

Psychometric Properties of the Test of Gross Motor Development, Third Edition (German Translation): Results of a Pilot Study

Matthias O. Wagner
University of Konstanz

E. Kipling Webster
Louisiana State University

Dale A. Ulrich
University of Michigan

The *Test of Gross Motor Development, 3rd Edition* (TGMD-3) is a process-oriented fundamental movement skill assessment to examine the movement patterns displayed by children between the ages of 3 and 10 years. Within this paper, results of a pilot study on the reliability, validity, and measurement invariance across gender of the TGMD-3 (German translation) are presented. In total, performances of 189 typically developing children (99 boys, 90 girls, 56 kindergarten children, 133 elementary school children, $M_{\text{age}} = 7.15 \pm 2.02$ years) are analyzed. Results provide preliminary evidence for test-retest, interrater and intrarater reliability, internal consistency, age- and gender- specific performance trends, factorial validity, measurement invariance across gender, divergent validity, and ball skill-related concurrent and predictive validity of the TGMD-3 (German translation). Subsequent research should be focused on a verification of the present findings on a representative database to foster the application of the TGMD-3 (German translation) in different settings.

Keywords: fundamental movement skills, process-oriented assessment, motor competencies

Fundamental motor skills are crucial for the development of perceived (motor) competence, health-related fitness, and physical activity (Robinson et al., 2015), and they become accessible via fundamental movement skills at the observational level (Burton & Miller, 1998). From a dimensional point of view, fundamental movement skills are commonly divided into two subscales, locomotor (e.g., running, jumping)

Wagner is with the Dept. of Sports Science, University of Konstanz, Konstanz, Germany. Webster is with the School of Kinesiology, Louisiana State University, Baton Rouge, LA. Ulrich is with the School of Kinesiology, University of Michigan, Ann Arbor, MI. Please address author correspondence to Matthias O. Wagner at matthias.wagner@uni-konstanz.de

and object control skills (e.g., catching, throwing; Haywood & Getchell, 2014). Fundamental movement skills generally emerge between 1 and 7 years of age, and the performance of such skills in an upright or bipedal position is based on the acquisition of early motor/movement milestones (rolling over, crawling, creeping, sitting, standing, walking, object manipulation) during the first year of life (Burton & Miller, 1998). Early movement milestones and fundamental movement skills are also referred to as phylogenetic skills because of their universal occurrence in all cultures of the world. As such, they differ from ontogenetic, specialized movement skills (e.g., pitching a baseball or shooting a free throw) that “are not performed by persons in all cultures, nor by all persons in any single culture” (Burton & Miller, 1998, p. 59). These movement skills are based on various foundational areas (e.g., balance/postural control) that have to be separated from skill-underlying motor abilities (e.g., cardiovascular/muscular endurance, flexibility) from an assessment standpoint, and they become functional skills when transferred to their natural and meaningful context (e.g., performing a baseball pitch or a free throw in a regular season game; Burton & Miller, 1998). The development of fundamental movement skills is characterized by a gradual increase across a variety of skills in childhood (Clark, 2005) as well as an early differentiation in object control skills in favor of boys (Goodway, Robinson, & Crowe, 2013).

The Test of Gross Motor Development (TGMD; Ulrich, 1985; TGMD-2; Ulrich, 2000) represents a well-established tool for the assessment of fundamental movement skills in 3- to 10-year-old boys and girls. A comprehensive overview of the psychometric properties of the TGMD-2 can be found in Cools, De Martelaer, Samaey, and Andries (2008). The third edition of the TGMD (TGMD-3; Ulrich, 2013) contains six locomotor and seven ball skill (i.e., object control) tasks that can be summarized in two respective subscale scores. For each of the 13 TGMD-3 movement tasks, three to five criteria are formulated to enable a differentiated assessment of the movement quality, such as the flight phase in running or the contralateral step in the overhand throw.

While Germany has a long tradition in motor ability-based physical fitness testing, motor skill assessments are comparatively underrepresented. The German version of the Movement Assessment Battery for Children 2 (M-ABC 2; Petermann, 2011) is presumably the most prominent approach. However, the M-ABC 2 subscales of manual dexterity, balance, and aiming and catching obviously deviate from the abovementioned structure of fundamental movement skills. The only existing procedure for the diagnosis of locomotor and ball skill performances published to date in Germany is the *Motorische Basiskompetenzen* (MOBAK, basic motor competencies) test battery (Herrmann, Gerlach, & Seelig, 2015). The MOBAK was developed for use in primary schools, contains four locomotor (balancing, rolls, jumping, and side stepping) and four ball skill (dribbling, bouncing, catching, and throwing) tasks, and represents a product-oriented assessment because the completion of each task is evaluated by binary “can do” statements (e.g., pupil is able to or is not able to run or catch). In extension to the MOBAK test battery and as indicated above, the TGMD-3, like its two previous versions, offers the possibility of a process-oriented assessment within these respective subdomains. Based on this qualitative information, targeted movement programs in early childhood (Logan, Robinson, Wilson, & Lucas, 2011), school age (Morgan et al., 2013) as well as in the occupational therapy context (Slater, Hillier, & Civetta, 2010) can be established,

evaluated and optimized, which emphasizes the necessity of a German translation of the TGMD-3 from an intervention standpoint. Moreover, since the TGMD-3 already is a commonly accepted assessment of fundamental movement skills in many different countries, a translation would enable German-speaking researchers to contribute to the international discussion on the promotion of fundamental movement skills as a primary pedagogical focus (see Barnett et al., 2016).

