Running head: French TGMD-3

Psychometric Properties of a French-Canadian Version of the Test of Gross Motor Development – Third Edition (TGMD-3): A Bifactor Structural Equation Modeling Approach

Christophe Maïano[a,b*], Alexandre J. S. Morin[b*], Johanne April[c], E. Kipling Webster[d], Olivier Hue[e], Claude Dugas[e], Dale Ulrich[f]

[a]Cyberpsychology Laboratory and Department of Psychoeducation and Psychology, Université du Université du Québec en Outaouais (UQO), Saint-Jérôme, Canada.
[b]Substantive-Methodological Synergy Research Laboratory, Department of Psychology, Concordia University, Montreal, Canada.
[c]Department of Education Sciences, Université du Québec en Outaouais (UQO|Campus de Saint-Jérôme), Saint-Jérôme, Canada.
[d]Institute of Public and Preventive Health, Augusta University, United States.
[e]Department of Physical Activity Sciences, Université du Québec à Trois-Rivières, Trois-Rivières, Canada
[f]School of Kinesiology, University of Michigan, Ann Arbor, United States.

* The order of appearance of the first and second authors (C.M. and A.J.S.M.) was determined at random: Both should be considered first authors.

Funding. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Corresponding author: Christophe Maïano, Université du Québec en Outaouais, Campus de Saint-Jérôme, Département de Psychoéducation et de Psychologie, 5 rue Saint-Joseph, Saint-Jérôme, Québec, J7Z 0B7, email: christophe.maiano@uqo.ca

This is the prepublication version of the following manuscript:

Maïano, C., Morin, A. J. S., April, J., Webster, K., Hue, O., Dugas, C., & Ulrich, D. (In Press, 2021). Psychometric properties of a French-Canadian version of the Test of Gross Motor Development – Third Edition (TGMD-3): A bifactor structural equation modeling approach. *Measurement in Physical Education and Exercise Science*. Doi: 10.1080/1091367X.2021.1946541

Abstract

The objective was to assess the psychometric properties of a French-Canadian version of the third edition of the Test of Gross Motor Development (TGMD-3). Participants were 127 French-speaking Canadian children. Results supported the validity-reliability of a bifactor exploratory structural equation modeling representation of the TGMD-3. Additionally, results supported a lack of differential item functioning as a function of age, body mass index (BMI), physical activity/sport practice (PA/SP), and sex. Finally, latent mean differences showed that: (a) older children score lower on specific skills and higher on the global motor skills factor than younger children; (b) children with a higher BMI score lower on locomotor skills than children with a lower BMI; (c) children with higher weekly frequency of PA/SP score higher on the global motor skills factor than children with a lower weekly frequency of PA/SP; and (d) boys score higher on ball skills and lower locomotor skills than girls.

Keywords: age; bifactor; body mass index; exploratory structural equation modeling; gross motor skills; physical activity; sex; sport.

Over the years, a multitude of tests have been developed to measure motor skills among children and adolescents. These tests generally rely either on a process- (i.e., how the skill is executed or performed) and/or a product-oriented (i.e., what is the results or the performance in terms of distance, velocity, etc.) scoring approach. Studies examining the validity and reliability of scores obtained on these tests have recently been synthetized in six systematic reviews (Cools et al., 2009; Eddy et al., 2020; Griffiths et al., 2018; Hulteen et al., 2020; Rey et al., 2020; Scheuer et al., 2019). According to these reviews, the original (Ulrich, 1985) and second edition (Ulrich, 2000) of the process-oriented Test of Gross Motor Development (TGMD) represents one of the most researched, and most psychometrically sound, motor skill tests available to date.

The TGMD has been recently revised, leading to the development of a third edition (TGMD-3; Ulrich, 2019; Webster & Ulrich, 2017). The TGMD-3 focuses on children's competence in six locomotor skills (i.e., running, hopping, sliding, galloping, skipping, and horizontal jumping) and seven ball skills (i.e., one-hand forehand striking of self-bounced ball, two-hand catching, kicking a stationary ball, two-hand striking of a stationary ball, one-hand stationary dribbling, overhand throw, and underhand throw). For each skill, children complete two trials. According to the type of skill, three to five performance criteria are assessed (Ulrich, 2019). When children perform a skill in a way that matches the performance criterion, they receive a score of one for this criterion (Ulrich, 2019). Conversely, when they do not meet the performance criterion, they receive a score of 0 (Ulrich, 2019). For each criterion, scores are summed across trials, before being summed across all performance criteria associated with a skill to obtain a raw score on that skill (Ulrich, 2019). As a result of the number of criteria incorporated to each skill, total skill scores range between 0 and 6 (i.e., one hand stationary dribble, two-hand catch, skip), 0 and 8 (i.e., gallop, hop, one-hand forehand, horizontal jump, kick a stationary ball, overhand throw, run, slide, underhand throw), and 0 and 10 (i.e., two-hand strike). Higher raw skill scores are indicative of higher motor skill competence. Finally, raw skill scores can be summed to obtain a global skills score (ranging from 0 to 100), as well as a total ball skills score (ranging from 0 to 54) and a total locomotor skills score (ranging from 0 to 46).

The psychometric properties of scores obtained on the TGMD-3 have been examined by Webster and Ulrich (2017) among a sample of 807 US children. The factor validity of scores obtained on the TGMD-3 was first examined using both exploratory factor analyses and confirmatory factor analyses (CFA). The results from these analyses supported a one-factor CFA solution (α= .96). Indeed, the a priori two-factor CFA (locomotor skills: α= .92; ball skills: α= .95) solution revealed factor correlations that were high enough (r = .96) to suggest the presence of a single underlying factor. Additional results revealed a significant and positive correlation between age and the global TGMD-3 factor (r = .45), as well as with the locomotor skills (r = .39) and ball skills factors (r = .47).

A Bifactor Representation of TGMD Ratings

The high correlations previously reported between the ball skills and locomotor skills factors of the TGMD call into question their true independence and suggest that these specific skills might in fact reflect a single overarching dimension, or at least co-exist with an overarching dimensions reflecting children's global levels of motor skills. However, another complementary explanation is also possible. The TGMD-3 indicators cover either ball skills or locomotor skills. However, even if these indicators are designed to measure a single type of skill, they may still involve some motor elements related to the other type of skill. This possibility highlights the value of relying on measurement models allowing for the free estimation of cross-loadings between each skill indicator and the other (non-target) factor. Statistical research has highlighted the fact that excluding cross-loadings, even as small as .10, from a measurement model led to an over-estimation of the factor correlations (Asparouhov et al., 2015). In contrast, including cross-loadings, even when none is present in the data, still results in accurate estimates of factor correlations. This possibility was recently examined by Garn and Webster (2018) using scores obtained on the TGMD-2 by a sample of 1,120 US children. These authors examined whether the factor structure of ratings obtained on the TGMD-2 would be better represented by a classical CFA solution (excluding cross-loadings) or by an exploratory structural equation modeling (ESEM; Asparouhov & Muthén, 2009) solution (including all cross-loadings). Their results supported the superiority of an ESEM representation of TGMD-2 scores, but also revealed that incorporating cross-loadings only resulted in a relatively small decrease in the size of the correlation obtained between scores on the two factors (r = .891 in CFA and .810 in ESEM). These results thus suggest that the need to incorporate cross-loadings to the measurement model underpinning children's

TGMD-2 scores might co-exist with the need to account for the dual global (global levels of motor skills) and specific (locomotor and ball skills) nature of TGMD-2 ratings.

