), model fit was evaluated using (1) χ^2 goodness-of-fit, (2) Root Mean Square Error of Approximation (RMSEA; with values ≤ .08 being indicative of acceptable fit to the data), and (3) Comparative Fit Index (CFI; ≥ .90). A change of ≥ -.01 in CFI between configural and metric invariance models, in addition to a change of ≥ -.02 in RMSEA indicated non-invariance, while a change of ≥ -.01 in CFI or ≥ -.02 in RMSEA would confute scalar or strict invariance (Chen, 2007; Vandenberg & Lance, 2000).

PLEASE INSERT FIGURE 1 ABOUT HERE

Item Response Theory

Finally, Item Response Theory analysis (IRT) using a Graded Response Model (GRM) estimated at which levels of the underlying traits of locomotor and ball skills children are likely to score at a given performance criterion scale point. The GRM is a cumulative categories approach to polytomous IRT modelling. Given the ordered-response nature of TGMD-3's items, where each increasing score represents an increasing quantity of gross motor skills attributes (Samejima, 2004), this was the preferred approach. Moreover, the GRM is an extension of the two-parameter logistic (2-PL) model. This model is preferable to the 1-PL or 3-PL models for our data because it does not have an assumption of equal discrimination across all items nor an adjustment for guessing, which is inappropriate for what the TGMD-3 is measuring (Chang, Lin, Gronholm, & Wu, 2016). The following parameters were estimated for every item: the discrimination item parameter (DI), representing the extent to which an item discriminates between different trait levels (higher values indicating a stronger association with the measured construct); and the difficulty or threshold parameters of each scoring level, referring to the latent trait score needed to have a .5 probability of scoring on a particular level or higher. The estimated information functions for each item were plotted based on the estimated parameter along with the conditional standard error and test information functions for each sub-scale as a whole to evaluate the fit of the model to the data.

Statistical software

SPSS (version 22; SPSS, Inc., Chicago, IL) and Amos (version 20; SPSS, Inc., Chicago, IL) software were used for the descriptive and multi-group factorial analyses, respectively. IRT analysis was carried out using STATA software (version 15; StataCorp LP, College Station, TX).

Results

Descriptive analyses

Absolute skewness and kurtosis values across all TGMD-3 tasks ranged respectively from .225 to 1.548 and .150 to 2.100 among children without disability and from .032 to 1.455 and .049 to 2.210 among children with disability. These results show that the assumptions of normality were met in both groups. Table 1 reports descriptive statistics and significant differences in socio-demographic variables between typically developing children and children with mental and behavioural disorders.

PLEASE INSERT TABLE 1 ABOUT HERE

Measurement invariance across disability status

Table 2 summarizes the results of measurement invariance tests across the two groups. Overall, according to the magnitude of changes in RMSEA and CFI between nested models, the assumption of measurement invariance across groups is valid. More specifically, the proposed factorial structure is tenable across typically developing and children with mental and behavioural disorders. Results from the configural invariance model indicate that the two-factor solution was tenable across groups. Table S2 reports standardized factor loadings from both groups according to the configural model. Loadings of the manifest indicators on locomotor skills and ball skills were all significant at $p < .001$ in both groups, ranging from $\lambda = .53$ to $\lambda = .82$ among children with mental and behavioural disorders and from $\lambda = .44$ to $\lambda = .80$ among typically developing children. Overall, the pattern of factor loadings of the skill areas was similar for children with and without disability and the differences between groups were small. Reliability analysis of test score based on Cronbach's alphas (α) yielded good results for locomotor skills ($\alpha = .80$ for both groups) and ball skills ($\alpha = .86$ and $\alpha = .78$) for the children with mental and behavioural disorders and the typically developing group, respectively) subscales. Correlations between subscales were equal to .89 among children with mental and behavioural disorders and to .81 among typically developed children.

In the model testing metric invariance, the factor loadings were constrained to be equal for children with and without disability to determine whether the skill areas function in a similar way on the two general factors for the two groups. Differences in goodness-of-fit indices with the configural model solution were small (ΔRMSEA = -.001; ΔCFI = -.005), indicating the model with equal loadings across groups fits the data as well as the model without equality constraints and that factor loadings differences are negligible.

Through constraining the item intercepts to be equal across groups in the scalar invariance model, we obtained fit results substantially similar to the previous model with only equal factor loadings (ΔRMSEA = -.002; ΔCFI = -.002). Furthermore, the intercepts for children with and without disability are small, thus supporting the assumption of strong factorial invariance.

