Reliability and Validity of the TGMD-2 in Primary-School-Age Children With Visual Impairments

Suzanne Houwen, Esther Hartman, Laura Jonker, and Chris Visscher

University Medical Center Groningen

This study examines the psychometric properties of the Test of Gross Motor Development-2 (TGMD-2) in children with visual impairments (VI). Seventy-five children aged between 6 and 12 years with VI completed the TGMD-2 and the Movement Assessment Battery for Children (Movement ABC). The internal consistency of the TGMD-2 was found to be high (alpha = 0.71–0.72) and the interrater, intrarater, and test-retest reliability acceptable (ICCs ranging from 0.82 to 0.95). The results of the factor analysis supported internal test structure and significant age and sex effects were observed. Finally, the scores on the object control subtest of the TGMD-2 and the ball skills subtest of the Movement ABC correlated moderately to high ($r = 0.45$ to $r = 0.80$). Based on the current results, it is concluded that the TGMD-2 is an appropriate tool to assess the gross motor skills of primary-school-age children with VI.

Visual impairments (VI; best corrected visual acuity ≤ 6/18 in the better eye; World Health Organization, WHO, 2007) may act as a constraint, slowing down motor skill acquisition and/or leading to qualitatively deviant motor skills (Jan, Sykanda, & Groenveld, 1990; Whitall, 2003). Earlier studies indeed showed that children with VI generally have inferior motor skill performance compared with sighted children (Bouchard & Tetreault, 2000; Houwen, Visscher, Hartman, & Lemmink, 2007; Houwen, Visscher, Lemmink, & Hartman, 2008; Reimer, Smits-Engelsman, & Siemonsma-Boom, 1999). As for children with VI, motor skill proficiency is as important for daily living and sports activities as it is for any child (Graf et al., 2004; Houwen, Hartman, & Visscher, 2009; Houwen et al., 2007; Morgan, Okely, Cliff, Jones, & Baur, 2008); it is crucial that health professionals are able to accurately identify the strengths and weaknesses in any child's motor skills to provide appropriate interventions.

As typically developing children grow older, their motor skill performance improves due to maturation, experience, age, and heredity (Branta, Haubenstricker, & Seefeldt, 1984; Davies & Rose, 2000). Moreover, sex differences in gross motor skill performance have been established in both children and adolescents; the performance of boys generally exceeds that of girls, with boys obtaining higher object control skill scores (Okely & Booth, 2004; Woodard & Surburg, 1997). The observed sex-specific differences may result from differences in body composition during growth and maturation and social influences on physical activity (Thomas & French, 1985); however, motor development in children with VI may progress more slowly and differently (Jan et al., 1990; Whitall, 2003), but to date, little is known about the actual progression of gross motor skills in this population.

Motor skill performance in children with VI has been measured by various tests, such as the Bruininks-Oseretsky Test of Motor Proficiency (Bouchard & Tetreault, 2000), the Movement Assessment Battery for Children (Movement ABC; Houwen et al., 2008), and the Test of Gross Motor Development-2 (TGMD-2; Houwen et al., 2007; Houwen et al., 2009; Houwen, Hartman, Visscher, in press). The latter test was specifically designed to assess children's motor skills associated with everyday sports and games played in physical education classes. As the TGMD-2 evaluates performance on locomotor skills and object control skills, it may hence be particularly suitable to test the motor skill performance of primary-school-age children with VI (Shapiro, Lieberman, & Moffett, 2003). But, like any instrument, the reliability and validity of the TGMD-2 first need to be evaluated in this population before it can be recommended as an appropriate means of motor skill assessment.

Reliability refers to the "consistency of scores on a particular instrument" (Gliner, Morgan, & Harmon, 2001) and validity to "the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of the test" (American Educational Research Association, AERA, et al., 1999). Yun and Ulrich (2002) emphasized that the three distinct approaches to providing validity evidence, i.e., (a) content-related evidence, (b) criterion-related evidence, and (c) construct-related evidence, are not three different categories of validity but that these three categories together provide validity evidence.

Reliability and validity evidence for the TGMD-2 has been reported for typically developing children (Evaggelinou, Tsigilis, & Papa, 2002; Ulrich, 2000) and children with intellectual disabilities (Simons et al., 2008), but given that evidence of a measure's reliability and validity cannot be assumed to generalize across populations (AERA, 1999; Yun & Ulrich, 2002), it remains to be established whether the reported estimates also hold for children with VI. To our knowledge, there is only one study to date that reported reliability evidence for the TGMD-2 in this population (Houwen et al., 2007). In this study, the performance of 20 children with VI attending mainstream primary schools was assessed, and good interrater reliability for the total test was found. While these reliability results are satisfactory, internal consistency, intrarater reliability, and test-retest reliability for the total TGMD-2 have not been tested. Furthermore, as reliability levels may vary for different populations, the reliability of the TGMD-2 for primary-school-age children with VI needs further investigation. To date, no studies have investigated the validity of the TGMD-2 in children with VI. The aim of this present study accordingly was to examine the psychometric properties of the TGMD-2 in children with VI at primary school age.