In sport and exercise measurement, bifactor models have recently been proposed as an efficient way to account for the dual global and specific nature of psychometric ratings (e.g., Myers et al., 2014, 2016; for a review see Myers et al., 2018). Bifactor models assume that ratings obtained on a multidimensional instrument simultaneously reflect a global overarching dimension (global levels of motor skills), while also reflecting subscale specificities left unexplained by the global dimension (specific levels of balls skills and locomotor skills, reflecting the extent to which these skills deviate from the global component; see Morin et al., 2020). As such, bifactor models provide direct tests of whether it is possible to estimate a global construct underlying all skills rated in the TGMD, as well as of whether meaningful specificity remains associated with each facet of the TGMD once this global level is taken into account. The bifactor-ESEM framework makes it possible to simultaneously consider both sources of psychometric multidimensionality (i.e., cross-loadings, and the co-existence of global and specific dimensions). Furthermore, the incorporation of cross-loadings to a bifactor model also tends to result in a more accurate estimate of all factors (Morin et al., 2016, 2020).

The possible value of the bifactor-ESEM framework to account for the multidimensionality of TGMD-3 ratings has recently been investigated by Garn and Webster (2021) among a sample of 862 US children. In their study, these authors contrasted three alternative CFA models (i.e., one factor, two-factor, and a bifactor model) and two exploratory structural equation models (i.e., two-factor and a bifactor model). These results supported the superiority of a bifactor-ESEM structure, revealing the presence of a global motor skills factor defined by all indicators coexisting with two orthogonal specific factors defined by either the ball skills or locomotor skills factors reflecting the unique nature of each type of skill left unexplained by the global factor. This model resulted in acceptable levels of composite reliability for the global factor (ω= .916), as well as for the locomotor skills (ω= .836) and ball skills (ω= .878) factors.

Psychometric Properties in Other Languages and Differential Item Functioning

The psychometric properties of scores obtained on the TGMD-3 have also been examined among other linguistic groups, such as German (Wagner et al., 2017), Italian (Magistro et al., 2020), Persian (Mohammadi et al., 2019), Brazilian (Valentini et al., 2017), and Spanish (Estevan et al., 2017). Generally, these studies have supported either the two-factor structure of the TGMD-3 (Magistro et al., 2020; Valentini et al., 2017; Wagner et al., 2017), or both one-factor and two-factor structures (Estevan et al., 2017; Mohammadi et al., 2019). These studies have generally supported the scale score reliability of global motor skill ratings (α= .91-.96 in relation to a one-factor model; Estevan et al., 2017; Mohammadi et al., 2019) and of locomotor skills (α= .63-.85) and ball skills (α= .76-.89) ratings (in relation to the two-factor model; Estevan et al., 2017; Magistro et al., 2020; Mohammadi et al., 2019; Valentini et al., 2017; Wagner et al., 2017). However, replicating results previously reported by Webster and Ulrich (2017), most of the cross-validation studies (Estevan et al., 2017; Mohammadi et al., 2019; Valentini et al., 2017; Wagner et al., 2017) revealed very high correlations between scores obtained on the locomotor skills and ball skills factors (r = .82-.95), thus again calling into question the relevance of the TGMD-3 two-factor structure. Additional results obtained in these studies also revealed that: (a) older children tended to score significantly higher than younger children on the locomotor skills and ball skills factors (Estevan et al., 2017; Mohammadi et al., 2019; Wagner et al., 2017), as well as on the global motor skills factor (Estevan et al., 2017; Mohammadi et al., 2019), and that (b) boys tended to score significantly higher than girls on the ball skills factor (Mohammadi et al., 2019; Wagner et al., 2017) and on the global motor skills factor (Mohammadi et al., 2019).

The ability of scores obtained on the TGMD-3 to retain their measurement properties (i.e., lack of differential item functioning, or measurement bias) irrespective of children main characteristics (i.e., age, body mass index, physical activity/sport involvement, sex) remain understudied. Only three of the aforementioned studies (Magistro et al., 2020; Valentini et al., 2017; Wagner et al., 2017) examined the measurement invariance of the TGMD-3 factor structure as a function of age and/or sex. Two of these studies (Magistro et al., 2020; Valentini et al., 2017) found evidence of weak invariance (i.e., equivalence of the factor loadings) of the two-factor solution of the TGMD-3 as function of children's age and sex. Another one (Wagner et al., 2017) found support for the weak invariance of the TGMD-3, but not for its strong invariance (i.e., equivalence of the response intercepts) as a function of children's sex. Furthermore, none of these studies has examined the presence of differential item functioning in

TGMD-3 ratings, and possible latent mean differences on the TGMD-3 factors, as a function of children's body mass index and involvement in physical activity/sport practice. This is surprising given that body massnindex (or weight status) and physical activity/sport practice are significantly associated with gross motor competence or fundamental movement skills among children (for systematic reviews and meta-analysis see Barnett et al., 2016; Cattuzzo et al., 2016; Holfelder & Schott, 2014; Logan et al., 2015; Lubans et al., 2010).

Indeed, results from previous systematic reviews or meta-analyses indicate that heavier youth (e.g., higher body mass index) tended to display lower levels of gross motor competence (Cattuzzo et al., 2016; Lubans et al., 2010), stability, or locomotor competencies (Barnett et al., 2016). However, no effect of weight status was found in relation to object control competence (Barnett et al., 2016). According to Barnett et al. (2016, p. 1682), this may be explained by the fact that in *"[...] object control skills, which tend to be more static, locomotor and stability skills involve shifting or controlling a larger body mass, which impedes functional movement [103] and contributes to the higher rate of lower limb problems among obese children (e.g., tibia varus, plantar pressure) [104]*."

Likewise, previous systematic reviews or meta-analyses also indicate that youth involved in physical activity (e.g., time spent in organized physical activity, body movement assessed by pedometers or accelerometers, moderate-vigorous physical activity) tended to display higher levels of gross motor competence, locomotor competencies, and object control competencies (Barnett et al., 2016; Holfelder & Schott, 2014; Logan et al., 2015; Lubans et al., 2010). However, these interpretations all rely on the assumption that there is no evidence of differential item functioning on motor skills ratings in relation to these youth characteristics (i.e., that comparison can be conducted in an unbiased manner), which has never before been systematically evaluated. More precisely, it thus remains unknown whether the specifics (i.e., locomotor and ball skills) and/or the global factors of the TGMD-3 could be used to reliably assess, and compare, the motor skills of children with different body mass index and levels of physical activity/sport practice. Then, assuming a lack of measurement bias, it would also be informative to verify whether these previously reported motor skills differences in relation to weight status and physical activity/sport practice would be replicated with the TGMD-3.

Study's Objectives

To our knowledge, the psychometric properties of a French-Canadian version of the TGMD-3 have yet to be examined. A French-Canadian version of the TGMD-3 would greatly facilitate the development of cross-cultural studies in French speaking countries (e.g., Belgium, Canada, France, Morocco, Switzerland, Tunisia, etc.), as well as the comparison with children speaking other languages or from other cultures. Therefore, the main objective of this study was to investigate the psychometric properties of scores obtained on a French-Canadian version of the TGMD-3 among a sample of French-speaking Canadian children. First, we examined the factor structure of scores on the TGMD-3 using bifactor-CFA and ESEM. Second, the presence of differential item functioning and latent mean differences on scores obtained on the TGMD-3 were examined as a function of children's age, body mass index, weekly frequency of physical activity/sport practice and sex.

Method

Participants

Participants were 127 French-speaking Canadian children recruited in four primary schools located in the Canadian Province of Québec. As illustrated in Table 1, this sample was almost evenly composed of boys (55.1%) and girls (44.9%), aged from 5 to 11 years old. The majority of these participants had a normal weight (73.4%), based on the sex- and age-specific body mass index revised cut-off scores provided by the International Obesity Task Force (Cole & Lobstein, 2012), and were involved in a physical activity/sport practice outside of school (75.4%) at least 1.5 times per week.