Finally, the model testing strict invariance was used to examine whether the skill areas measure locomotor and ball skills with the same accuracy (by comparing the unique variance of skill areas) between the two groups. The similarities of fit results between the strict invariance and the scalar invariance model (ΔRMSEA = -.001; ΔCFI = -.005), indicate that the unique variances did not differ significantly between children with and without disability and that the skill areas measure locomotor and ball skills with similar accuracy across groups.

An independent t test showed that the total TGMD-3 score ($p < 0.05$) and the locomotor subscale score ($p < 0.01$) were statistically different between groups. Conversely, there were no significant differences between groups in ball skills score (see Table 3).

PLEASE INSERT TABLE 2 ABOUT HERE

PLEASE INSERT TABLE 3 ABOUT HERE

IRT analysis

The parameter estimates of the GRM for all the items of the two TGMD-3's sub-scales are shown in Table S3. Figures S2 and S3 show the item information functions for the locomotor and ball skills sub-scales, respectively. In the locomotor skills sub-scale, the first three items provide the highest amount of information to measure locomotor skills in the current sample ("Run", DI = 1.968; "Gallop", DI = 1.669; and "Hop", DI = 1.575) and appeared to discriminate best between children with low and high levels of locomotor skills. For the ball skills sub-scale, the highest DI were observed for the items measuring “Forehand strike of self-bounced ball” (DI = 1.877), “One-hand stationary dribble” (DI = 1.589), and “Kick a stationary ball” (DI = 1.552).

The conditional standard error and test information functions for the two sub-scales are shown in Figures 2 and 3. Both batteries of items have low levels of standard error and concurrently high levels of test information across a broad range of locomotor and ball skills. These results indicate that the TGMD-3’s sub-scales demonstrate strong discriminative power across the spectrum of gross motor skills.

PLEASE INSERT FIGURE 2 ABOUT HERE

PLEASE INSERT FIGURE 3 ABOUT HERE

Discussion

The aim of this study was to assess the measurement invariance characteristics of the TGMD-3 across typically developing children and children with mental and behavioural disorders. Based on our results, the TGMD-3 can be considered an appropriate instrument to simultaneously evaluate and compare general motor skills in children with mental and behavioural disorders and typically developing children.

The present study took a further step in assessing the validity of the TGMD-3 score, confirming its factorial structure and demonstrating its generalizability across typically developing children and children with mental and behavioural disorders. Moreover, IRT analysis showed that this instrument can discriminate efficiently between children with low and high levels of gross motor skills. Results obtained from this study have significant implications for future research aiming to compare typically developing children and children with disabilities. Multi-factorial item tests measuring physical abilities across relevant groups (Van De Schoot, Schmidt, De Beuckelaer, Lek, & Zondervan-Zwijnenburg, 2015) rely on the assumption of factorial structural validity for valid between groups comparisons (Mellenbergh, 1989). Our results highlighted that the TGMD-3 had a good fit in terms of strict invariance across typically developing children and children with mental and behavioural disorders, situation observable when measurement parameters are precisely the same across groups (Borsboom, 2006; Van De Schoot et al., 2015). These results emphasize the validity of the TGMD-3, comparable with TGMD-2 (Simons et al., 2008), to measure general motor skills in children with mental and behavioural disorders. Females were underrepresented in the mental and behavioural disorders group (27.8%) in the present sample and there was a small but significant difference in BMI between the two groups. However, the percentages in each BMI classification reflect the regional and national distributions in both groups (Nardone et al., 2016). As expected, typically developed children, in general, performed better than their peers with mental and behavioural disorders on the whole TGMD-3 and on the four locomotor tasks (hop, skip, horizontal jump, and slide), and consequently reported overall better on the aggregate locomotor subscale. Conversely, regarding ball skills tasks, children with mental and behavioural disorders performed worse than their typically developed peers only in the underhand throw while there were no significant differences at this second aggregated subscale level. This second results could be related to the small number (12) in our sample of children with mental and behavioural disorders under 7 years old. Indeed, children of this age usually show worst performance on object control tasks compare to older children (Payne & Isaacs, 2017). Our results suggest that future assessments using TGMD-3 can include children with and without mental and behavioural disorders. In turn, this will facilitate the evaluation of specific interventions across a more inclusive sample and promoting new evidence-based strategies to social inclusion of children with mental and behavioural disorders.