The aim of this paper is to initially test the psychometric properties of the TGMD-3 (German translation) based on pilot-study data. Regarding reliability, it is hypothesized that the TGMD-3 (German translation) will provide similar results when repeated in the same sample within a two-week interval (test-retest reliability; hypothesis 1), rated by two independent raters at the same time (interrater-reliability; hypothesis 2), or rated twice by the same rater within a four-week interval (intrarater-reliability; hypothesis 3). In addition, it is assumed that the TGMD-3 (German translation) items reflecting the same construct will yield a strong coherence (internal consistency; hypothesis 4). Regarding validity, it is assumed that both subscale scores of the TGMD-3 (German translation) are positively correlated with age (hypothesis 5) and that boys outperform girls in ball skills (hypothesis 6). Further, it is assumed that the postulated two factor structure is recoverable in the given set of test scores (factorial validity; hypothesis 7) and that these factors will prove to be equivalent across gender (measurement invariance; hypothesis 8). Finally, it is assumed that the TGMD-3 (German translation) will show only a slight overlap with the German version of the M-ABC 2 on the subscale level (divergent validity; hypothesis 9) and that TGMD-3 (German translation) subscale performances are substantially correlated with selected German Youth Games performances at the time of (concurrent validity; hypothesis 10) as well twelve months after (predictive validity; hypothesis 11) their respective assessment.

Methods

Participants

A total of 189 typically developing children (99 boys, 90 girls, 56 kindergarten children, 133 elementary school children) with a mean age of 7.15 years ($SD = \pm 2.02$ years; range: 3.17–10.67 years) and a mean body-mass index (BMI) of 16.03 kg/m^2 ($SD = \pm 1.92 \text{ kg/m}^2$; range: 10.45–27.68 kg/m^2 ; $SDS_{LMS \text{ boys}} = + 0.4$, $SDS_{LMS \text{ girls}} = + 0.4$; $SDS =$ standardized deviation score [Kromeyer-Hauschild, Dortschy, Stolzenberg, Neuhauser, & Rosario, 2011]) participated in this study. The recruitment and testing of these children took place in five kindergartens and two elementary schools in southern Germany. This sample was used for the evaluation of internal consistency (hypothesis 4), age-specific (hypothesis 5) and gender-specific (hypothesis 6) performance trends, factorial validity (hypothesis 7), and measurement invariance across gender (hypothesis 8) of the TGMD-3 (German translation).

The total sample of 189 typically developing children included the following subsamples:

One hundred and four children (52 boys, 52 girls, 50 kindergarten children, 54 elementary school children) with a mean age of 6.72 years ($SD = \pm 2.27$ years, range = 3.17–10.67 years) and a mean BMI of 16.12 kg/m^2 ($SD = \pm 1.49 \text{ kg/m}^2$,

range = 13.63–22.23 kg/m², $SDS_{LMS\ boys} = + 0.5$, $SDS_{LMS\ girls} = + 0.4$) completed the TGMD-3 (German translation) twice within a two-week interval; this sample was used for the analysis of test-retest reliability (hypothesis 1).

Randomly selected TGMD-3 (German translation) performances of 30 children (14 boys, 16 girls, 10 kindergarten children, 20 elementary school children) with a mean age of 6.96 years ($SD = \pm 2.16$ years, range = 3.50–10.33 years) and a mean BMI of 15.91 kg/m² ($SD = \pm 1.52$ kg/m², range = 13.82–20.72 kg/m², $SDS_{LMS\ boys} = + 0.4$, $SDS_{LMS\ girls} = + 0.2$) were rated by two independent raters at the same time (interrater-reliability; hypothesis 2) and rated twice by the same rater after a four-week interval (intrarater-reliability; hypothesis 3), respectively.

One hundred forty seven children (77 boys, 70 girls, 54 kindergarten children, 93 elementary school children) with a mean age of 7.19 years ($SD = \pm 2.25$ years, range = 3.17–10.67 years) and a mean BMI of 16.10 kg/m² ($SD = \pm 1.86$ kg/m², range = 12.46–27.68 kg/m², $SDS_{LMS\ boys} = + 0.5$, $SDS_{LMS\ girls} = + 0.4$) also completed the German version of the M-ABC 2 at the time of their respective TGMD-3 (German translation) assessment; this sample was used for the analysis of divergent validity (hypothesis 9).

Out of the 133 elementary school children, 91 children (49 boys, 42 girls) with a mean age of 8.63 years ($SD = \pm 1.19$ years, range = 6.33–10.67 years) and a mean BMI of 16.19 kg/m² ($SD = \pm 2.05$ kg/m², range = 12.46–27.68 kg/m², $SDS_{LMS\ boys} = + 0.3$, $SDS_{LMS\ girls} = \pm 0$) also completed the German Youth Games at the time of and 68 children (37 boys, 31 girls) with a mean baseline age of 8.21 years ($SD = \pm 0.98$ years, range = 6.33–9.75 years) and a mean baseline BMI of 16.09 kg/m² ($SD = \pm 1.71$ kg/m², range = 12.46–21.52 kg/m², $SDS_{LMS\ boys} = \pm 0$, $SDS_{LMS\ girls} = \pm 0$) completed the German Youth Games twelve months after their TGMD-3 (German translation) assessment; respective subsamples were used for the analysis of concurrent (hypothesis 10) or predictive (hypothesis 11) validity, respectively.

Measures

The TGMD-3 (German translation), which results from the translation of the American original version (i.e., in English), was validated in terms of coherence in an independent back translation and contains six locomotor and six ball skill tasks (see Table 1).

The item “Stationary strike from a tee” was excluded from the German translation of the TGMD-3 because the authors did not see a comprehensive phylogenetic relevance within this particular task. Following the original publication and the herein described developmental character of the TGMD-3, all tasks and corresponding criteria remained unchanged across all age groups. The TGMD-3 (German translation) is conducted as a single test by specifically trained test instructors, takes between 15–20 min, contains a demonstration, a practice trial and two scoring trials per task, and only requires little material effort (e.g., various balls and cones).