Measures

Children's characteristics. Parents or legal representatives reported children's age and sex. In addition, following a common practice in research involving children, parents or legal representatives were asked to report their children's height and weight (Ghosh-Dastidar et al., 2016; Shields et al., 2011). This information was then used to estimated children's body mass index [(Weight/(Height^2)] in kg/m^2. Finally, also in line with previous research (for a review see Chinapaw et al., 2010), parents or legal representatives were asked to report their children's weekly frequency of physical activity/sport practice outside of the school context (i.e., "*Does your child practice a physical activity or a sport outside of school?*"; If yes, "*Overall, how many times per week does he/she practice this physical* *activity or sport outside of school?*"). These questions were specifically developed for this study but are highly similar to those used in previous research.

TGMD-3. Two independent professional bilingual translators and four members of the research team were involved in the development of the French-Canadian version of the TGMD-3. To ensure the generalizability of this version, this team included people born in France and in the French part of Canada. This version was developed using standardized translation back-translation techniques (Hambleton, 2005). A first professional bilingual translator (not familiar with the TGMD-3) translated the original English items into a universal French language. These items were then back-translated into English by a second independent professional bilingual translator that was also not aware of the original English items. Finally, the back-translated items of the TGMD-3 were compared with the original English items and any discrepancies were resolved in a committee including both translators and four members of the research team. The testing material of the TGMD-3 is copyrighted and cannot be reproduced. Authors seeking to obtain a complete copy of the testing material for research purpose should directly contact the first author.

Procedures

This research was authorized by the research ethics committee of the first author's University and by the school board of the participating primary schools. Parents/legal representatives were informed by letter of the study. Although the TGMD-3 can be used to assess motor skills competence among typically developing children as well as children presenting a developmental delay, the present study more specifically focused on typically developing children to initially assess the psychometric properties of this instrument among a more "normative" population. As such, parents were informed that children presenting a characteristic likely to limit their ability to complete the TGMD-3 (i.e., if they needed assistance to move, presented a neurological disorder or developmental delay, or presented a sensory or physical disabilities) were not eligible for this study. Parents/legal representatives where then asked to sign an informed consent form, whereas participating children were asked to actively and verbally indicate their agreement to participate to the study. Participating children were eligible for a draw of 20 gift certificates (each with a value of $20 CAD).

Parents/legal representatives were asked to complete a series of questions about their child's characteristics. The TGMD-3 was administered, at school, by two members of the research team. Before testing, the administrators trained together using online videos available at the TGMD-3's website (https://sites.google.com/a/umich.edu/tgmd-3/home). Testing was realized in groups of 2 children. During the administration, each skill was first demonstrated by one of the administrators. The child was then invited by the administrators to practice the skill one time, followed by the two official trials. All trials were coded live, separately and simultaneously, by the two administrators. After each trial, the administrators compared and discussed their assessment and a final score for each performance criteria was attributed to the participant. It is this final score that is used in our analyses.

Analyses

Interrater reliability was examined using the initial scores provided for each trial separately by the two administrators, through the estimation of intra-class correlation (ICC) coefficients and of their 95% confidence interval (Hallgren, 2012). More specifically, ICC were estimated for each raw skill scores and factor scores using a two-way mixed, consistency and average-measures configuration. According to Hallgren (2012), ICC values are considered as good to excellent when they respectively fall between .60 and .74 (good), or between .75 and 1.00 (excellent).

The factor analyses of the TGMD-3 were conducted using Mplus 8.4's (Muthén & Muthén, 2019) robust weighted least squares estimator with mean and variance adjusted statistics (WLSMV) using the raw score obtained on each skill. This estimator is best suited to the ordinal nature, and asymmetric response thresholds, of the TGMD-3 scores (Finney & Di Stephano, 2013). The few missing responses at the item level (0.76%; M = 0.06%) were handle using algorithms implemented in Mplus in conjunction with the WLSMV estimator (Asparouhov & Muthén, 2010). First, the *a priori* two-factor model measuring ball skills (7 skills) and locomotor skills (6 skills) was examined using correlated factors CFA and ESEM solutions. The bifactor-CFA and bifactor-ESEM representations of the TGMD-3 corresponded to their correlated-factors counterparts by included one more factor defined by all items. More precisely, in these solutions, all items were specified as having a main loading on both a global motor skills factor (G-factor) and on their *a priori* specific factors (S-factors; ball skills and locomotor skills). Matching typical bifactor solutions (Morin et al., 2016; Reise, 2012), all factors were specified as orthogonal Moreover, a one-factor solution was also examined.

The ESEM and bifactor-ESEM solutions relied on a confirmatory approach to factor rotation named target (Asparouhov & Muthén, 2009). In the ESEM solution, an oblique target rotation procedure was utilized (Browne, 2001), whereas an orthogonal bifactor-target rotation (Reise, 2012) was used for the bifactor-ESEM solution. In both of these models, all cross-loadings between the ESEM factors, or between the bifactor-ESEM S-factors, were freely estimated by "targeted" to be as close to zero as possible, using the most commonly applied form of target rotation (Morin et al., 2020). Although it is theoretically possible to improve the accuracy of the rotation procedures by relying on more informed targets (i.e., specifying the expected values of the target loadings and cross-loadings; Myers et al., 2013, 2015), information was lacking to rely on these more informed specifications in the present study. Recent statistical simulation studies have supported the robustness of target rotations procedures when compared to other types of rotations (Myers et al., 2015), while also highlighting the risks associated with using erroneous informed target values (Guo et al., 2019), which supported the present, and more generic, approach.

The goodness-of-fit of this models was examined using (e.g., Hu & Bentler, 1999; Marsh et al., 2005; Yu, 2002): The comparative fit index (CFI $\geq$ .90 and >. 95), the Tucker-Lewis index (TLI; same thresholds as for the CFI), and the root mean square error of approximation (RMSEA $\leq$.08 and $\leq$. 06). Despite that the WLSMV chi-square test of exact fit (Wχ^2) was reported it has not been considered for model fit given that it is over sensitive to sample size and minor misspecifications (e.g., Hu & Bentler, 1999; Marsh et al., 2005). For all models, the composite reliability of the TGMD-3 factors was estimated using McDonald's (1970) omega (ω) coefficient.

As noted by Morin et al. (2016, 2020) goodness-of-fit assessment is insufficient to guide model selection when CFA, ESEM, bifactor-CFA and bifactor-ESEM solutions are contrasted. These authors recommend a careful examination of the parameter estimates (i.e., loadings, cross-loadings, latent correlations, composite reliability) obtained from these various models. This examination should start with a comparison of the CFA and ESEM models, where the observation of reduced factor correlations in ESEM coupled with generally well-defined factors could be taken as evidence in favor of the ESEM solution over a similarly fitting CFA solution. Then, the retained model should be contrasted to its bifactor counterpart. In this second comparison, the observation of a well-defined G-factor coupled with at least a subset of well-defined S-factor could be taken as evidence supporting a bifactor solution over a similarly fitting first-order solution. The model providing the most optimal representation of the data on the basis of this examination was retained for subsequent analyses.

The presence of measurement bias in the TGMD-3 as function of children's age, body mass index, week frequency of physical activity/sport practice, and sex was examined using tests of differential item functioning and latent mean differences. These tests were conducted separately for children's characteristics (i.e., age, body mass index, weekly frequency of physical activity/sport practice, and sex) using a multiple indicators multiple causes approach. As recommended by Marsh et al. (2013) and Morin et al. (2013) the following model were examined: (a) null effects (the paths from the predictors to the TGMD-3 latent factors and raw skill scores were constrained to be zero); (b) saturated (the paths from the predictors to the raw skill scores were freely estimated, while the paths from the predictors to the latent factors were constrained to be zero); and (c) factors-only (the paths from the predictors to the TGMD-3 latent factors were freely estimated, while the paths from the predictors to the raw skill scores were constrained to be zero). Improvement in fit (ΔCFIs/TLIs $\geq$ .01 and ΔRMSEAs $\geq$. 015) between the factors-only models and the saturated models relative to the null effects model suggest the presence of associations between predictors and the TGMD-3 latent factors and raw skill scores. However, improvement in fit between the saturated model relative to the factors-only model indicates differential item functioning.