Method

Participants

The current reliability and validity study was part of a larger project examining the motor skills, physical fitness, and physical activity of children with VI. Part of the children participating in the current study participated also in other studies (Houwen et al., 2008; Houwen et al., 2009; Houwen et al., in press). Participants were 75 children with VI attending either special-needs or mainstream primary schools. Children were included if they were between 6 and 12 years of age, were diagnosed with VI (best corrected visual acuity ≤ 6/18 in the better eye; WHO, 2007), and had no other known impairment(s) that could interfere with the aim of the study. Recruitment was conducted through Visio and Sensis, two regional organizations in the Netherlands, providing special education for children with VI or ambulant educational support (i.e., visiting teachers and resource services) for children with VI and their teachers at mainstream schools. The children's parents gave informed consent for their children's participation and all study procedures were in accordance with the ethical standards of the Faculty of Medical Sciences of the University Medical Center Groningen, University of Groningen.

To establish interrater reliability, two examiners independently rated the video recordings of the TGMD-2 performance of 50 of the children. To determine its test-retest reliability, a subsample of 23 children was tested twice by the same examiner over a 2-week interval, with each child performing the two measurements on the same weekday and at the same time of day. For the intrarater reliability study, another group of 25 children was recruited whose videotapes were scored twice by the same examiner with an interval of one month, an interval we considered long enough to minimize the examiner's memory bias. For the internal consistency and validity analyses, the data of the two groups of children were combined resulting in a total sample of 75 children.

The demographic data of all participating children are listed in Table 1. The group consisted of 29 girls and 46 boys. Of these 75 children, eight children attended special-needs schools and 67 children attended mainstream primary schools. Seventy-one children were Caucasian while four were of Asian descent. Socioeconomic status (SES) was based on the participant's residential postal code. The VI data were extracted from the children's medical records. Children were classified as having severe VI when their visual acuity was less than 20/200 (0.10 with correction; i.e., the visual acuity of the better eye at 20 ft is equal to what a child without VI can see at a distance of 200 ft) but greater than 20/400 (0.05). When visual acuity was less than 20/60 (0.30) but more than 20/200 with correction this was denoted as moderate VI.

Measures

Test of Gross Motor Development-2 (TGMD-2).   The TGMD-2 (Ulrich, 2000) is a qualitative measure to assess the gross motor skills of children aged 3–10 years. The 12 skills tested are subdivided into two skill areas: locomotor skills (run, gallop, hop, leap, horizontal jump, and slide) and object control skills (two-hand strike, stationary bounce, catch, kick, overhand throw, and underhand roll). Participants execute each skill twice. To indicate skill performance qualitative performance criteria (three to five, depending on the skill) are scored, with 1

Table 1   Demographic Data for the Participating Children per Test Conducted (*N* = 75)

	Assessment of interrater reliability (*n* = 50)	Assessment of test-retest reliabililty (*n* = 23)	Assessment of intrarater reliability (*n* = 25)	Assessment of validity and internal consistency (*n* = 75)
Age (yrs)				
Mean (*SD*)	8.3 (1.8)	8.8 (2.1)	8.8 (1.8)	8.5 (1.8)
Range	6–12	6–12	6–12	6–12
Sex, n (%)				
Female	20 (40)	7 (30)	9 (36)	29 (39)
Male	30 (60)	16 (70)	16 (64)	46 (61)
Ethnicity, n (%)				
Caucasian	46 (92)	22 (96)	25 (100)	71 (95)
Asian	4 (8)	1 (4)	0 (0)	4 (5)
Education, n(%)				
SpEd	8 (16)	8 (35)	0 (0)	8 (11)
MainsEd	42 (84)	15 (65)	25 (100)	67 (89)
SES, n (%)				
Lower SES	22 (44)	13 (57)	7 (28)	29 (39)
Medium SES	19 (38)	7 (30)	12 (48)	31 (41)
Upper SES	9 (18)	3 (13)	6 (24)	15 (20)
Degree of VI, n (%)				
Moderate	30 (60)	16 (70)	18 (72)	48 (64)
Severe	20 (40)	7 (30)	7 (28)	27 (36)

Note. SpEd = special education; MainsEd = mainstream education; SES = socioeconomic status; VI = visual impairment.

indicating its presence and 0 its absence. If a skill is assessed using three performance criteria, the raw scores can thus vary between 0–6. The highest raw total subtest score for the locomotor as well as the object control skills is 48.

Movement Assessment Battery for Children (Movement ABC).   The Movement ABC (Henderson & Sugden, 1992) is a test battery designed for the assessment of the everyday motor competence of children aged 4–12 years. There are four age-related item sets (age band 1: 4–6 years, age band 2: 7–8 years, age band 3: 9–10 years, and age band 4: 11–12 years), each consisting of eight items, with three items measuring manual dexterity, two ball skills, and three static and dynamic balance. The items are scored 0–5, 0 being the optimal score. The test has acceptable validity and reliability (Henderson & Sugden; Smits-Engelsman, Fiers, Henderson, & Henderson, 2008). We tested the instrument's test-retest reliability in a pilot-study with 15 6-to-12-year-old children (10 boys, 5 girls) with VI (none of whom were included in the current study). Five children had severe

VI and 10 had moderate VI. We found the Movement ABC test-retest reliability over a 2-week interval to be good (ICC > 0.80).