On the basis of the binary encoding exemplified in Table 2, children could score a maximum of 46 locomotor and 44 balls skill points in two scoring trials (see Table 1).

Table 1 Subscales, Skills, and Scores Within the TGMD-3 (German Translation)

Locomotor skills	MS	Ball skills	MS
Run	8	One-hand forehand strike	8
Gallop	8	One-hand stationary dribble	6
One-legged hop	8	Two-hand catch	6
Skip	6	Kick a stationary ball	8
Horizontal jump	8	Overhand throw	8
Slide	8	Underhand throw	8
Total	46	Total	44

Note. MS = maximum score.

Table 2 Evaluation of Movement Quality Within the TGMD-3 (German Translation) Using the Example of Kicking a Stationary Ball

Performance criteria	Trial 1	Trial 2	Score
Rapid, continuous approach to the ball	0/1	0/1	0–2
Child takes an elongated stride or leap just prior to ball contact	0/1	0/1	0–2
Nonkicking foot placed close to the ball	0/1	0/1	0–2
Kicks ball with instep or inside of preferred foot (not the toes)	0/1	0/1	0–2
Skill Score			0–8

In order to establish divergent validity of the TGMD-3 (German translation; hypothesis 9), we administered the German version of the M-ABC 2 (age band 1 and 2; Petermann, 2011), which covers the motor subscales of manual dexterity, balance, and aiming and catching in terms of a product-oriented assessment. Thus, divergence between both measures was assumed not only from a content-related perspective but also from a methodological one. Further, to establish concurrent (hypothesis 10) and predictive (hypothesis 11) validity of the TGMD-3 (German translation), we recorded elementary school children's 50 m sprinting time and their ball-throwing distance performance within the German Youth Games in terms of specialized movement skills.

Procedure

All institutions were recruited out of a regional database containing information on child and youth studies on the basis of their voluntary availability. Children's testing took place between May 2014 and May 2016 and was approved in written form by the ethical commission of the involved universities, by the management

personnel of the kindergartens and schools, and by the children's parents. A central part of the written consent was the parents' agreement with the videotaping of their children when they were performing the tasks of the TGMD-3 (German translation). Two independent and specifically trained assessors (assessor 2 was consulted in terms of interrater reliability; hypothesis 2) viewed the resulting video sequences at the normal speed and analyzed the performances per the TGMD-3 (German translation) criteria (see Table 2).

Statistical Analysis

Subsequent results on the reliability, validity, and measurement invariance across gender of the TGMD-3 (German translation) were calculated using SPSS 23.0 and AMOS 23.0 (Arbuckle, 2014). The significance level for all statistical tests was set a priori to $\alpha = .05$.

Intraclass correlation coefficient (ICC) was used to evaluate test-retest (hypothesis 1), interrater (hypothesis 2), and intrarater (hypothesis 3) reliability of the TGMD-3 (German translation). For the evaluation of internal consistency of both TGMD-3 (German translation) subscales (hypothesis 4), Cronbach's α was calculated.

Simple linear regressions as well as a one-way analysis of variance were applied for the identification of age-specific (hypothesis 5) and gender-specific (hypothesis 6) performance trends within both TGMD-3 (German translation) subscales. To verify factorial validity of the TGMD-3 (German translation; hypothesis 7), a confirmatory factor analysis (CFA) was conducted. The Mardia-Test was used for the assessment of multivariate normal distribution. Latent variables were scaled by use of the fixed factor method, which was found to be superior to the common marker method (e.g., Little, 2013). The nonnormalized model parameters were estimated by maximum likelihood (ML) using the covariance matrix. ML was chosen with reference to Hancock and Mueller (2006) who state that "if the observed data have many categories (e.g., at least five ordered categories) and are approximately normal, use of ML estimation techniques does not result in severe levels of bias in fit indices, parameter estimates or standard errors." To evaluate the global model fit, selected incremental fit indices were referenced, including root-mean-square error of approximation (RMSEA), standardized root mean residual (SRMR), and comparative fit index (CFI). The selection of indices is oriented on Beauducel and Wittmann (2005), and their evaluation is based on the recommendations by Hu and Bentler (1999). However, despite a good global fit, structural equation models can still be problematic within their substructures. Therefore, the model's local fit was assessed more comprehensively in a second step by evaluating particular inner-method convergent and divergent measures. The inner-method convergent measures assessed were indicator reliability (IR), factor reliability (FR), and the average variance extracted (AVE). Fornell-Larcker ratios (FLR) and χ^2 -difference ($\Delta\chi^2$) tests were used as inner-method divergent measures. The selection and evaluation of inner-method convergent and divergent measures was based on the recommendations by Homburg, Klarmann, and Pflesser (2008).

Measurement invariance across gender of the TGMD-3 (German translation; hypothesis 8) was examined by testing and comparing four (nested) models using multiple-group analysis, with each successive model including the previous model

restrictions as follows (see Milfont & Fischer, 2010): model A tested equivalence of the structure across gender (configural invariance); model B tested equivalence of the factor loadings across gender (metric invariance); model C tested equivalence of the item intercepts across gender (scalar invariance), and model D tested equivalence of the error variance across gender (error variance invariance). The successive, nested models were compared by $\Delta\chi^2$ tests. Due to the sensitivity of the $\Delta\chi^2$ test to sample size, the difference of CFI (ΔCFI) was considered in addition with a value of $\Delta CFI \leq .01$, indicating that the null hypothesis of invariance should not be rejected (see Cheung & Rensvold, 2002).