Results

Descriptive Statistics and Interrater Reliability of TGMD-3's Skills and Subscales

Descriptive statistics and interrater reliability coefficients for scores obtained in the TGMD-3 skills and subscales are reported in Table 2. These results strongly support the interrater reliability (with all interrater reliability coefficients >.98) of scores obtained on the global (motor skills), and subscale-specific (locomotor skills and ball skills) dimensions of the TGMD-3.

Factor Validity and Reliability

The goodness-of-fit indices of all measurement models are reported in the top section of Table 3.

The one-factor solution resulted in a poor (CFI-TLI< .90, RMSEA > .08) level of fit to the data, whereas the two-factor CFA and the bifactor-CFA solutions resulted in an acceptable level of fit to the data (CFI-TLI > .90, RMSEA < .08). Additionally, the ESEM and the bifactor-ESEM solutions resulted in an excellent level fit to the data (CFI-TLI > .95, RMSEA ≤.06), and their level of fit was substantially improved relative to that of their CFA (ΔCFI = +.034; ΔTLI = +.031; ΔRMSEA= -.014) or bifactor-CFA (ΔCFI = +.050; ΔTLI = +.072; ΔRMSEA= -.043) counterparts. Likewise, the bifactor-CFA (ΔCFI = +.016; ΔTLI = +.004; ΔRMSEA= -.001) and most importantly the bifactor-ESEM (ΔCFI = +.032; ΔTLI = +.045; ΔRMSEA= -.030) solutions both resulted in an improved level of fit relative to their correlated factors counterparts. Taken together, these results thus seem to support the bifactor-ESEM solution. However, following Morin et al.'s (2016) recommendations, the selection of the optimal solution needs to be based on a careful examination of parameter estimates. To this end, CFA and ESEM solutions were first contrasted, and then the most optimal of these representations with its bifactor counterpart.

The detailed parameter estimates from the CFA, ESEM, and bifactor-CFA solutions are respectively reported in Tables S1, S2, and S3 of the online supplements, whereas those from the bifactor-ESEM solution are reported in Table 4. In the CFA solution, the factor loadings of the TGMD-3 (λ = .447-.719, M_{λ} = .609) were all satisfactory and associated with acceptable composite reliability coefficients (ω = .835 and .739, M_{ω} = .787). However, the factor correlations remained high enough (r = .616) to suggest some level of conceptual overlap between the two factors. In contrast, this factor correlation was substantially reduced in the ESEM solution (r = .391), suggesting that the slight level of overlap could in fact be explained by the presence of cross-loadings ($|\lambda|$ = .013-.398, $M_{|\lambda|}$ = .171) rather than true conceptual overlap at the factor level. More precisely, although most cross-loadings remain negligible, the results suggest that some skills tended to present a similar level of association with both types of skills (i.e., skipping and sliding), suggesting that they might tap into some similar motor processes. Finally, the ESEM solutions also resulted in generally well-defined (λ = .253-.844, M_{λ} = .577) and reliable latent factors (ω = .838 and .712, M_{ω} = .775).

For this reason, the ESEM solution was favored, and contrasted with its bifactor-ESEM counterpart. This bifactor-ESEM solution resulted in a well-defined and reliable G-factor (λ = .097-.811, M_{λ} = .477; ω = .849) reflecting participants' global motor skills. The results revealed that the two S-factors also remained reasonably well-defined (ball skills: λ = -.043-.542, M_{λ} = .327, ω = .620; locomotor skills: λ = .260-.756, M_{λ} = .474, ω = .691), although slightly weaker than the G-factor. Furthermore, most cross-loadings were reduced in the bifactor-ESEM (relative to the ESEM) solution, which also revealed that the items with the highest cross-loadings in the ESEM solution best serve to define the G-factor (rather than the S-factors) in the bifactor-ESEM solution. This is consistent with the idea that these skills might more directly tap into global motor skills than into specific locomotor skills or specific ball skills. Therefore, the present results support the bifactor-ESEM representation of the data, which resulted in (Morin et al., 2020): (a) an improved level of fit to the data compared to all other solutions (CFA, bifactor-CFA, ESEM); (b) a well-defined global motor skills G-factor; and (c) the presence of reasonably well-defined S-factors reflecting ball skills and locomotor skills.

Differential Item Functioning and Latent Mean Differences

The results from the models used to assess differential item functioning are reported in the bottom section of Table 3. For age, body mass index, physical activity/sport practice, and sex, the results showed that both the saturated and the factors-only models resulted in a substantial improvement in model fit relative to the null effects model. Furthermore, for all of these characteristics, the factors-only and saturated models resulted in a comparable level of fit (ΔCFI/ΔTLI ≤ .01 and ΔRMSEA ≤ .015). These results are consistent with a lack of differential item functioning, but with the presence of meaningful associations between scores on the TGMD-3 and children's age, body mass index, physical activity/sport practice, and sex. More precisely, the results showed that: (a) older children tended to score lower on both S-factors (ball skills: -.579, p = .004; locomotor skills: -.589, p = .002) and higher on the G-factor (.637, p < .001) relative to younger children; (b) children with a higher body mass index tended to score lower on the locomotor skills S-factor (-.234, p = .007) relative to children with a lower body mass index; (c) children with higher levels of physical activity/sport practice tended to score higher on the G-factor (.284, p < .001) relative to children with lower levels of physical activity/sport practice; and (d) boys tended to score higher on the ball skills S-factor (.547, p = .001) and lower on the locomotor skills S-factor (-.455, p = .001) relative to girls.

Discussion

The main objective of the study was to assess the psychometric properties of scores obtained on a French-Canadian version of the TGMD-3 among a sample of French-speaking Canadian children. To achieve this objective, we first examined the factor structure of scores obtained on the TGMD-3 using a combination or bifactor and ESEM analyses. Similar to Garn and Webster (2018), who retained an ESEM solution in their analyses of scores obtained on the TGMD-2 among a sample of US children, our results also supported the superiority of an ESEM representation of scores obtained on the TGMD-3. Moreover, this incorporation of cross-loadings (via ESEM) resulted in a much more marked decrease in the size of the factor correlations relative to those obtained in a CFA solution (r = .616 to .391) than the decrease (r = .891 to .810) reported by Garn and Webster (2018), suggesting that scores on the TGMD-3 ball skills and locomotor skills factors might be more independent from one another than scores obtained on the TGMD-2, although part of this difference could also be linked to the linguistic version considered in both studies (English-US in Garn and Webster, 2018; French-Canadian in the present study). Despite these differences, both studies highlight the importance to incorporate cross-loadings in order to obtain a more accurate estimate of factor correlations, something which has already been documented in statistical research (Asparouhov et al., 2015).

Furthermore, and supporting the results reported by Garn and Webster (2021) among a sample of US children in relation to the TGMD-3, the present results also supported the importance of incorporating a bifactor component (i.e., bifactor-ESEM) to the measurement structure of scores obtained on this instrument. More specifically, our results revealed that children’s scores on the TGMD-3 can be used to reflect their global levels of motor skills. Additionally, beyond this global factor, the results also suggest that enough specificity remained at the subscale level to also estimate children’s specific levels of ball skills and locomotor skills beyond their global motor skills. In consequence, the present results revealed that the factor structure of scores obtained on the TGMD-3 seem to follow a hierarchical (global *vs.* specific) and multidimensional (ball skills *vs.* locomotor skills) structure, confirming thus that a global factor measuring gross motor skills could coexist with locomotor skills and ball skills factors.