There are some limitations to the current study. First, as pointed out by Brown (2015), it is possible to conduct MGCFA with unequal group sizes but at the same time the power to detect violation of invariance decreases as the ratio of the sample size between the compared groups increases. Although one of the advantages of the current study was the fairly large sample of children with and without disabilities from the same context, future studies aiming to replicate our invariance results regarding the use of TGMD-3 across this type of groups should use more balanced group sizes so to reduce the risk of biases. Second, to fully demonstrate the utility of the TGMD-3, future research should examine whether the validity of TGMD-3 test score interpretations, such as predictive concurrent and content validity, are equivalent across both groups of children. Third, this study did not control for the differences in chronological and biological (physical) age of the children. Lastly, children with different mental and behavioural disorders were pooled in order to optimize sample size. To enhance the generalizability of these findings, further studies need to recruit a larger sample size of children with mental and behavioural disorders (i.e., > 200; see Su, Ng, Yang, & Lin, 2014).

This will enable a more robust examination of measurement invariance characteristics of the TGMD-3 across a broader range of mental and behavioural disorders.

Conclusions

This study adds to our knowledge of how the assessment of general motor skills may be facilitated in children with mental and behavioural disorders. The TGMD-3 was shown to be a reliable tool to assess gross motor skills in children with mental and behavioural disorders. From a clinical point of view, the identification of delayed gross motor skill development in children with mental and behavioural disorders in comparison to their peers may suggest a need for adaptive intervention to mitigate the limitations which may prevent possible lifelong involvement in physical activities. The TGMD-3 might also be useful as a guideline to design motor development programs and goal setting to specifically address the skills in which the performance of the children with mental and behavioural disorders may be lacking.

References

Allen, K. A., Bredero, B., Van Damme, T., Ulrich, D. A., & Simons, J. (2017). Test of Gross Motor Development-3 (TGMD-3) with the Use of Visual Supports for Children with Autism Spectrum Disorder: Validity and Reliability. *Journal of Autism and Developmental Disorders*, *47*(3), 813–833. https://doi.org/10.1007/s10803-016-3005-0

Benedict, R. E., Patz, J., Maenner, M. J., Arneson, C. L., Yeargin-Allsopp, M., Doernberg, N. S., Van Naarden Braun, K., Kirby, R. S. & Durkin, M. S. (2011). Feasibility and reliability of classifying gross motor function among children with cerebral palsy using population-based record surveillance. *Paediatric and Perinatal Epidemiology*, *25*(1), 88–96.

Borsboom, D. (2006). When does measurement invariance matter? *Medical Care*, *44*(11), S176–S181.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research - Second edition*. New York, NY: Guilford Publications.

Burton, A. W., & Miller, D. E. (1998). *Movement skill assessment*. Human Kinetics.

Cacciari, E., Milani, S., Balsamo, A., Spada, E., Bona, G., Cavallo, L., … Cicognani, A. (2006). Italian cross-sectional growth charts for height, weight and BMI (2 to 20 yr). *Journal of Endocrinological Investigation*, *29*(7), 581–593.

Chang, C.-C., Lin, C.-Y., Gronholm, P. C., & Wu, T.-H. (2016). 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*, 1073191116658547.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504.

Estevan, I., Molina-García, J., Queralt, A., Álvarez, O., Castillo, I., & Barnett, L. (2017). Validity and Reliability of the Spanish Version of the Test of Gross Motor Development-3. *Journal of Motor Learning and Development*, 1–21.

Goodway, J. D., & Branta, C. F. (2003). Influence of a motor skill intervention on fundamental motor skill development of disadvantaged preschool children. *Research Quarterly for Exercise and Sport*, *74*(1), 36–46.

Goodway, J. D., Crowe, H., & Ward, P. (2003). Effects of motor skill instruction on fundamental motor skill development. *Adapted Physical Activity Quarterly*, *20*(3), 298–314.

International Standard ISO/TR 7250-2. (2010). *Basic human body measurements for technological design. Part 2: Statistical summaries of body measurements from individual ISO populations* (ISO). Geneva, Switzerland.

Jürimäe, T., & Jürimäe, J. (2000). *Growth, physical activity, and motor development in prepubertal children*. CRC Press Boca Raton, FL.