For the evaluation of outer-method divergent validity (Hypothesis 9) and following the results by Logan, Robinson, and Getchell (2011) as well as Logan, Robinson, Rudisill, Wadsworth and Morera (2012), we calculated Spearman's rho (one-tailed) between the German version of the M-ABC 2 and the TGMD-3 (German translation) subscale scores expecting low to moderate correlations.

Finally, to evaluate concurrent (Hypothesis 10) and predictive (Hypothesis 11) validity of both TGMD-3 (German translation) subscales, children's 50 m sprinting time and their ball-throwing distance performance at the time of as well as twelve months after the TGMD-3 (German translation) assessment were transformed into age- and gender-specific scores on the basis of the existing German Youth Games normative data. Spearman's rho (one-tailed) was calculated assuming positive relations between the TGMD-3 (German translation) locomotor score and the 50 m sprinting time score or the TGMD-3 (German translation) ball skill score and the ball-throwing distance performance score at both measurement points, respectively.

Results

The age- and gender- specific distribution of participants as well as their respective TGMD-3 (German translation) subscale scores can be found in Table 3.

Reliability

The TGMD-3 (German translation) provides excellent test-retest (hypothesis 1, interval: 2 weeks, $N = 104$; locomotor skills: ICC = .94, 95% CI [.91, .96], $p < .001$; balls skills: ICC = .98, 95% CI [.97, .99], $p < .001$), interrater (hypothesis 2, 2 raters, $N = 30$ observations; locomotor skills: ICC = .88, 95% CI [.76, .95], $p < .001$; ball skills: ICC = .97, 95% CI [.94, .99], $p < .001$) as well as intrarater (hypothesis 3, $N = 30$ observations, interval: 4 weeks; locomotor skills: ICC = .97, 95% CI [.94, .99], $p < .001$; ball skills: ICC = .99, 95% CI [.98, 1.00], $p < .001$) reliability. Internal consistency (hypothesis 4) of both TGMD-3 (German translation) subscales ($N = 189$) ranges between acceptable (locomotor skills: Cronbach's $\alpha = .76$) and good (ball skills: Cronbach's $\alpha = .89$). In summary, results provide strong evidence for the reliability of the TGMD-3 (German translation).

Age- and Gender-Specific Performance Trends

Results indicate an expected performance increase with age in both TGMD-3 (German translation) subscales (hypothesis 5, locomotor skills: $\beta = .64$, $t(186) = 11.41$, $p < .001$, $R^2 = .41$, $F(1, 186) = 130.24$, $p < .001$; ball skills: $\beta = .79$, $t(186)$

Table 3 Age- and Gender-Specific Distribution of Participants and Corresponding TGMD-3 (German Translation) Subscale Scores

	Boys (<i>N</i> = 99)		Girls (<i>N</i> = 90)		Total sample (<i>N</i> = 189)	
	<i>N</i>	<i>M</i> (<i>SD</i>)	<i>N</i>	<i>M</i> (<i>SD</i>)	<i>N</i>	<i>M</i> (<i>SD</i>)
Locomotor skills						
3 years	9	17.11 (5.06)	8	22.13 (5.84)	17	19.47 (5.86)
4 years	11	21.82 (9.24)	8	19.00 (10.39)	19	20.63 (9.56)
5 years	8	27.50 (5.90)	9	30.44 (4.00)	17	29.06 (5.06)
6 years	11	29.27 (9.32)	12	32.25 (6.68)	23	30.83 (8.00)
7 years	26	34.65 (7.07)	22	34.09 (5.26)	48	34.40 (6.24)
8 years	12	36.83 (6.19)	8	34.00 (8.52)	20	35.70 (7.14)
9 years	17	35.06 (4.92)	12	37.67 (3.70)	29	36.14 (4.57)
10 years	5	36.60 (2.88)	11	35.82 (5.00)	16	36.06 (4.36)
Ball skills						
3 years	9	8.33 (3.91)	8	6.13 (4.05)	17	7.29 (4.01)
4 years	11	12.18 (5.34)	8	9.25 (5.12)	19	10.95 (5.32)
5 years	8	18.13 (3.44)	9	12.33 (5.17)	17	15.06 (5.24)
6 years	11	24.36 (6.85)	12	21.75 (8.59)	23	23.00 (7.75)
7 years	26	32.27 (7.82)	22	25.45 (8.40)	48	29.15 (8.71)
8 years	12	38.58 (3.37)	8	27.38 (5.04)	20	34.10 (6.91)
9 years	17	35.47 (5.68)	12	34.50 (4.10)	29	35.07 (5.02)
10 years	5	36.00 (2.00)	11	31.55 (6.27)	16	32.94 (5.64)

Note. *N* = number of participants; *M* = mean; *SD* = standard deviation.

= 17.86, $p < .001$, $R^2 = .63$, $F(1, 186) = 318.88$, $p < .001$) as well as at least small advantages when comparing the ball skill performance of the boys (hypothesis 6, $N = 99$, $M = 27.34$, $SD = \pm 11.67$) to the ball skill performance of the girls ($N = 90$, $M = 22.61$, $SD = \pm 11.15$, $F(1,147) = 8.09$, $df = 1$, $p < .01$, $\eta_p^2 = .041$). Both results reflect the assumptions derived from the literature and can be considered an initial indication for the construct-related validity of the TGMD-3 (German translation).

Factorial Validity

Applying maximum likelihood requires that the included indicators be interval-scaled and multivariate normally distributed. Interval scale level was assumed for all indicators with reference to Hancock and Mueller (2006); their intercorrelations can be found in Table 4.