Another objective of this study was to examine differential item functioning and latent mean differences of the TGMD-3 as a function of children’s characteristics frequently considered in research as being linked to motor skills development. In this regard, our results first revealed a lack of differential item functioning in TGMD-3 ratings as a function of children’s age and sex, in accordance with the results from previous psychometric investigations of scores obtained on other linguistic versions of the TGMD-3 (Magistro et al., 2020; Valentini et al., 2017; Wagner et al., 2017). Subsequent analyses also revealed meaningful latent mean differences on the TGMD-3 factors as function of children’s age and sex. More precisely, in accordance with previous studies (Estevan et al., 2017; Mohammadi et al., 2019) the results showed that older children tended to score significantly higher on the global TGMD-3 factor relative to younger children. However, the latent mean differences found in relation to scores on the specific factors of the TGMD-3 were in the opposite direction of those found in previous studies (Estevan et al., 2017; Mohammadi et al., 2019; Wagner et al., 2017), showing that older children tended to score lower than younger children on both S-factors and that boys, relative to girls, tended to score higher on the ball skills S-factor but lower on the locomotor skills S-factor.

Nevertheless, given our reliance on a bifactor representation of scores obtained on the TGMD-3, our results related to these S-factors cannot be directly compared to those obtained in previous studies relying on observed scores, or on CFA or ESEM solutions excluding a bifactor component (Morin et al., 2020). Indeed, whereas factors from a correlated factors model reflect the covariance between the items forming a subscale, the S-factors from a bifactor model reflect the residual covariance between these items once the covariance between all items has been absorbed by the G-factor. Thus, rather than reflecting the net ball skills and locomotor skills manifested by the children, these S-factors rather reflect the extent to which these skills deviate from their global levels of motor skills. The present results provide a vivid illustration of this difference by revealing associations between age and scores in these S-factors that are in the opposite direction of association typically reported in relation to scores on these complete subscales. In this regard, it is important to note that children’s scores on the TGMD-3 global motor skills G-factor did show an increase as a function of age, in accordance with the results from these previous studies. Thus, what the current results suggest is not only that older children tend to present higher global motor skills scores, but also that their levels of locomotor and ball skills tend to become increasingly differentiated from one another as children get older (leading to an increase in the scores obtained on the S-factors, which reflect the extent to which each specific skill deviates from children's global motor skills). Given that it is the first study to examine these differences while relying on a bifactor model the present results should be considered with caution pending their replication.

Finally, to our knowledge, the present study was the first to test for evidence of differential item functioning and latent mean differences in TGMD-3 ratings as a function of children's body mass index and physical activity/sport practice. In this regard, our results first revealed a lack of differential item functioning between TGMD-3 scores and children's body mass index and physical activity/sport practice. In addition, our results also revealed latent mean differences on the TGMD-3 latent factors in relation to these two characteristics. More precisely, and matching the results obtained in previous studies (for systematic reviews or meta-analysis see Barnett et al., 2016; Holfelder & Schott, 2014; Logan et al., 2015), our results showed that children with a higher body mass index tended to score lower on the locomotor skills S-factor relative to children with a lower body mass index, and that children with a higher weekly frequency of physical activity/sport practice outside of school tended to score significantly higher on the global motor skills factor relative to children with a lower weekly frequency of physical activity/sport practice.

The present study has five main limitations that should be considered when interpreting the results. First, the current study relies on a small convenience sample of children where three-fourths of the participants had a normal weight and were involved in physical activity/sport outside school. Therefore, these results cannot be generalized to more representative and diversified samples of children. It would be important for future studies to assess whether and how the current results generalize to children presenting developmental delays, neurological disorders, or other forms of disabilities. In addition, as recommended by Myers et al. (2016, 2018) power estimation based on the model data fit from the present study sample was performed on an online calculator developed by Preacher and Coffman (2006). The following parameter were selected: $\alpha = .05$, df = 42, sample size = 127, Null RMSEA = .08, Alt. RMSEA = .033. Results revealed a power of .715, and indicated that 20 more participants would be necessary to reach a power of .80. Consequently, this lower power should be considered when interpreting the present results. Second, the psychometric equivalence of scores obtained on the French-Canadian version of the TGMD-3 in relation to other linguistic versions of the TGMD-3 (such as the original English version) was not examined in the present study. Therefore, it is currently unknown whether the current results would generalize to samples of children from other cultures (European and North-African) or languages (e.g., English, Italian, German, Persian, Portuguese, Spanish, etc.). Likewise, although the current version was developed to be suitable for any French-speaking populations, its generalizability outside of the French-Canadian context remains to be systematically assessed. Third, children's height and weight were not directly measured, but reported by their parents or legal representatives. Therefore, as noted in previous research (e.g., Ghosh-Dastidar et al., 2016; Shields et al., 2011) the BMI values used in the present study might have been slightly underestimated. Future research should thus examine the extent to which our results would generalize to objective measures of BMI. Fourth, the convergent validity of scores obtained on the current French-Canadian version of the TGMD-3 was not examined in this study. Therefore, future research should assess their convergent validity in relation to other measures of objective (e.g., Barnett & Ulrich, 2021) or perceived (e.g., van Veen et al., 2020) motor skills competence. Finally, scores obtained on our measure of physical activity/sport practice have never been psychometrically validated. As a result, the present results may have been biased by our reliance on a new measure, which cannot really be considered to be fully representative of children's level of physical activity, or of their level of involvement in moderate to vigorous physical activity. It is thus important to consider replicating the present results using validated measures of physical activity/sport practice.

In conclusion, results from the current study provided preliminary support to the psychometric properties of scores obtained on the French-Canadian version of the TGMD-3 and suggested that this test can be confidently used to measure, and compare, gross motor skills among children irrespective of their age, sex, body mass index, and level of physical activity/sport involvement. However, given that the psychometric properties of scores obtained on the French-Canadian version of the TGMD-3 were not examined as function of other language/culture or time (e.g., test-retest reliability, longitudinal stability, and longitudinal invariance) it remains premature to recommend its use in cross-linguistic/culture or longitudinal studies. These psychometric properties should thus be more thoroughly investigated in future research. Furthermore, in practice, the TGMD-3 is scored manually. However, the use of summed scores to estimate the global factor and specific factors of the TGMD-3 is inconsistent with the bifactor structure of the TGMD-3 (Brown et al., 2011; Maïano et al., 2021). Practically, two solutions are offered (Maïano et al., 2019, 2021; Perreira et al., 2018): (1) relying on bifactor-ESEM latent variable models as in the present study, which remains the optimal approach for research purposes; and (2) using the *Mplus's* FSCORE algorithm to obtain scores on the global motor skills factor and specific ball skills and locomotor skills factors. Pending the development of computerized scoring algorithm, these two approaches remain the only way to account for the TGMD-3 bifactor structure.

References

Asparouhov, T., & Muthén, B. (2010). Weighted least squares estimation with missing data. *Mplus Technical Appendix*, Los Angeles, CA: Muthén & Muthén.

Asparouhov, T., & Muthén, B.O. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16*, 397-438. https://doi.org/10.1080/10705510903008204

Asparouhov, T., Muthén, B.O., & Morin, A.J.S. (2015). Bayesian Structural equation modeling with cross-loadings and residual covariances. *Journal of Management*, *41*, 1561-1577.