Karabourniotis, D., Evaggelinou, C., Tzetzis, G., & Kourtessis, T. (2002). Curriculum enrichment with self-testing activities in development of fundamental movement skills of first-grade children in Greece. *Perceptual and Motor Skills*, *94*(3_suppl), 1259–1270.

Kenny, D. A. (2015). *Measuring model fit*.

Kim, C.-I., Han, D.-W., & Park, I.-H. (2014). Reliability and validity of the Test of Gross Motor Development-II in Korean preschool children: applying AHP. *Research in Developmental Disabilities*, *35*(4), 800–807.

Kim, H., Duran, C. A. K., Cameron, C. E., & Grissmer, D. (2017). Developmental Relations Among Motor and Cognitive Processes and Mathematics Skills. *Child Development*, n/a-n/a. https://doi.org/10.1111/cdev.12752

Kim, J.-T. (2001). *The effects of a physical education program on the standing long jump performance of preschool-aged children with cognitive delays*. Microform Publications, University of Oregon.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. London: Guilford publications.

Krebs, P. L., & Winnick, J. (2005). Intellectual disabilities. *Adapted Physical Education and Sport*, *1*, 133.

Lopes, V. P., Stodden, D. F., Bianchi, M. M., Maia, J. A., & Rodrigues, L. P. (2012). Correlation between BMI and motor coordination in children. *Journal of Science and Medicine in Sport*, *15*(1), 38–43.

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

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, *13*(2), 127–143.

Nardone, P., Spinelli, A., Buoncristiano, M., Lauria, L., Pizzi, E., Andreozzi, S., & Galeone, D. (2016). Il Sistema di sorveglianza OKkio alla SALUTE: risultati 2014. *Istituto Superiore Di Sanità. Roma*.

Netelenbos, J. B. (2005). Teachers' ratings of gross motor skills suffer from low concurrent validity. *Human Movement Science*, *24*(1), 116–137.

Payne, V. G., & Isaacs, L. D. (2017). *Human motor development: A lifespan approach*. Routledge.

Piek, J. P., Dawson, L., Smith, L. M., & Gasson, N. (2008). The role of early fine and gross motor development on later motor and cognitive ability. *Human Movement Science*, *27*(5), 668–681.

Robinson, L. E., Rudisill, M. E., & Goodway, J. D. (2009). Instructional climates in preschool children who are at-risk. Part II: Perceived physical competence. *Research Quarterly for Exercise and Sport*, *80*(3), 543–551.

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*, *45*(9), 1273–1284.

Simons, J., Daly, D., Theodorou, F., Caron, C., Simons, J., & Andoniadou, E. (2008). Validity and reliability of the TGMD-2 in 7–10-year-old Flemish children with intellectual disability. *Adapted Physical Activity Quarterly*, *25*(1), 71–82.

Simons, J., & Eyitayo, G. (2016). Aspects of reliability and validity of the TGMD-3 in 7-10 year old children with intellectual disability in Belgium. *European Psychomotricity Journal*, *8*(1), 3–16.

Skinner, R. A., & Piek, J. P. (2001). Psychosocial implications of poor motor coordination in children and adolescents. *Human Movement Science*, *20*(1), 73–94.

Smith, L., Fisher, A., & Hamer, M. (2015). Prospective association between objective measures of childhood motor coordination and sedentary behaviour in adolescence and adulthood. *International Journal of Behavioral Nutrition and Physical Activity*, *12*, 75. https://doi.org/10.1186/s12966-015-0236-y

Son, S.-H., & Meisels, S. J. (2006). The relationship of young children's motor skills to later reading and math achievement. *Merrill-Palmer Quarterly (1982-)*, 755–778.

Su, C.-T., Ng, H.-S., Yang, A.-L., & Lin, C.-Y. (2014). Psychometric evaluation of the Short Form 36 Health Survey (SF-36) and the World Health Organization Quality of Life Scale Brief Version (WHOQOL-BREF) for patients with schizophrenia. *Psychological Assessment*, *26*(3), 980–989. https://doi.org/10.1037/a0036764

Ulrich, D. A. (2016). *Test of gross motor development (3rd ed.)*. Austin, TX: Pro-Ed.

Ulrich, D. A., & Sanford, C. B. (1985). *Test of gross motor development*. Pro-ed Austin, TX.