Even though skewness and kurtosis of all subtests fell within the limit recommended by West, Finch, and Curran (1995), the Mardia-Test showed a slight

Table 4 Subtest Intercorrelation Matrix of the TGMD-3 (German Translation; N = 189)

Subtests	1	2	3	4	5	6	7	8	9	10	11	12
Run (1)	1											
Gallop (2)	.26	1										
One-legged hop (3)	.32	.33	1									
Skip (4)	.29	.26	.44	1								
Horizontal jump (5)	.45	.25	.40	.36	1							
Slide (6)	.38	.25	.52	.48	.37	1						
One-hand forehand strike (7)	.40	.28	.46	.47	.46	.47	1					
One-hand stationary dribble (8)	.38	.28	.45	.45	.42	.48	.63	1				
Two-hand catch (9)	.41	.21	.46	.41	.27	.45	.59	.54	1			
Kick a stationary ball (10)	.39	.20	.39	.39	.40	.36	.66	.66	.52	1		
Overhand throw (11)	.38	.25	.40	.43	.39	.37	.66	.61	.45	.66	1	
Underhand throw (12)	.34	.28	.39	.43	.31	.36	.59	.56	.45	.53	.65	1

violation of the multivariate normal distribution assumption (critical ratio > 2.58; see Table 5). Therefore, the Bollen–Stine bootstrap procedure (200 samples) was applied to achieve necessary correction of p values.

The resulting and proposed two-factor solution showed an appropriate global model-fit ($\chi^2[53, N = 189] = 71.31$, Bollen–Stine $p = .164$, $\chi^2/df = 1.35$, CFI = .98, RMSEA = .04, 90% CI [.01, .07], SRMR = .04), which basically supports the assumption of factorial validity of the TGMD-3 (German translation; hypothesis 7).

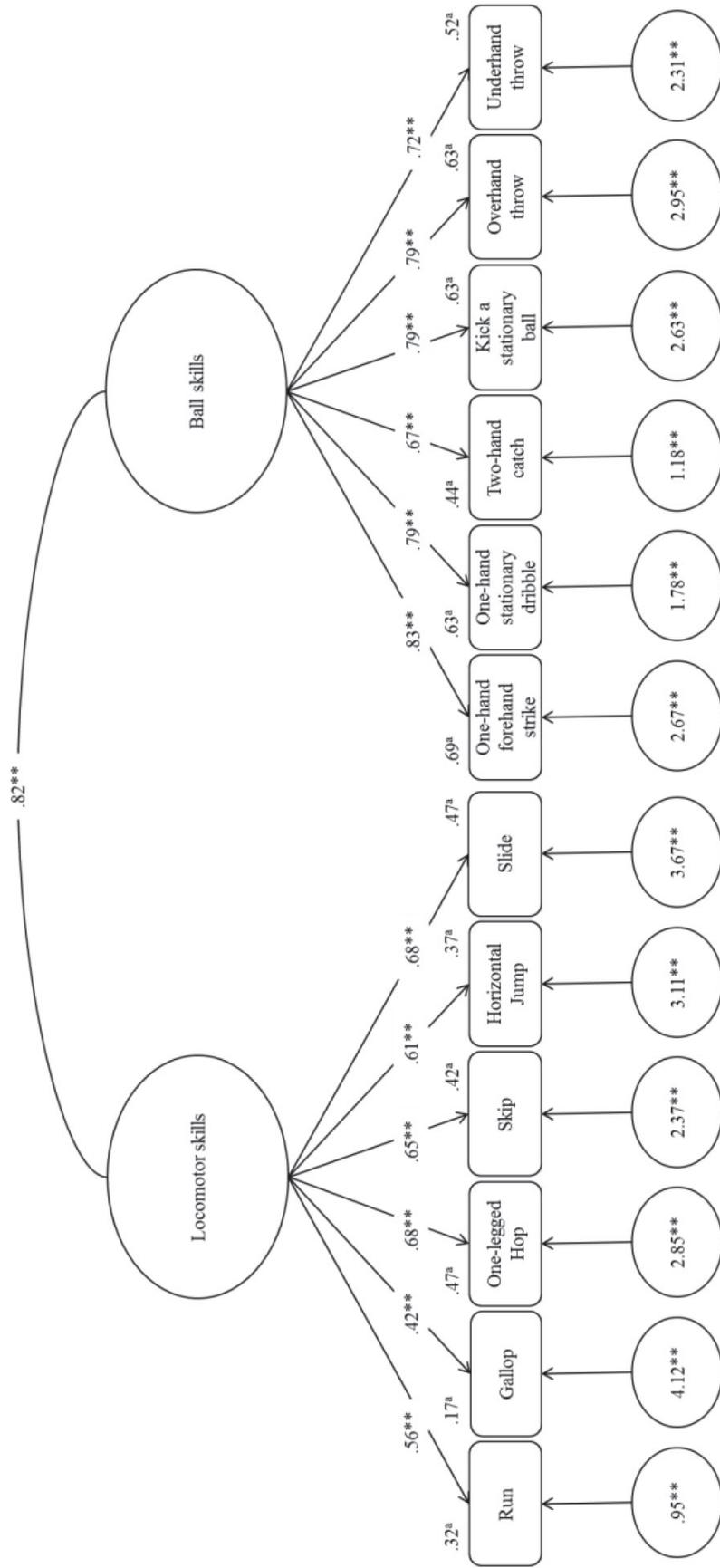
For inner-method convergent measures, each estimated factor loading differed significantly from zero ($p < .001$; see Figure 1) and FR of both latent variables exceeded the required minimum of .60 (see Table 6). In addition, the latent variable ball skills explained more than 50% of the variance (AVE) of its subordinated indicators whereupon the latent variable locomotor skills did not show a comparable and sufficient explanatory power (see Table 6). Besides running and horizontal jump, indicator gallop proved to be particularly critical here because considerably less than 40% of its respective variance (IR) was assessed by the superordinate latent variable (see Figure 1).