Barnett, L.M., Lai, S.K., Veldman, S.L., Hardy, L.L., Cliff, D.P., Morgan, P.J., ... & Okely, A.D. (2016). Correlates of gross motor competence in children and adolescents: A systematic review and meta-analysis. *Sports Medicine*, *46* 1663-1688. https://doi.org/10.1007/s40279-016-0495-z

Barnett, L.M., & Ulrich, B.D. (2021). The Michigan State University Motor Performance study, look to the past to shed light on the future. *Measurement in Physical Education & Exercise Science, 25*, 1-6. https://doi.org/10.1080/1091367X.2020.1860993

Brown, A., Finney, S., & France, M. (2011). Using the bifactor model to assess the dimensionality of the Hong Psychological Reactance Scale. *Educational & Psychological Measurement*, *71*, 170-185. https://doi.org/10.1177/0013164410387378

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*, 111–150. https://doi.org/10.1207/S15327906MBR3601_05

Cattuzzo, M. T., dos Santos Henrique, R., Ré, A. H. N., de Oliveira, I. S., Melo, B. M., de Sousa Moura, M., ... & Stodden, D. (2016). Motor competence and health related physical fitness in youth: A systematic review. *Journal of Science and Medicine in Sport*, *19*, 123-129. http://doi.org/10.1016/j.jsams.2014.12.004

Chinapaw, M. J., Mokkink, L. B., van Poppel, M. N., van Mechelen, W., & Terwee, C. B. (2010). Physical activity questionnaires for youth. *Sports Medicine*, *40*(7), 539-563. https://doi.org/10.2165/11530770-000000000-00000

Cole, T.J., & Lobstein, T. (2012). Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. *Pediatric Obesity*, *7*, 284-294. https://doi.org/10.1111/j.2047-6310.2012.00064.x

Cools, W., De Martelaer, K., Samaey, C., & Andries, C. (2009). Movement skill assessment of typically developing preschool children: A review of seven movement skill assessment tools. *Journal of Sports Science & Medicine*, *8*, 154-168.

Eddy, L.H., Bingham, D.D., Crossley, K.L., Shahid, N.F., Ellingham-Khan, M., Otteslev, A., ... & Hill, L.J. (2020). The validity and reliability of observational assessment tools available to measure fundamental movement skills in school-age children: A systematic review. *PloS One*, *15*, e0237919. https://doi.org/10.1371/journal. pone.0237919

Estevan, I., Molina-García, J., Queralt, A., Álvarez, O., Castillo, I., & Barnett, L.M. (2017). Validity and reliability of the Spanish version of the test of gross motor development–3. *Journal of Motor Learning and Development*, *5*, 69-81. https://doi.org/10.1123/jmld.2016-0045

Finney, S.J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In G.R. Hancock & R.O. Mueller (Eds.), Quantitative methods in education and the behavioral sciences: Issues, research, and teaching. *Structural equation modeling: A second course* (pp. 439-492). IAP Information Age Publishing.

Garn, A. C., & Webster, E. K. (2018). Reexamining the factor structure of the test of gross motor development – second edition: Application of exploratory structural equation modeling. *Measurement in Physical Education and Exercise Science*, *22*, 200-212 https://doi.org/10.1080/1091367X.2017.1413373

Garn, A.C., & Webster, E.K. (2021). Bifactor structure and model reliability of the Test of Gross Motor Development—3rd edition. *Journal of Science and Medicine in Sport*, *24*, 67-73. https://doi.org/10.1016/j.jsams.2020.08.009

Griffiths, A., Toovey, R., Morgan, P.E., & Spittle, A.J. (2018). Psychometric properties of gross motor assessment tools for children: a systematic review. *BMJ Open*, *8*, e021734. http://dx.doi.org/10.1136/bmjopen-2018-021734

Guo, J., Marsh, H.W., Parker, P.D., Dicke, T., Lüdtke, O., & Diallo, T.M.O. (2019). A systematic evaluation and comparison between exploratory structural equation modeling and Bayesian structural equation modeling. *Structural Equation Modeling, 26,* 529-556. https://doi.org/10.1080/10705511.2018.1554999

Ghosh-Dastidar, M., Haas, A.C., Nicosia, N., & Datar, A. (2016). Accuracy of BMI correction using multiple reports in children. *BMC Obesity*, *3*, 37. https://doi.org/10.1186/s40608-016-0117-1

Hallgren, K.A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*, 23. https://doi.org/ 10.20982/tqmp.08.1.p023

Hambleton, R.K. (2005). Issues, designs, and technical guidelines for adapting tests to languages and cultures. In R.K. Hambleton, P. Merenda, & C. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3-38). Mahwah, NJ: Erlbaum.

Holfelder, B., & Schott, N. (2014). Relationship of fundamental movement skills and physical activity in children and adolescents: A systematic review. *Psychology of Sport and Exercise*, *15*, 382-391. https://doi.org/10.1016/j.psychsport.2014.03.005

Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55. https://doi.org/10.1080/10705519909540118

Hulteen, R.M., Barnett, L.M., True, L., Lander, N.J., del Pozo Cruz, B., & Lonsdale, C. (2020). Validity and reliability evidence for motor competence assessments in children and adolescents: A systematic review. *Journal of Sports Sciences, 38,* 1717-1798. https://doi.org/10.1080/02640414.2020.1756674

Logan, S.W., Webster, K.E., Getchell, N., Pfeiffer, K.A., & Robinson, L.E. (2015). Relationship between fundamental motor skill competence and physical activity during childhood and adolescence: A systematic review. *Kinesiology Review, 4*, 416-426. https://doi.org/10.1123/kr.2013-0012

Lubans, D.R., Morgan, P.J., Cliff, D.P., Barnett, L.M., & Okely, A.D. (2010). Fundamental movement skills in children and adolescents. *Sports Medicine*, *40*, 1019-1035. https://doi.org/10.2165/11536850-000000000-00000

Maïano, C., Aimé, A., Lepage, G., ASPQ Team, & Morin, A.J.S. (2019). Psychometric properties of the Weight Self-Stigma Questionnaire (WSSQ) among a sample of overweight/obese French-speaking adolescents. *Eating and Weight Disorders – Studies on Anorexia, Bulimia and Obesity*, *24*, 575-583. https://doi.org/10.1007/s40519-017-0382-0

Maïano, C., Morin, A.J.S., Aimé, A. Lepage, G., & Bouchard, S. (2021). Psychometric Properties of the Body Checking Questionnaire (BCQ) and of the Body Checking Cognitions Scale (BCCS): A bifactor-exploratory structural equation modeling approach. *Assessment*, *28*, 632-646. https://doi.org/10.1177/1073191119858411

Magistro, D., Piumatti, G., Carlevaro, F., Sherar, L.B., Esliger, D. W., Bardaglio, G., ... & Musella, G. (2020). Psychometric proprieties of the Test of Gross Motor Development–Third Edition in a large sample of Italian children. *Journal of Science and Medicine in Sport*. https://doi.org/10.1016/j.jsams.2020.02.014

Marsh, H.W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares, & J. McArdle (Eds.), *Contemporary psychometrics* (pp. 275-340). Hillsdale, NJ: Erlbaum.