Valentini, N. C. (2012). Validity and Reliability of the TGMD-2 for Brazilian Children. *Journal of Motor Behavior*, *44*(4), 275–280. https://doi.org/10.1080/00222895.2012.700967

Valentini, N. C., Zanella, L. W., & Webster, E. K. (2016). Test of Gross Motor Development—Third Edition: Establishing Content and Construct Validity for Brazilian Children. *Journal of Motor Learning and Development*, *5*(1), 15–28. https://doi.org/10.1123/jmld.2016-0002

Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Editorial: Measurement Invariance. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01064

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70.

Viholainen, H., Ahonen, T., Lyytinen, P., Cantell, M., LicSSc, A. T., & Lyytinen, H. (2006). Early motor development and later language and reading skills in children at risk of familial dyslexia. *Developmental Medicine & Child Neurology*, *48*(5), 367–373.

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

Westendorp, M., Hartman, E., Houwen, S., Smith, J., & Visscher, C. (2011). The relationship between gross motor skills and academic achievement in children with learning disabilities. *Research in Developmental Disabilities*, *32*(6), 2773–2779. https://doi.org/10.1016/j.ridd.2011.05.032

Wong, K. Y. A., & Yin Cheung, S. (2010). Confirmatory factor analysis of the Test of Gross Motor Development-2. *Measurement in Physical Education and Exercise Science*, *14*(3), 202–209.

Woodard, R. J., & Surburg, P. R. (1997). Fundamental gross motor skill performance by girls and boys with learning disabilities. *Perceptual and Motor Skills*, *84*(3), 867–870.

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

Woodward, R. J., & Surbrug, P. R. (2001). The performance of fundamental movement skills by elementary school children with learning disabilities. *Physical Educator*, *58*(4), 198.

World Health Organization. (2004). *ICD-10, International statistical classification of diseases and related health problems.* (Vol. 1). World Health Organization.

Table 1. Descriptive statistics and results of tests for significant differences between typically developing children and children with mental and behavioural disorders across socio-demographic variables. Values are frequencies (percentages) unless stated otherwise

	Children with mental and behavioural disorders (n = 98)	Typically developing children (n = 977)	*p*
Gender			< .001[a]
Female	28 (28%)	482 (50%)	
Male	70 (72%)	495 (50%)	
Age *Mean years*	8.28 (*SD* 1.98)	8,68 (*SD* 1.84)	.348[b]
Age groups			
3 years	0	6 (1%)	
4 years	1 (1%)	26 (3%)	
5 years	3 (3%)	50 (4%)	
6 years	8 (8%)	115 (12%)	
7 years	12 (12%)	175 (18%)	
8 years	12 (12%)	115 (12%)	
9 years	17 (17%)	159 (16%	
10 years	23 (24%)	205 (21%)	
11 years	22 (23%)	126 (13%)	
Nationality			.936[a]
Italian	93 (95%)	928 (95%)	
Other	5 (5%)	49 (5%)	
Residence			.013[a]

Table continues

Urban	85 (88.5%)	756 (77.6%)	
Rural	12 (11.5%)	221 (22.4%)	
Weight *Mean kg*	31 (*SD* 9.95)	32,4 (*SD* 11.02)	.195[b]

Height *Mean cm*	130 (*SD* 13.07)	133.44 (*SD* 11.40	.758[b]
BMI			.017[a]
Underweight	8 (8%)	24 (2%)	
Normal weight	71 (73%)	771 (79%)	
Overweight/Obese	19 (19%)	182 (19%)	

[a] Significant difference between groups according to χ^2 test. [b] Significant difference between groups according to t-test for independent samples. *SD* = Standard Deviation.

Table 2. Results of measurement invariance testing across typically developing children and children with mental and behavioural disorders

Models	χ^2	$\Delta\chi^2$	*df*	Δdf	RMSEA (90% CIs)	ΔRMSEA	CFI	ΔCFI	Comparison
Single-group solutions									
Disabled (n = 98)	79.140		64		.049 (.000, .082)		.970		
Not disabled (n = 977)	473.010		64		.081 (.074, .088)		.902		
Model 1. Configural invariance	552.421		128		.056 (.051, .060)		.909		
Model 2. Metric invariance	589.481	37.060^{***}	139	11	.055 (.050, .060)	-.001	.904	-.005	Model 2 vs. Model 1
Model 3. Scalar invariance	612.691	23.21^{**}	152	13	.053 (.049, .058)	-.002	.902	-.002	Model 3 vs. Model 2
Model 4. Strict invariance	649.527	36.836^{***}	168	16	.052 (.048, .056)	-.001	.897	-.005	Model 4 vs. Model 3

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.