Concerning inner-method divergent measures (see Table 6), the AVE of both latent variables was lower than the maximum squared intercorrelation (MSI) resulting in FLRs greater than 1. This result can be interpreted as an initial indication for a lack of inner-method discriminance. However, fixing the intercorrelation of both factors to one resulted in a significantly inferior fit of the model ($\chi^2(54, N = 189) = 111.31$, Bollen–Stine $p < .01$, $\chi^2/df = 2.06$, CFI = .94, RMSEA = .08, 90% CI

Table 5 Subtests Descriptives Within the TGMD-3 (German Translation; $N = 189$)

Subtests	<i>M</i>	<i>SD</i>	Min	Max	Skewness	<i>c. r.</i>	Kurtosis	<i>c. r.</i>
Run	7.33	1.18	2	8	-1.75	-9.84	2.81	7.88
Gallop	3.93	2.24	0	8	0.46	2.58	-1.22	-3.43
Hop	4.60	2.32	0	8	-0.33	-1.87	-0.85	-2.38
Skip	3.53	2.02	0	6	-0.67	-3.77	-0.70	-1.95
Horizontal jump	5.52	2.22	0	8	-0.63	-3.53	-0.61	-1.71
Slide	6.50	2.63	0	8	-1.63	-9.13	1.18	3.32
One-hand forehand strike	4.20	2.94	0	8	-0.26	-1.46	-1.39	-3.99
One-hand stationary dribble	2.57	2.19	0	6	0.19	1.09	-1.38	-3.86
Two-hand catch	4.70	1.46	0	6	-1.12	-6.28	0.84	2.34
Kick a stationary ball	4.51	2.70	0	8	0.02	0.12	-1.48	-4.15
Overhand throw	3.70	2.82	0	8	-0.06	-0.32	-1.34	-3.75
Underhand throw	5.47	2.20	0	8	-0.67	-3.76	-0.29	-0.82
Multivariate	—	—	—	—	—	—	10.53	3.95

Note. *M* = mean scale score; *SD* = standard deviation; Min = minimum; Max = maximum; *c. r.* = critical ratio.



$\chi^2(53, N = 189) = 71.31$, Bollen-Stine $p = .164$, $\chi^2/df = 1.35$, $CFI = .98$, $RMSEA = .04$, 90% CI [.01, .07], $SRMR = .04$

Figure 1 — Factorial validity of the TGMD-3 (German translation). Results of the confirmatory factor analysis (standardized solution); a = IR.

Table 6 Inner-method convergent and divergent measures within the TGMD-3 (German translation; N = 189)

Subscales	FR	AVE	MSI		FLR
			Locomotion	Ball skills	
Locomotor skills	.77	.38	—	.67	1.77
Ball skills	.90	.62	—	—	1.09

Note. FR = factor reliability; AVE = average variance extracted; MSI = maximum squared intercorrelation; FLR = Fornell-Larcker ratio.

[.06, .10], SRMR = .05); $\Delta\chi^2 = 40.01, p < .001$), which counteracts the existence of a superior one-factor solution. Thus, subsequent analysis of measurement invariance across gender of the TGMD-3 (German translation; hypothesis 8) was conducted with the proposed two-factor model. Corresponding results can be found in Table 7.

Measurement Invariance Across Gender

Results on the measurement invariance across gender of the TGMD-3 (German translation; hypothesis 8, see Table 7) support the assumption of configural (model A) and metric (model B) invariance across gender, which means that the postulated two-factor structure is equivalent in both groups and that boys and girls show a comparable item responsiveness. However, the $\Delta\chi^2$ test was significant ($p < .001$) when comparing models B and C with a substantial decrease in CFI (Δ CFI = .03). The derived rejection of the scalar invariance null-hypothesis counteracts the objective of comparing the latent means between both groups at least when considering the presented pilot data.

Divergent Validity

When comparing the TGMD-3 (German translation) subscales of locomotor skills and ball skills to the M-ABC 2 (German version) subscales of manual dexterity, balance, and aiming and catching, only low to moderate correlations can be identified (see Table 8), which supports the assumption of outer-method divergence between both measures (hypothesis 9).

Concurrent and Predictive Validity

Indications for the concurrent and predictive validity of the TGMD-3 (German translation) can be derived from the relationship between the TGMD-3 (German translation) ball skill score and the German Youth Games ball-throwing distance performance at the time of (hypothesis 10, $r_s(89) = .36, p < .001$) as well as twelve months after (hypothesis 11, $r_s(66) = .39, p < .001$) the TGMD-3 (German translation) assessment. The identified correlations are in line with our hypothesis: the higher a child's TGMD-3 (German translation) ball skill score, the higher his or her ball-throwing distance performance score. However, no significant relationship can be identified between the TGMD-3 (German translation) locomotor skill

Table 7 Analysis of Measurement Invariance Across Gender Within the TGMD-3 (German Translation; 99 boys, 90 girls)

Models	χ^2	df	p	CFI	RMSEA	$\Delta\chi^2$	pa
A	132.61	106	.041	.97	.04	—	—
B	142.75	116	.046	.97	.04	10.14	.428
C	184.26	128	.001	.94	.05	41.51	.000
D	190.10	140	.003	.95	.04	5.83	.924

Note. χ^2 = chi square; df = degrees of freedom; p = probability value; CFI = comparative fit index; RMSEA = root mean square error of approximation; $\Delta\chi^2$ = chi square difference; model A = configural invariance; model B = model A + metric invariance; model C = model B + scalar invariance; model D = model C + error variance invariance.

^ap value to compare nested models.