Procedure

All test sessions were conducted by two human movement scientists, both adequately trained in the use of the tests. The principal examiner had earlier evaluated TGMD-2 and Movement ABC outcomes of more than 200 children, while the second examiner was thoroughly trained in the two tests before the data collection (training included familiarization with all procedures and scoring methods and the viewing and scoring of existing video-recordings of test sessions).

The children performed the TGMD-2 and Movement ABC individually at the school gym with their performance being videotaped throughout. Both tests had been slightly adapted to meet the needs of our target group. Care was taken that none of the modifications involved changes likely to influence the actual motor skills required to complete the task. The adaptations included the use of brightly colored materials that contrasted sharply (e.g., bright yellow balls in the object control test). If appropriate, the children were also allowed to feel the items before performing the task. When necessary, the examiner let the child "feel" the required movement and gave additional instructions before the actual test.

Data Analysis

Reliability.   Internal consistency of the locomotor and object control subtests was established by calculating Cronbach's alpha, corrected item-total correlations, and interitem correlations. The resultant alpha (if an item was deleted) was determined. A Cronbach's alpha of 0.70 was considered acceptable and item-total correlations were set at > 0.30 (Nunally, 1978; Nunally & Bernstein, 1994). A high interitem correlation (> 0.80) is an indication of redundancy and is therefore undesirable. In contrast, if all correlations are near 0, there is no meaningful construct.

Interrater, intrarater, and test-retest reliability were assessed using the intraclass correlation coefficient (ICC) and the Bland and Altman method simultaneously (Rankin & Stokes, 1998). The dependent variables were the locomotor and object control subtest raw scores. The ICC gives a relative index of the ratio of variance among participants to the variance among participants plus error variance. As an indication of reliability the ICC (model 2,1) also accounts for all the (systematic and random) variance in the measurements. Ninety-five percent confidence intervals (95% CI) were determined for all the ICCs (Rankin & Stokes, 1998). There are no universally accepted criteria for reliability coefficients, but we have adopted the often used criteria recommended by Fleiss (1999) that consider ICCs of ≥ 0.75 excellent, < 0.40 are poor, and those between these two ranges moderate to good.

The Bland and Altman method consists of several analyses that are designed to detect differences between measurements. An estimate of the strength of any bias in the measurement is obtained by calculating the 95% CI for the mean difference, using the standard error of the mean difference. If zero fell within the 95% CI of the mean difference, we concluded that no bias existed between the measurements (Rankin & Stokes, 1998). The Bland and Altman method also includes the formation of plots. The differences between measurement 2 and 1 (2 minus 1) are plotted against the average of test measurement 1 and 2, i.e., the mean. The possibility of

heteroscedasticity was addressed by computing the Pearson's correlation coefficient of the absolute differences (i.e., ignoring any sign) between measurements 2 and 1 and the mean of the two test measurements.

Validity.   The construct validity of the TGMD-2 was addressed, i.e., the degree to which the underlying traits of a test could be identified and the extent to which these traits reflect the theoretical model on which the test is based (Streiner & Norman, 1996). Given that the TGMD-2 has a proposed factor structure (Ulrich, 2000), we employed confirmatory factor analysis (CFA) to empirically test the validity of this factor structure in the study sample. CFA using maximum likelihood estimation methods was conducted on correlation matrices using LISREL 8.51 (2001).

The goodness-of-fit of the model was assessed using several criteria: chi-square, chi-square/Df ratio, the goodness of fit index (GFI), and the root-mean-square error of approximation (RMSEA). A nonsignificant *p* value for chi-square means that there is no reliable difference (i.e., the model is a good representation of the data) between the sample matrix (i.e., the data) and the fitted matrix (i.e., the model) and thus provides evidence in favor of the model under investigation (especially in smaller models). With regard to the chi-square/Df ratio, a value close to 1 indicates that the model is correct (Byrne, Shavelson, & Muthén, 1989). Typically, RMSEA values at or below 0.08 are considered acceptable and values at or below 0.05 good (e.g., Browne & Cudeck, 1992). Generally, a GFI value greater than 0.90 is indicative of a good model fit.

Evidence of the construct validity of the test's capacity to detect developmental changes in motor skill performance should be supported when performance on the TGMD-2 is correlated with chronological age. We adjusted the TGMD-2 raw scores for variables that differed among the categories (sex) and variables that might affect motor skill performance (degree of VI). Therefore, partial correlation analyses were used to examine the relationship between age and locomotor and object control raw subtest scores. To determine sex-specific performance, we used analysis of covariance (ANCOVA) with age and degree of VI as covariates. All the assumptions for this parametric analysis were met.