$^{**}p < .01$; $^{***}p < .001$.

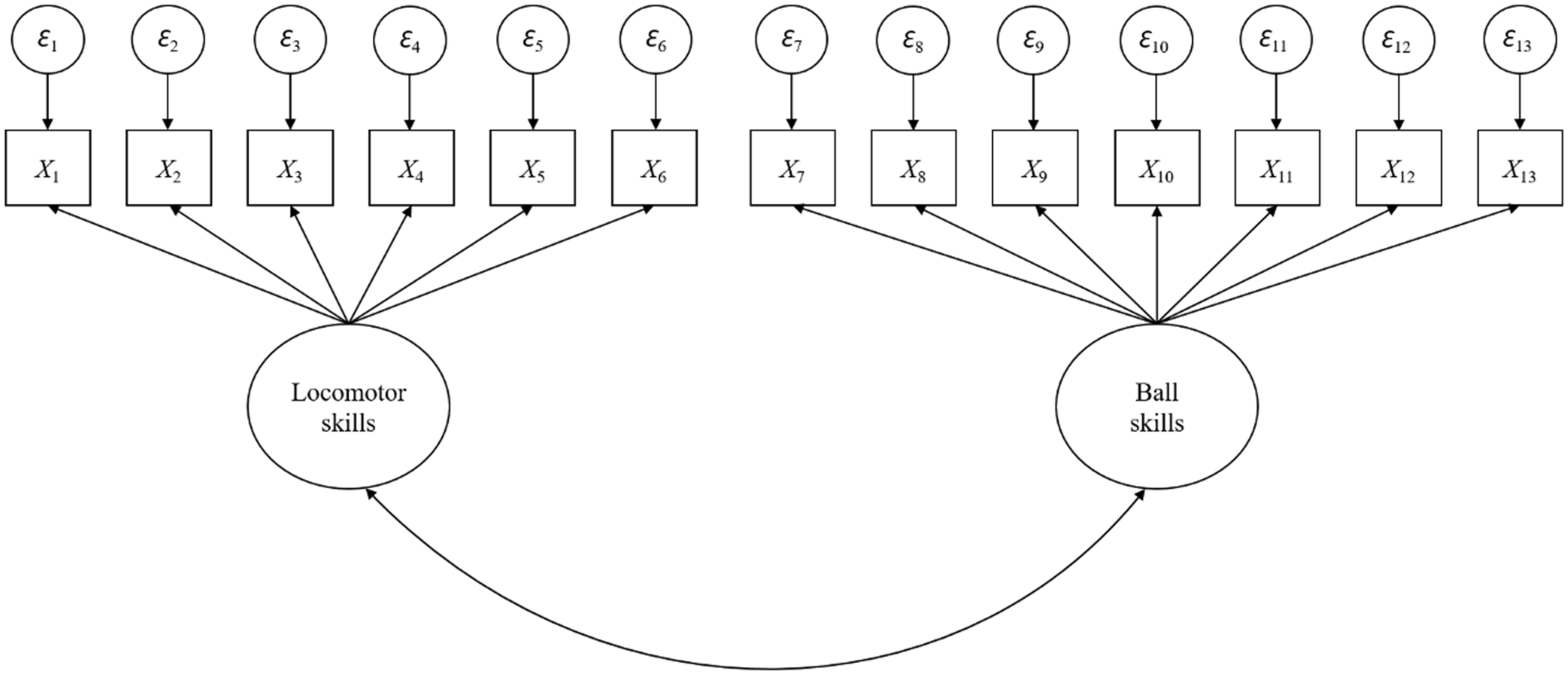
Table 3. Descriptive statistics and results of tests for significant differences between typically developing children and children with mental and behavioural disorders across TGMD-3 tasks