Marsh, H.W., Nagengast, B., & Morin, A.J.S. (2013). Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. *Developmental Psychology*, *49*, 1194-1218. https://doi.org/10.1037/a0026913

McDonald, R.P. (1970). Theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, *23*, 1–21.

https://doi.org/10.1111/j.2044-8317.1970.tb00432.x
Mohammadi, F., Bahram, A., Khalaji, H., Ulrich, D.A., & Ghadiri, F. (2019). Evaluation of the Psychometric Properties of the Persian Version of the Test of Gross Motor Development–3rd Edition. *Journal of Motor Learning and Development,* 7, 106-121. https://doi.org/10.1123/jmld.2017-0045
Morin, A. J. S., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling, 23*, 116-139. https://doi.org/10.1080/10705511.2014.961800
Morin, A.J.S., Marsh, H.W., & Nagengast, B. (2013). Exploratory structural equation modeling. In G.R. Hancock & R.O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 395-438). Charlotte, NC: Information Age.
Morin, A.J.S., Myers, N.D., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, exploratory structural equation modeling (ESEM) and bifactor-ESEM. In G. Tenenbaum & R.C. Eklund (Eds.), *Handbook of sport psychology, 4th Ed*. (pp. 1044-1073). London: Wiley
Muthén, L.K., & Muthén, B. (2019). *Mplus user's guide.* Los Angeles, CA: Muthén & Muthén.
Myers, N.D., Ahn, S., & Jin, Y. (2013). Rotation to a partially specified target matrix in exploratory factor analysis: How many targets? *Structural Equation Modeling, 20*, 131-147. doi: 10.1080/10705511.2013.742399
Myers, N.D., Martin, J.J., Ntoumanis, N., Celimli, S., & Bartholomew, K.J. (2014). Exploratory bi-factor analysis in sport, exercise, and performance psychology: A substantive-methodological synergy. *Sport, Exercise, and Performance Psychology, 3*, 258-272. https://doi.org/10.1037/spy0000015
Myers, N.D., Ntoumanis, N., Gunnell, K.E., Gucciardi, D.F., & Lee, S. (2018). A review of some emergent quantitative analyses in sport and exercise psychology. *International Review of Sport and Exercise Psychology*, *11*, 70-100. https://doi.org/10.1080/1750984X.2017.1317356
Myers, N.D., Jin, Y., Ahn, S., Celimli, S., & Zopluoglu, C. (2015). Rotation to a partially specified target matrix in exploratory factor analysis in practice. *Behavior Research Methods*, *47*, 494-505. https://doi.org/10.3758/s13428-014-0486-7
Myers, N.D., Park, S.E., Lefevor, G.T., Dietz, S., Prilleltensky, I., & Prado, G.J. (2016). Measuring multidimensional subjective well-being with the I COPPE scale in a Hispanic sample. *Measurement in Physical Education and Exercise Science, 20*, 230-243. https://doi.org/10.1080/1091367X.2015.1111220
Perreira, T.A., Morin, A.J.S., Hebert, M., Gillet, N., Houle, S.A., & Berta, W. (2018). The short form of the Workplace Affective Commitment Multidimensional Questionnaire (WACMQ-S): A bifactor-ESEM approach among healthcare professionals. *Journal of Vocational Behavior, 106,* 62-83. https://doi.org/10.1016/j.jvb.2017.12.004
Preacher, K. J., & Coffman, D. L. (2006, May). Computing power and minimum sample size for RMSEA [Computer software]. Available from http://quantpsy.org/.
Rey, E., Carballo-Fazanes, A., Varela-Casal, C., Abelairas-Gómez, C., & ALFA-MOV Project collaborators. (2020). Reliability of the test of gross motor development: A systematic review. *PloS One, 15,* e0236070. https://doi.org/10.1371/journal.pone.0236070
Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47,* 667-696. https://doi.org/10.1080/00273171.2012.715555
Scheuer, C., Herrmann, C., & Bund, A. (2019). Motor tests for primary school aged children: A systematic review. *Journal of Sports Sciences, 37,* 1097–1112. https://doi.org/10.1080/02640414.2018.1544535
Shields, M., Gorber, S. C., Janssen, I., & Tremblay, M. S. (2011). Obesity estimates for children based on parent-reported versus direct measures. *Health Reports, 22,* 47-58. https://doi.org/10.1016/S1499-2671(11)52024-6
Ulrich, D.A. (1985). *Test of Gross Motor Development.* Austin, Pro-ED.
Ulrich, D.A. (2000). *Test of Gross Motor Development (2nd Edition).* Austin, Pro-ED.
Ulrich, D.A. (2019). *Test of Gross Motor Development (3rd Edition).* Austin, Pro-ED.
Valentini, N.C., Zanella, L.W., & Webster, E.K. (2017). Test of Gross Motor Development—Third edition: Establishing content and construct validity for Brazilian children. *Journal of Motor*

Learning and Development, *5*, 15-28. https://doi.org/10.1123/jmld.2016-0002

van Veen, C., Schott, N., Lander, N., Tietjens, M., Hinkley, T., Dreiskämper, D., . . . & Barnett, L. M. (2020). The stability of perceived motor competence of primary school children from two countries over one year. *Measurement in Physical Education and Exercise Science, 24*, 74-80. doi:10.1080/1091367X.2019.1675665

Wagner, M.O., Webster, E.K., & Ulrich, D. (2017). Psychometric properties of the Test of Gross Motor Development, (German translation): Results of a pilot study. *Journal of Motor Learning and Development*, *5*, 29-44. https://doi.org/10.1123/jmld.2016-0006

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, 45-58. https://doi.org/10.1123/jmld.2016-0003

Yu, C.Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Los Angeles, CA: University of California.

Table 1

Descriptive Statistics of Participants

Characteristics of participants		M or %	SD	Range
Age (in years)		7.40	1.67	5-11
	5 years	12.7%	-	-
	6 years	23.0%	-	-
	7 years	21.4%	-	-
	8 years	14.3%	-	-
	9 years	15.1%	-	-
	10 years	10.3%	-	-
	11 years	3.2%	-	-
Sex				
	Boys	55.1%	-	-
	Girls	44.9%	-	-
Body mass index (in kg/m^2)		15.98	2.23	9.61-23
Body mass index *categories*				
	Underweight	15.6%	-	-
	Normal weight	73.4%	-	-
	Overweight	9.2%	-	-
	Obese	1.8%	-	-
Physical activity/sport practice				
	Involved	75.4%	-	-
	Not involved	24.6%	-	-
	Weekly frequency	1.48	1.35	0-7

Notes. M = mean; SD = standard deviation.

Table 2

Descriptive Statistics and Interrater Reliability of TGMD-3's Skills and Subscales

Skills	Range	M	SD	ICC	95%CI
Run	0-8	7.11	1.59	.986	.981-.990
Gallop	0-8	5.09	2.33	.996	.995-.997
Hop	0-8	6.67	1.66	.993	.990-.995
Skip	0-6	4.14	1.81	.996	.995-.997
Horizontal jump	3-8	6.67	1.47	.986	.979-.990
Slide	1-8	7.28	1.23	.994	.991-.996
Total locomotor skills	21-46	36.92	6.04	.997	.995-.998
Two-hand strike of a stationary ball	3-10	8.14	1.75	.996	.994-.997
One-hand forehand strike of self-bounced ball	0-8	5.75	1.79	.993	.989-.995
One hand stationary dribble	0-6	4.78	1.83	.997	.996-.998
Two-hand catch	0-6	5.54	1.17	.982	.974-.987
Kick a stationary ball	2-8	7.17	1.23	.984	.977-.989
Overhand throw	0-8	6.44	1.97	.994	.991-.996
Underhand throw	2-8	6.86	1.48	.979	.971-.985
Total ball skills	17-54	44.68	7.39	.996	.995-.997
Total TGMD-3	50-100	81.60	11.30	.998	.997-.998

Notes. CI = confidence interval; ICC = intraclass correlations; M = mean; SD = standard deviation; TGMD-3 = Test of Gross Motor Development - Third edition.