In addition, we analyzed the convergent validity (criterion-related evidence for validity) of the TGMD-2, which indicates the relationship between test scores and outcome criteria that are analyzed at the same time (Morgan, Gliner, & Harmon, 2001). Convergent validity of the object control subtest was addressed by examining Spearman Rho correlations between the object control subtest of the TGMD-2 and the ball skills subtest of the Movement ABC. The analyses were conducted separately for the four age bands (6, 7–8, 9–10, and 11–12 years) because the Movement ABC has different formats for each age band. The analyses were also additionally adjusted for degree of VI. Data analysis was conducted using SPSS for Windows 15.0 and the significance level was set at 0.05.

Results

Reliability

Cronbach's alpha for the TGMD-2 locomotor subtest was 0.71 with alpha ranging from 0.61 to 0.68 if an item was deleted. For the object control subtest this was 0.72 with alpha ranging from 0.63 to 0.71 if an item was deleted. Item-total correlations ranged from 0.36 to 0.54 for the locomotor items and from 0.30 to 0.54

for the object control items. Interitem correlations ranged from 0.11 to 0.45 (with mean interitem correlation of 0.29) for the locomotor items and from 0.14 to 0.51 (with mean interitem correlation of 0.30) for the object control items.

Table 2 shows the ICCs and corresponding 95% CIs for interrater, intrarater, and test-retest reliability. The range of the values was 0.82 to 0.95. Excellent levels for all three reliability measures were seen for both subtests (Fleiss, 1999).

Table 3 shows the between-raters and the within-rater means and SDs for the locomotor and object control subtests at the first and second assessment, as well as the mean difference and SD_{diff}, standard error, and 95% CIs of the mean difference. The analysis of the interrater, intrarater, and test-retest reliability according to the Bland and Altman method showed that the mean differences between the two measurements were low and that there was no systematic bias for either the locomotor or the object control subtests.

Figures 1 and 2 show the Bland and Altman between-raters and within-rater plots, respectively, while Figure 3 comprised the Bland and Altman plots for the test-retest data. With regard to interrater reliability, the plots demonstrated a roughly equal distribution above and below the zero line for the TGMD-2 object control subtest. For the locomotor subtest, there were slightly more values below than above the zero line, suggesting that the second examiner had allotted somewhat higher scores than the first examiner. In regard to the intrarater reliability results, there were slightly more values above than below the zero line for the locomotor subtest, suggesting that the examiner's first round of ratings were somewhat higher than the second round. The opposite was seen for the object control subtest. For the test-retest data, the plots demonstrated a roughly equal distribution above and below the zero line for the locomotor subtest. For the object control subtest, there were

Table 2   Intraclass Correlation Coefficients (ICC) and 95% Confidence Interval (95% CI) for Interrater Reliability (*n* = 50), Intrarater Reliability (*n* = 25), and Test-Retest Reliability (*n* = 23) of the Two TGMD-2 Subtests and the Total Test

	ICC	95% CI
Interrater reliability		
Locomotor	0.82	0.70–0.90
Object control	0.93	0.88–0.96
Total test	0.89	0.81–0.93
Intrarater reliability		
Locomotor	0.85	0.69–0.93
Object control	0.93	0.84–0.97
Total test	0.95	0.88–0.98
Test-retest reliability		
Locomotor	0.86	0.70–0.94
Object control	0.87	0.72–0.94
Total test	0.92	0.82–0.91

Note. TGMD-2 = Test of Gross Motor Development-2.

Table 3   Results of the Bland and Altman Method for the Interrater (*n* = 50), Intrarater (*n* = 25), and Test-Retest (*n* = 23) of the TGMD-2

	T1		T2		d		SE of d	95%CI
	M	*SD*	*M*	*SD*	*M*	*SD*		
Interrater								
Locomotor	31.86	5.24	32.36	5.19	-0.50	3.12	0.44	-1.38–0.39
Object control	29.36	7.43	28.98	7.20	0.38	2.77	0.39	-0.41–1.17
Total test	61.26	10.87	61.34	10.73	-0.08	5.13	0.72	-1.54–1.38
Intrarater								
Locomotor	33.00	4.20	32.48	4.70	0.52	2.43	0.49	-0.48–1.52
Object control	30.96	7.21	30.92	6.39	0.04	2.61	0.52	-1.12–1.12
Total test	63.40	9.63	63.96	10.16	0.56	3.29	0.66	-0.86–1.92
Test-Retest Reliability								
Locomotor	31.00	7.79	30.83	7.44	0.17	4.06	.85	-1.58–1.93
Object control	28.30	9.41	29.70	7.97	-1.39	4.47	0.93	-3.32–0.54
Total test	59.30	16.28	60.52	13.94	-1.22	6.07	1.27	-3.84–1.41

Note. M = mean; *SD* = standard deviation; T1 = measurement 1; T2 = measurement 2; d = difference between measurements; *SE* of d = standard error of measurements; CI = confidence interval.

slightly more values below than above the zero line, suggesting that performance was slightly better during the second test session than during the first test session. The Pearson's correlation coefficients between the absolute difference between measurements 2 and 1 and the mean of the two measurements were low and not significant ($p > 0.05$) for all three reliability measures.