Tasks	Number of performance criteria	Range score	Children with mental and behavioural disorders (n = 98) *M*	*SD*	Typically developing (n = 977) *M*	*SD*	*p*
Locomotor skills		0–46	28.24	9.67	30.96	9.25	.006
Run	4	0–8	6.35	1.89	6.48	1.91	.485
Gallop	4	0–8	5.54	2.08	5.90	2.17	.119
Hop	4	0–8	4.63	2.28	5.11	1.96	.023
Skip	3	0–6	3.17	1.90	3.67	1.78	.007
Horizontal jump	4	0–8	4.54	2.39	5.07	2.28	.030
Slide	4	0–8	4.01	2.97	4.70	2.76	.019
Ball skills		0–54	36.61	11.67	38.22	9.20	.109
Two-hand strike of a stationary ball	5	0–10	6.54	2.45	6.67	2.36	.614

Forehand strike of self-bounced ball	4	0–8	5.23	2.48	5.36	2.37	.630
One-hand stationary dribble	3	0–6	4.14	2.22	4.25	2.02	.617
Two-hand catch	3	0–6	4.57	1.71	4.62	1.41	.725
Kick a stationary ball	4	0–8	5.73	1.98	5.94	1.63	.230
Overhand throw	4	0–8	4.91	2.36	5.30	2.16	.086
Underhand throw	4	0–8	5.47	2.52	6.07	1.79	.003

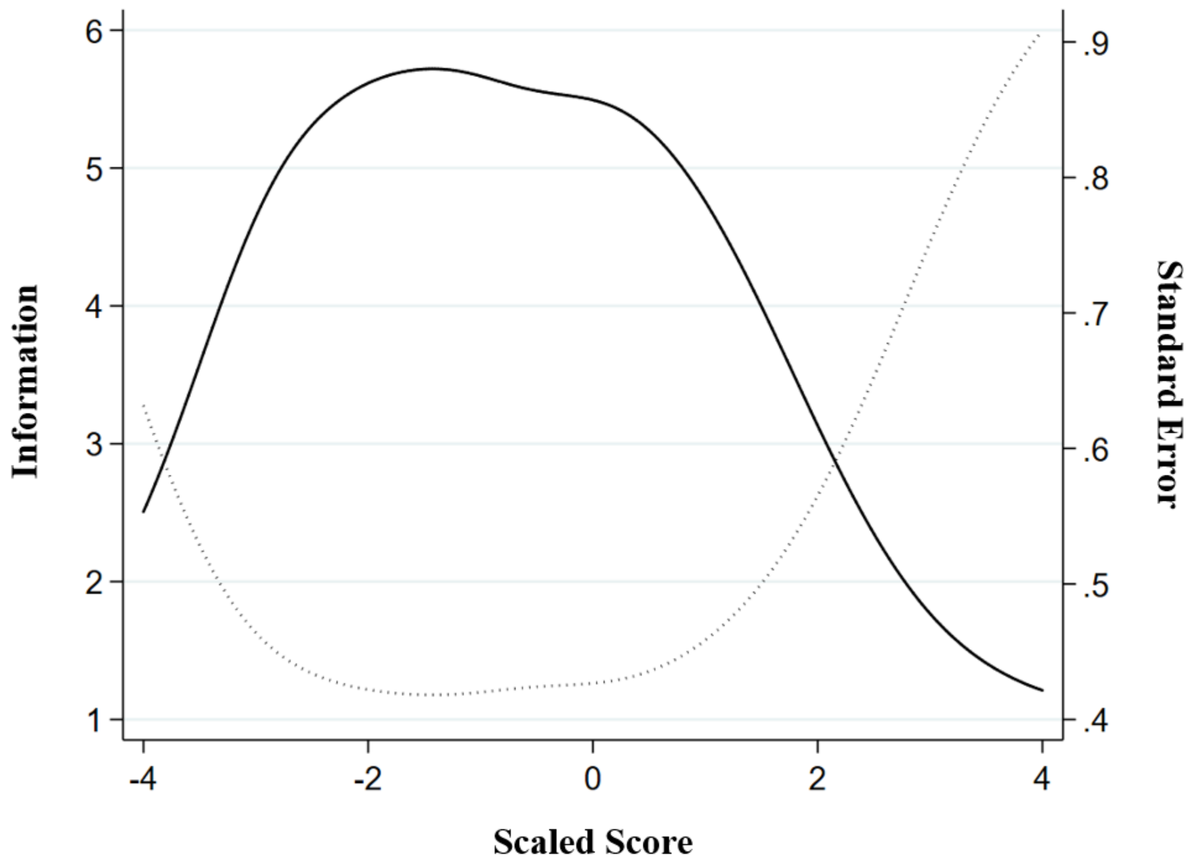
Note. TGMD-3: Test of Gross Motor Development – 3rd edition. M = Mean; SD = Standard Deviation. The significant difference between groups was tested according to t-test for independent samples.