Table 8 Bivariate Correlations (Spearman's Rho) Between the TGMD-3 (German Translation) and the German M-ABC 2 Subscales (N = 147)

Subscales	M-ABC 2 manual dexterity	M-ABC 2 balance	M-ABC 2 aiming and catching
TGMD 3 locomotion	.29**	.33**	.22*
TGMD 3 ball skills	.23*	.25*	.30**

*p (one-tailed) < .01; **p (one-tailed) < .001.

score and the 50 m sprinting performance at the time of (hypothesis 10, $r_s(89) = .15$, $p = .086$, $1-\beta = .42$) or twelve months after the TGMD-3 (German translation) assessment (hypothesis 11, $r_s(66) = .08$, $p = .253$, $1-\beta = .16$).

Discussion

The purpose of this study was to initially test reliability, validity, and measurement invariance across gender of the TGMD-3 (German translation) using a pilot-study sample of kindergarten and elementary school children. Results provide preliminary evidence for test-retest (hypothesis 1), interrater (hypothesis 2) and intrarater (hypothesis 3) reliability, internal consistency (hypothesis 4), age-specific (hypothesis 5) and gender-specific (hypothesis 6) performance trends, and factorial validity (hypothesis 7) of the TGMD-3 (German translation). Hence, results are in line with those of recently published studies on the American version (Webster & Ulrich, 2017) as well as the Brazilian translation of the TGMD-3 (Valentini, Zanella, & Webster, 2017). However, because the item “Stationary strike from a tee” was dropped from the German translation of the TGMD-3, cross-cultural comparisons are of only limited significance. Concerning factorial validity, a detailed inner-method convergence analysis indicated punctual improvement potential within the model's substructure. Subtests gallop, run, and horizontal jump proved

to be particularly problematic and should therefore be revised with respect to both the test instructions as well as the performance criteria used. For example, it should be clarified that galloping from one cone to the other does not have to be performed as fast but rather as accurate as possible. Moreover, combined criteria such as “Arms move in opposition to legs with elbows bent,” in the case of running, or “Both feet come off the floor together and land together,” in the case of the horizontal jump, should be avoided. However, revisions should be conducted under reserve since the underlying analysis of factorial validity of the TGMD-3 (German translation) is based only on the observation of 189 primarily elementary school children, which leaves open the possibility of age bias. On the basis of the locally verified two-factor model, nested model comparisons provided at least weak evidence for the measurement invariance of the TGMD-3 (German translation) across gender (hypothesis 8). Corresponding results again should be interpreted with caution since they are based only on the observation of 99 boys and 90 girls. In terms of a further verification of construct validity, results indicate outer-method divergence between the TGMD-3 (German translation) and the German version of the M-ABC 2 subscales (hypothesis 9). Respective results are in line with Logan et al. (2011a) and Logan et al. (2012) and indicate that—as with the comparison of other established motor assessments (e.g., Van Waelvelde, Peersman, Lenoir, & Smits-Engelsman, 2007, Bardid et al., 2016, Fransen et al., 2016)—both tests actually provide different information about motor skill performance and are therefore not interchangeable. In future studies, we propose to verify factorial validity and measurement invariance across age and gender on the basis of an enlarged, representative sample and to examine the outer-method convergent and divergent validity of the TGMD-3 (German translation) based on the multitrait-multimethod approach using the MOBAK test battery (see Herrmann et al., 2015).

Concerning criterion-related validity, our results indicate that the TGMD-3 (German translation) ball skill subscale score is significantly correlated with the German Youth Games ball-throwing distance performance score in a concurrent (hypothesis 10) as well as a predictive (hypothesis 11) sense. However, we did not find similar results when comparing the TGMD-3 (German translation) locomotor subscale score with the German Youth Games 50 m sprinting time score, suggesting that respective performances predominantly depend on the particular foundations of movement (e.g., motivation and neurological functioning) or motor abilities (e.g., speed and strength).

Conclusion

The results presented here on the psychometric properties of the TGMD-3 (German translation) are promising and should encourage other researchers to implement necessary replication studies. However, current results should be interpreted with caution in terms of preliminary evidence since corresponding data were assessed in the course of a pilot study including rather small, nonrepresentative subsamples. Thus, the collection of age- and-gender specific normative data as well as the thereon based further verification of psychometric properties is considered the next logical step before being able to evaluate the usability of the TGMD-3 (German translation) in different settings.