Table 3

Goodness-of-Fit Statistics of Confirmatory Factor Analyses (CFA) and Exploratory Structural Equation Modeling (ESEM) for the TGMD-3

Models	N°	Description	χ^2 (df)	CFI	TLI	RMSEA	RMSEA 90% CI	CM	$\Delta W\chi^2$ (df)	ΔCFI	ΔTLI	ΔRMSEA
Measurement	1-1	CFA - 1 factor	156.853(65)*	.858	.830	.105	.085-.127	-	-	-	-	-
	1-2	CFA - 2 factors	112.698(64)*	.925	.908	.077	.053-.101	-	-	-	-	-
	1-3	BCFA - 2 factors	89.949(52)*	.941	.912	.076	.048-.102	-	-	-	-	-
	1-4	ESEM - 2 factors	79.778(53)*	.959	.939	.063	.031-.090	-	-	-	-	-
	1-5	BESEM - 2 factors	47.694(42)	.991	.984	.033	.000-.071	-	-	-	-	-
DIF: Age	2-1	Null effects	143.873(55)*	.865	.777	.113	.090-.136	-	-	-		-
	2-2	Saturated	43.934(42)	.997	.994	.019	.000-.064	2-1	75.98(13)*	+.132	+.217	-.094
	2-3	Factors-only	54.954(52)	.996	.992	.021	.000-.061	2-1	34.09(3)*	+.131	+.215	-.092
DIF: Body mass index	3-1	Null effects	71.840(55)	.974	.957	.049	.000-.078	-	-	-		-
	3-2	Saturated	48.290(42)	.990	.979	.034	.000-.072	3-1	21.19(13)	+.016	+.022	-.015
	3-3	Factors-only	55.255(52)	.995	.991	.022	.000-.062	3-1	8.93(3)	+.021	+.034	-.027
DIF: PA/sport frequency	4-1	Null effects	107.450(55)*	.921	.869	.087	.062-.111	-	-	-		-
	4-2	Saturated	47.273(42)	.992	.983	.031	.000-.070	4-1	45.90(13)*	+.071	+.114	-.056
	4-3	Factors-only	60.670(52)	.987	.977	.036	.000-.070	4-1	22.37(3)*	+.066	+.108	-.051
DIF: Sex	5-1	Null effects	157.574(55)*	.849	.750	.121	.099-.144	-	-	-		-
	5-2	Saturated	49.845(42)	.988	.975	.038	.000-.075	5-1	79.44(13)*	+.139	+.225	-.083
	5-3	Factors-only	57.768(52)	.991	.985	.030	.000-.066	5-1	40.50(3)*	+.142	+.235	-.091

Notes. χ^2 = robust weighed least square (WLSMV) chi-square; Δ = change from previous model; $\Delta W\chi^2$ = WLSMV chi square difference test (calculated with the Mplus DIFFTEST function); CFI = comparative fit index; CM = comparison model; df = degrees of freedom; DIF = differential item functioning; PA = physical activity; RMSEA = root mean square error of approximation; 90% CI = 90% confidence interval of the RMSEA; TGMD-3 = Test of Gross Motor Development - Third edition; TLI = Tucker-Lewis index. The fact that WLSMV χ^2 values are not exact, but "estimated" as the closest integer necessary to obtain a correct *p* value explains the fact that the χ^2 and the resulting CFI values can be non-monotonic with model complexity. *$p < .01$.

Table 4

Standardized Parameters Estimates from the Bifactor Exploratory Structural Equation Modeling Representation of the TGMD-3

Skills	Ball Skills (λ)	Locomotor Skills (λ)	G-factor	δ
Two-hand strike of a stationary ball	.324	*.144*	.596	.519
One-hand forehand strike of self-bounced ball	.532	*-.026*	.542	.422
One hand stationary dribble	*-.043*	-.199	.811	.301
Two-hand catch	*-.229*	*-.095*	.753	.372
Kick a stationary ball	.385	*-.124*	.525	.561
Overhand throw	.542	.240	.448	.448
Underhand throw	*.231*	.209	.570	.578
Run	*.114*	.627	.335	.482
Gallop	*.037*	.482	.320	.664
Hop	.194	.756	*.097*	.382
Skip	-.207	.260	.432	.703
Horizontal jump	*.004*	.435	.233	.757
Slide	*.037*	.285	.535	.631
ω	.620	.691	.849	

Notes. λ = Factor loadings (target loadings are in greyscale); δ = Uniquenesses; ω = McDonald's omega coefficient of composite reliability; G-factor = global factor; TGMD-3 = Test of Gross Motor Development - Third edition. Non-significant parameters ($p > .05$) are in italics and underlined.

Online Supplements for:

Psychometric Properties of a French-Canadian Version of the Test of Gross Motor Development – Third Edition (TGMD-3): A Bifactor Structural Equation Modeling Approach

Table S1

Standardized Parameters Estimates from the 2-Factor Confirmatory Factor Analytic Representation of the TGMD-3

Skills	Ball Skills (λ)	Locomotor Skills (λ)	δ
Two-hand strike of a stationary ball	.719		.483
One-hand forehand strike of self-bounced ball	.670		.551
One hand stationary dribble	.649		.578
Two-hand catch	.572		.672
Kick a stationary ball	.578		.666
Overhand throw	.676		.543
Underhand throw	.668		.553
Run		.681	.536
Gallop		.588	.654
Hop		.547	.701
Skip		.453	.795
Horizontal jump		.447	.800
Slide		.667	.555
ω	.835	.739	
Latent Factor Correlation	.616		

Notes. λ = Factor loadings (target loadings are in greyscale); δ = Uniquenesses; ω = McDonald's omega coefficient of composite reliability; TGMD-3 = Test of Gross Motor Development - Third edition. All parameters are statistically significant ($p \leq .01$).

Table S2

Standardized Parameters Estimates from the Exploratory Structural Equation Modeling Representation of the TGMD-3

Skills	Ball Skills (λ)	Locomotor Skills (λ)	δ
Two-hand strike of a stationary ball	.597	.195	.515
One-hand forehand strike of self-bounced ball	.657	*.054*	.538
One hand stationary dribble	.844	-.220	.385
Two-hand catch	.692	-*.139*	.577
Kick a stationary ball	.645	*-.075*	.617
Overhand throw	.477	.312	.559
Underhand throw	.508	.253	.577
Run	*.065*	.689	.486
Gallop	*.089*	.525	.680
Hop	-.195	.834	.393
Skip	.211	.253	.850
Horizontal jump	*.013*	.478	.766
Slide	.398	.299	.659
ω	.838	.712	
Latent Factor Correlation	.391		

Notes. λ = Factor loadings (target loadings are in greyscale); δ = Uniquenesses; ω = McDonald's omega coefficient of composite reliability; TGMD-3 = Test of Gross Motor Development - Third edition. Non-significant parameters ($p > .05$) are in italics and underlined.

Table S3

Standardized Parameters Estimates from the Bifactor Confirmatory Factor Analytic Representation of the TGMD-3

Skills	Ball Skills (λ)	Locomotor Skills (λ)	G-factor	δ
Two-hand strike of a stationary ball	.395		.575	.514
One-hand forehand strike of self-bounced ball	.588		.428	.471
One hand stationary dribble	.551		.427	.514
Two-hand catch	.421		.407	.656
Kick a stationary ball	.583		.320	.557
Overhand throw	*.287*		.610	.545
Underhand throw	*.273*		.609	.554
Run		.511	.522	.467
Gallop		.469	.421	.603
Hop		.551	.372	.557
Skip		*.121*	.417	.811
Horizontal jump		.423	.306	.727
Slide		*.023*	.693	.519
ω	.716	.544	.833	

Notes. λ = Factor loadings (target loadings are in greyscale); δ = Uniquenesses; ω = McDonald's omega coefficient of composite reliability; G-factor = global factor; TGMD-3 = Test of Gross Motor Development - Third edition. Non-significant parameters (p > .05) are in italics and underlined.