Figure 1. Factorial model of TGMD-3



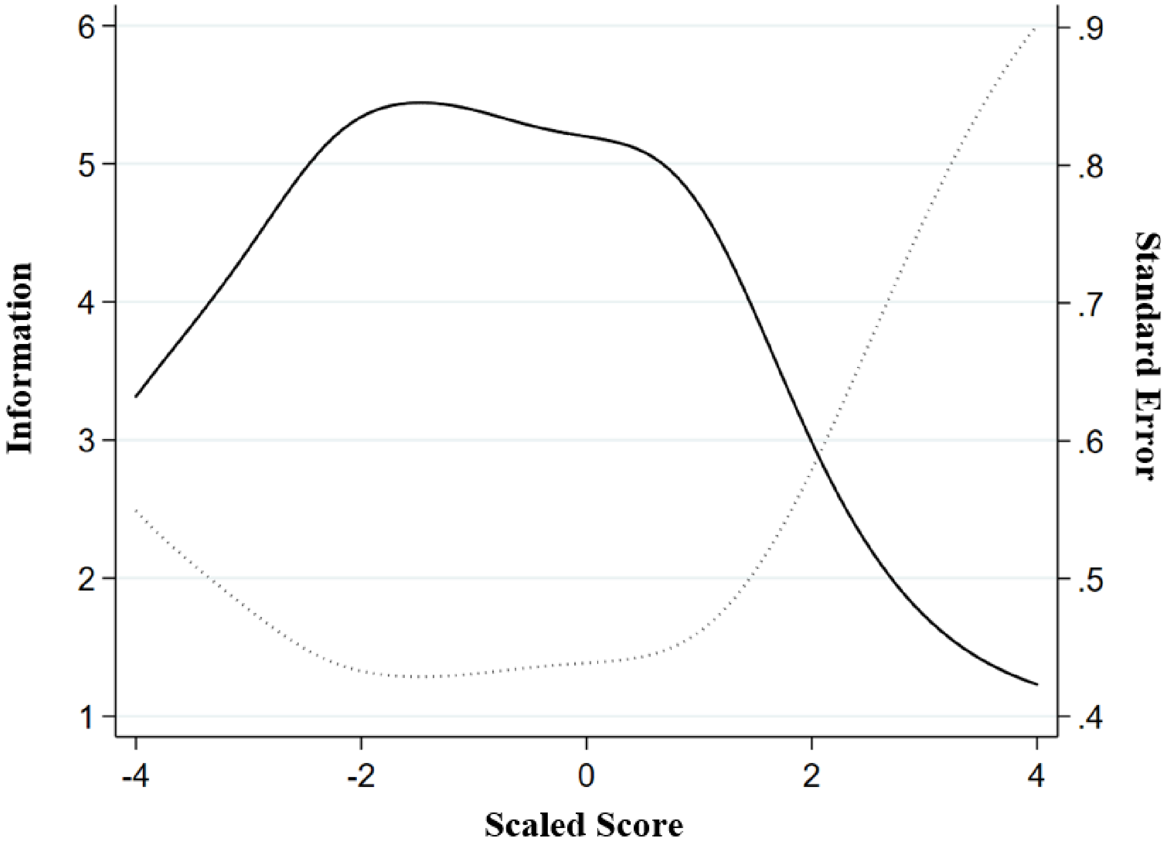
Note. TGMD-3: Test of Gross Motor Development – 3rd edition. X_1: Run; X_2: Gallop; X_3: Hop; X_4: Skip; X_5: Horizontal jump; X_6: Slide; X_7: Two-hand strike of a stationary ball; X_8: One-hand forehand strike of self-bounced ball; X_9: One-hand stationary dribble; X_{10}: Two-hand catch; X_{11}: Kick a stationary ball; X_{12}: Overhand throw; X_{13}: Underhand throw

Figure 2. Information graph showing the TGMD-3's locomotor skills sub-scale: test information function (solid line) and conditional standard error curve (dotted line) (N = 1,075)



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

Figure 3. Information graph showing the TGMD-3's ball skills sub-scale: test information function (solid line) and conditional standard error curve (dotted line) (N = 1,075)



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