Construct Validity

The fit indices showed a sufficient fit: Chi-square (Df = 53) = 79.55, $p = 0.01$, Chi-square/Df ratio = 1.50, RMSEA = 0.07, GFI = 0.85. As is displayed in Figure 4, all factor loadings were statistically significant ($p < 0.05$) and they ranged from 0.32 to 0.76. In addition, the correlation between the two factors yields a high positive value (0.81). The construction of an alternative model in which correlations between error terms for certain items were allowed did not result in substantial improvements in fit.

There were significant partial correlations between age and the locomotor ($r = 0.36, p = 0.002$) and object control ($r = 0.53, p < .001$) subtests, with the older children performing better than the younger children. ANCOVA showed a significant effect for sex in the object control subtest, $F(1, 71) = 8.54, p = 0.005$, but not in the locomotor subtest, $F(1, 71) = 0.03, p = 0.855$, with the boys doing better than the girls on the object control subtest. Age was also a significant covariate for both the locomotor and object control subtests: $F(1, 71) = 10.61, p = 0.002; F(1, 71) = 27.55, p < .001$, respectively. The degree of VI exerted no significant effect as a covariate on the locomotor and object control scores: $F(1, 71) = 0.22, p = 0.644; F(1, 71) = 2.44, p = 0.123$, respectively.
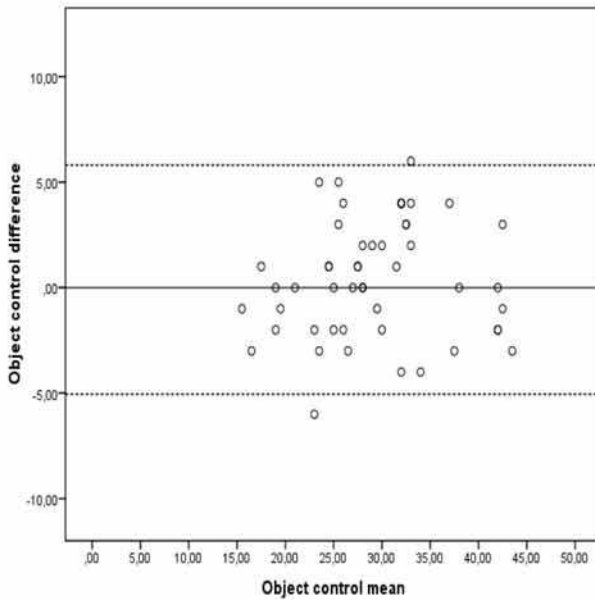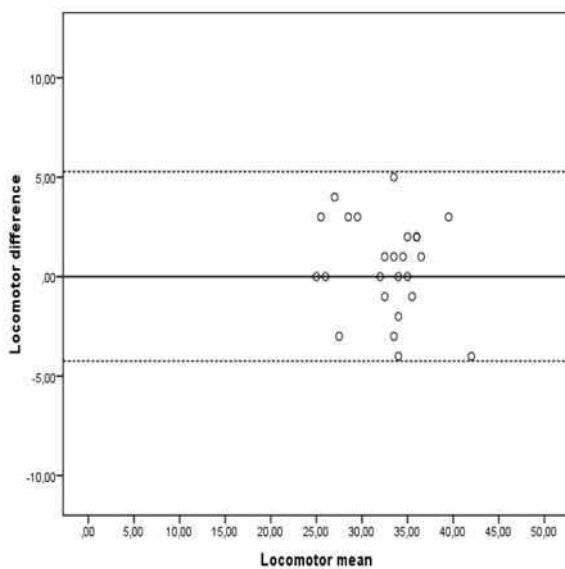
A
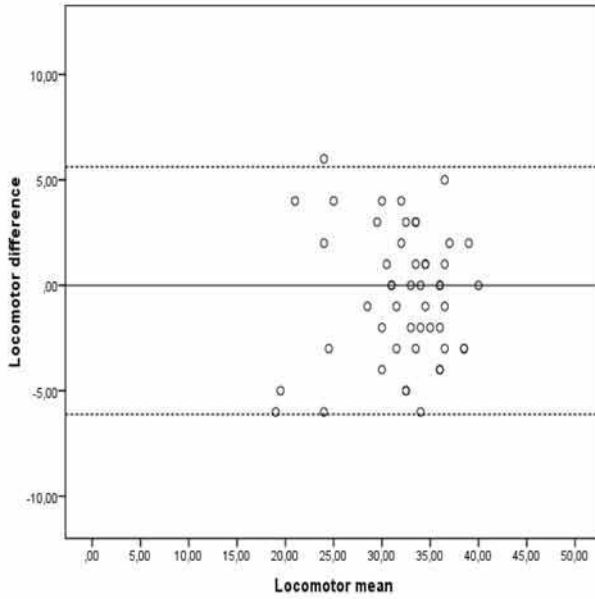
B