References

- Arbuckle, J. (2014). *IBM SPSS Amos™ 23 user's guide*. Chicago: SPSS.
- Bardid, F., Huyben, F., Deconinck, F.J., De Martelaer, K., Seghers, J., & Lenoir, M. (2016). Convergent and divergent validity between the KTK and MOT 4-6 Motor Tests in early childhood. *Adapted Physical Activity Quarterly*, *33*(1), 33–47. doi:10.1123/APAQ.2014-0228
- Barnett, L.M., Stodden, D., Cohen, K.E., & Smith, J. J., Lubans, D.R., Lenoir, M., ... Morgan, P.J. (2016). Fundamental movement skills: An important focus. *Journal of Teaching in Physical Education*, *35*(3), 219–225.
- Beauducel, A., & Wittmann, W.W. (2005). Simulation study on fit indices in confirmatory factor analysis based on data with slightly distorted simple structure. *Structural Equation Modeling*, *12*, 41–75. doi:10.1207/s15328007sem1201_3
- Burton, A.W., & Miller, D.E. (1998). *Movement skill assessment*. Champaign, IL: Human Kinetics.
- Cheung, G.W., & Rensvold, R.B. (2002). Evaluation goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255. doi:10.1207/S15328007SEM0902_5
- Clark, J.E. (2005). From the beginning: A developmental perspective on movement and mobility. *Quest*, *57*, 37–45. doi:10.1080/00336297.2005.10491841
- Cools, W., De Martelaer, K., Samaey, C., & Andries, C. (2008). Movement skill assessment of typically developing preschool children: A review of seven movement skill assessment tools. *Journal of Sports, Science, and Medicine*, *8*, 154–168.
- Fransen, J., D'Hondt, E., Bourgois, J., Vaeyens, R., Phillipaerts, R., & Lenoir, M. (2016). Motor competence assessment in children: Convergent and discriminant validity between the BOT-2 Short Form and KTK testing batteries. *Research in Developmental Disabilities*, *35*(6), 1375–1383. doi:10.1016/j.ridd.2014.03.011
- Goodway, J.D., Robinson, L.E., & Crowe, H. (2013). Gender differences in fundamental motor skill development in disadvantaged preschoolers from two geographical regions. *Research Quarterly for Exercise and Sport*, *81*(1), 17–24. doi:10.1080/02701367.2010.10599624
- Hancock, G.R., & Mueller, R.O. (2006). *Structural equation modeling*. Greenwich, CT: Information Age Publishing.
- Haywood, K.M., & Getchell, N. (2014). *Lifespan motor development* (6th ed.). Champaign, IL: Human Kinetics.
- Herrmann, C., Gerlach, E., & Seelig, H. (2015). Development and validation of a test instrument for the assessment of basic motor competencies in primary school. *Measurement in Physical Education and Exercise Science*, *19*(2), 80–90. doi:10.1080/1091367X.2014.998821
- Homburg, C., Klarmann, M., & Pflesser, C. (2008). Konfirmatorische Faktorenanalyse [Confirmatory factor analysis]. In A. Herrmann, C. Homburg, & M. Klarmann (Eds.), *Handbuch Marktforschung [Handbook Marketing Research]* (3rd ed., pp. 271–303). Wiesbaden: Gabler.
- Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. doi:10.1080/10705519909540118
- Kromeyer-Hauschild, K., Dortschy, R., Stolzenberg, H., Neuhauser, H., & Rosario, A. S. (2011). Nationally representative waist circumference percentiles in German adolescents aged 11.0–18.0 years. *International Journal of Pediatric Obesity*, *6*(2-2), e129–e137. doi:10.3109/17477166.2010.490267
- Little, T.D. (2013). *Longitudinal structural equation modeling*. New York: The Guilford Press., doi:10.3102/00028312035004705.

- Logan, S.W., Robinson, L.E., & Getchell, N. (2011a). The comparison of performances of preschool children on two motor assessments. *Perceptual and Motor Skills*, 113(3), 715–723. doi:10.2466/03.06.25.PMS.113.6.715-723
- Logan, S.W., Robinson, L.E., Rudisill, M.E., Wadsworth, D., & Morera, M. (2012). The comparison of performances of at-risk school-age children on two motor assessments: The Test of Gross Motor Development and the Movement Assessment Battery for Children. *Physical Education and Sport Pedagogy*, 19(1), 48–59. doi:10.1080/17408989.2012.726979
- Logan, S.W., Robinson, L.E., Wilson, A.E., & Lucas, W.A. (2011b). Getting the fundamentals of movement: A meta-analysis of the effectiveness of motor skill intervention in children. *Child: Care, Health and Development*, 38(3), 305–315.
- Milfont, T.L., & Fischer, R. (2010). Testing measurement invariance across groups. Application in cross-cultural research. *International Journal of Psychological Research*, 3(1), 111–121.
- Morgan, P.J., Barnett, L.M., Cliff, D.P., Okely, A.D., Scott, H.A., Cohen, K.E., & Lubans, D.L. (2013). Fundamental movement skill interventions in youth: A systematic review and meta-analysis. *Pediatrics*, 132(5), e1361–e1383. doi:10.1542/peds.2013-1167
- Petermann, F. (Hrsg.). (2011). *Movement Assessment Battery for Children-2 (M-ABC-2)* (3. überarb. erw. Aufl.). Frankfurt/Main: Pearson Assessment.
- Robinson, L.E., Stodden, D.F., Barnett, L.M., Lopes, V.P., Logan, S.W., Rodrigues, L.P., & D'Hondt, E. (2015). Motor competence and its effect on positive developmental trajectories of health. *Sports Medicine (Auckland, N.Z.)*, 45, 1273–1284. doi:10.1007/s40279-015-0351-6
- Slater, L., Hillier, S., & Civetta, L. (2010). Clinimetric properties of performance based gross motor tests used in children with DCD: Systematic review. *Pediatric Physical Therapy*, 22, 170–179. doi:10.1097/PEP.0b013e3181dbeff0
- Ulrich, D.A. (1985). *Test of Gross Motor Development*. Austin, TX: Pro-Ed.
- Ulrich, D.A. (2000). *Test of Gross Motor Development* (2nd ed.). Austin, TX: Pro-Ed.
- Ulrich, D.A. (2013). The Test of Gross Motor Development–3 (TGMD-3): Administration, scoring, and international norms. *Spor Bilimleri Dergisi*, 24(2), 27–33.
- 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(1), 15–28. https://doi.org/10.1123/jmld.2016-0002
- Van Waelvelde, H., Peersman, W., Lenoir, M., & Smits-Engelsman, B. (2007). Convergent validity between two motor tests: Movement-ABC and PDMS-2. *Adapted Physical Activity Quarterly*, 24, 59–69. doi:10.1123/apaq.24.1.59
- Webster, E.K., & Ulrich, D.A. (2017). Evaluation of the psychometric properties of the Test of Gross Motor Development—3rd edition. *Journal of Motor Learning and Development*, 5(1), 45–58. https://doi.org/10.1123/jmld.2016-0003
- West, S.G., Finch, J.F., & Curran, P.J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R.H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56–75). Thousand Oaks, CA: SAGE Publications, Inc.