Figure 1 — Bland and Altman plots for interrater reliability of the TGMD-2 locomotor subtest (Panel A) and the TGMD-2 object control subtest (Panel B). The dotted lines indicate the limits of agreement (± 1.96 SD).
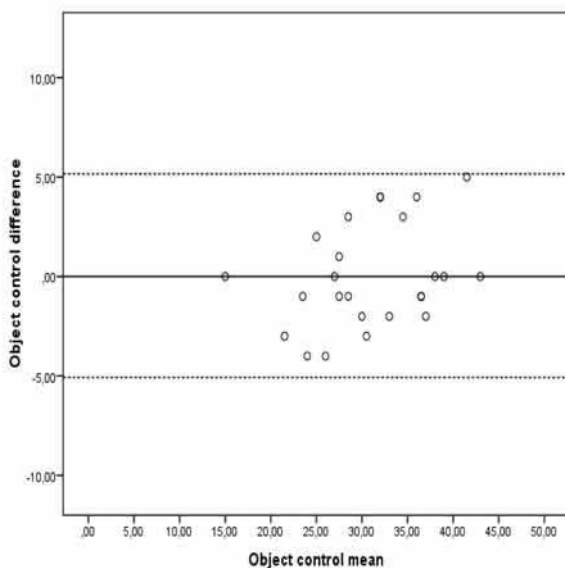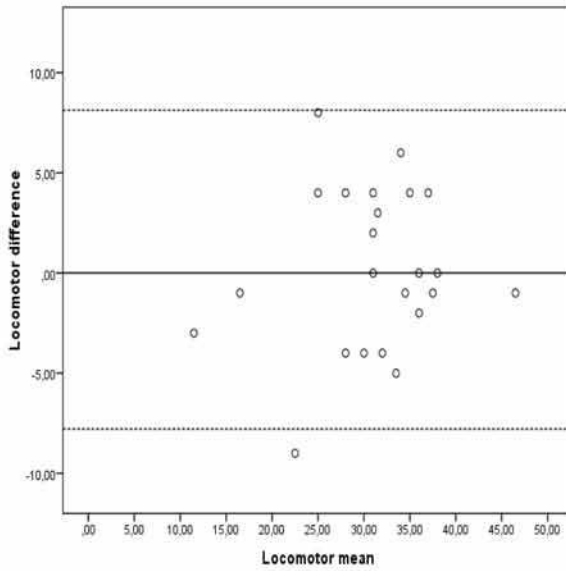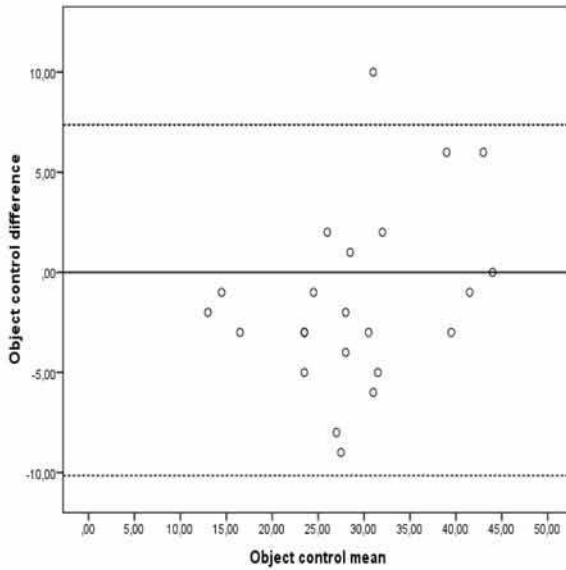
A

B

Figure 2 — Bland and Altman plots for intrarater reliability of the TGMD-2 locomotor subtest (Panel A) and the TGMD-2 object control subtest (Panel B). The dotted lines indicate the limits of agreement (± 1.96 SD).

A

B

Figure 3 — Bland and Altman plots for test-retest reliability of the TGMD-2 locomotor subtest (Panel A) and the TGMD-2 object control subtest (Panel B). The dotted lines indicate the limits of agreement (± 1.96 SD).
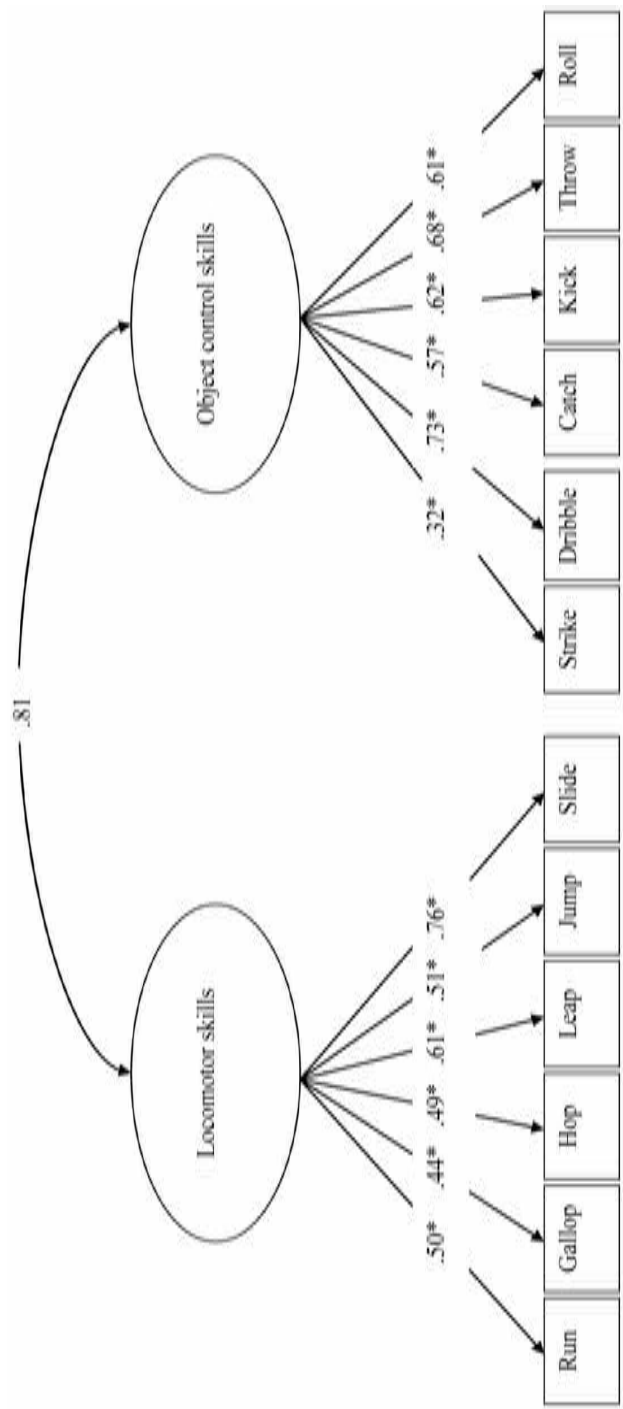
Figure 4 — Hypothesized two-factor model of the TGMD-2. * p < .05.

Convergent Validity

For age bands 1 and 4, the Spearman Rho correlation coefficient for the TGMD-2 object control subtest and the Movement ABC ball skills subtest indicated a strong relationship between the two measures ($r_s = 0.80$, $p = 0.006$ and $r_s = 0.76$, $p = 0.007$, respectively). For age bands 2 and 3, the correlations were fair ($r_s = 0.57$, $p = 0.001$ and $r_s = 0.45$, $p = 0.040$, respectively).

Discussion

In our study we examined the psychometric properties of the TGMD-2 to establish its suitability to assess the gross motor skills in primary-school-age children with VI.

Reliability

The internal consistency of the TGMD-2 was found to be high and the interrater, intrarater, and test-retest reliability acceptable. Our finding that the ICCs for intrarater reliability were generally somewhat higher than those for interrater reliability is in line with the literature (Portney & Watkins, 1993); it is more likely that an individual examiner will agree more consistently with him or herself than with other examiners. The variability in the ratings of the two examiners may be attributable to individual differences in interpretation of the scoring criteria and the instructions for particular items (Parkkinen & Rintala, 2004).

Although our trial in children with VI showed that the overall reliability of the TGMD-2 was satisfactory, superior values were obtained in populations without VI (Simons et al., 2008; Ulrich, 2000). Direct comparison of results is, however, difficult due to the differences in the statistical analysis methods applied. In their reliability studies, Simons and colleagues and Ulrich computed Pearson and Spearman correlation coefficients, while in the current study we used ICCs in combination with the Bland and Altman method, both considered the more stringent procedures and hence generating lower values than those obtained with Pearson and Spearman correlation coefficients. The Bland and Altman analyses revealed large interindividual variability, which is likely the result of the large age range and the heterogeneity (typical of VI populations) of our sample.

Validity

Two earlier studies investigated the test's underlying structure in typically developing children and found support for a two-factor model (Evaggelinou et al., 2002; Ulrich, 2000). We found this model to fit sufficiently well, although the results we obtained should be interpreted with caution given the relatively small sample size. The underlying factor structure of the TGMD-2 may uncover that the pattern of strengths and weaknesses in the motor skill performance of children with VI may be more complex and that a simple discrimination between locomotor and object control performance does not suffice.

The significant correlations between age and the locomotor and object control subtest scores we found were lower than the ones Ulrich (2000) reported for typically developing children. Possibly, the qualitative level of gross motor performance

of children with VI is lower than that of typically developing children (Houwen et al., 2007). In addition, the development of these skills may be slowed in children with VI (Jan et al., 1990; Whitall, 2003). Alternatively, our results may have been affected by the disproportion in the age groups and our relatively small sample size. Nevertheless, our results still suggest that the TGMD-2 has sufficient validity to assess age-related development of gross motor skills in children with VI.

The object control scores revealed sex-related differences, with the boys having higher scores than the girls, confirming the notion that performance in this area differs between the sexes (Okely & Booth, 2004; Woodard & Surburg, 1997) and supporting the validity of the TGMD-2 as a tool that distinguishes between the two groups. Our results suggest that especially girls with VI need encouragement and dedicated practice for optimal development of these specific skills.

Adjustment for degree of VI did not change the effects of age and sex on the TGMD-2. Although it is logical to assume that the severity of the impairment would play a major role in the ability to perform motor skills, there is no clear, linear relationship between the two, which was also corroborated by the absence of performance differences on the TGMD-2 between 6-to-11-year-old children with moderate and those with severe VI in our earlier study (Houwen et al., 2007); however, the homogeneity with regard to the degree of VI within the sample may account for the lack of an association between severity of VI and the TGMD-2 scores.

The results of our convergent validity study, with a mean correlation between the two subtests of 0.65, indicate that the object control subtest of the TGMD-2 and the ball skills subtest of the Movement ABC measure a similar construct. Although a correlation of this magnitude supports the convergent validity of the two tests, it should also be noted that a substantial part of the variance on one test cannot be explained by the scores on the other test. A major drawback to criterion validity is the problem of identifying and then being able to measure a suitable criterion (Morgan et al., 2001).

Study Limitations

In the current study, we used video-based performance assessments to determine the reliability of the test under review, which has the advantage that task performance can be scrutinized repeatedly and at leisure. Video assessments, however, have the drawback that the recordings show the child's performance from a single, fixed vantage point without the experimenter being able to choose the best possible viewpoint (Smits-Engelsman et al., 2008). Although we recognize the inherent limitation of the use of video recording, nevertheless, we also believe that the advantages of videotaping (stability of measurements and easy scoring) outweigh the disadvantages.

Another point of attention is the use of two raters. We might have obtained more generalizable results if we had chosen to involve more raters. The use of a larger number of raters could reduce the potential for the effects of between-rater variance on the precision of ICCs (Saito, Sozu, Hamada, & Yoshimura, 2006). Nevertheless, most other studies examining the interrater reliability of the TGMD-2 also used two raters (Butterfield, Angel, & Mason, 2008; Pollatou, Karadimou, & Gerodimos, 2005; Simons et al., 2008).

We have not addressed content-related validity evidence for the TGMD-2. For motor skills to pass the criterion of functionality, they must be usable in daily activities, including physical education and after-school leisure activities (Sherrill, 2004). Clinicians and educators should question whether a child with VI uses a certain skill in daily activities or play and sport situations. It has been suggested that the TGMD-2 may be particularly suitable to test the motor performance of primary-school-age children with VI (Shapiro et al., 2003).

Finally, we also tested 11- and 12-year-olds, an age group that falls outside the suggested age-range for TGMD-2 test application. Using raw scores rather than standardized scores allowed us to include these data in our analyses. As none of the children over age 10 obtained maximal scores, including their data did not induce ceiling effects.

Conclusion

By specifically focusing on children with VI between the ages of 6 and 12 years, the current study adds to the existing evidence of the reliability and validity of the TGMD-2 in testing gross motor skill performance in children and young adolescents. Based on our findings, we conclude that the TGMD-2 is an appropriate tool to examine the gross motor skills in this population. Nevertheless, additional studies using larger samples are needed before any definitive conclusions can be drawn.

Acknowledgments

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Bouchard, D. & Tetreault, S. (2000). The motor development of sighted children and children with moderate low vision aged 8-13. *Journal of Visual Impairment & Blindness, 94,* 564–573.

Branta, C., Haubenstricker, J., & Seefeldt, V. (1984). Age changes in motor skills during childhood and adolescence. *Exercise and Sport Sciences Reviews, 12,* 467–520.

Browne, M.W. & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21,* 230–258.

Butterfield, S., Angel, R.M., & Mason, C.A. (2008). Object control skill performances by children ages 5-14 years. *Research Quarterly for Exercise and Sport, 79,* A38.

Byrne, B.M., Shavelson, R.J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105,* 456–466.

Davies, P.L., & Rose, J.D. (2000). Motor skills of typically developing adolescents: Awkwardness or improvement? *Physical & Occupational Therapy in Pediatrics, 20,* 19–42.

Evaggelinou, C., Tsigilis, N., & Papa, A. (2002). Construct validity of the Test of Gross Motor Development: A cross-validation approach. *Adapted Physical Activity Quarterly, 19,* 483–495.

Fleiss, J.L. (1999). *The design and analysis of clinical experiments*. New York: Wiley.

Gliner, J.A., Morgan, G.A., & Harmon, R.J. (2001). Measurement reliability. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 486–488.

Graf, C., Koch, B., Kretschmann-Kandel, E., Falkowski, G., Christ, H., Coburger, S., et al. (2004). Correlation between BMI, leisure habits and motor abilities in childhood (CHILT-Project). *International Journal of Obesity, 28,* 22–26.

Henderson, S.E., & Sugden, D.A. (1992). *Movement Assessment Battery for Children: Manual*. London, UK: Psychological Corporation.

Houwen, S., Hartman, E., & Visscher, C. (2009). Physical activity and motor skills of children with and without visual impairments. *Medicine and Science in Sports and Exercise, 41,* 103–109.

Houwen, S., Hartman, E., & Visscher, C. (in press). Examining the relationship among motor proficiency, physical fitness, and body composition in children with visual impairments and sighted children. *Research Quarterly for Exercise and Sport*.

Houwen, S., Visscher, C., Hartman, E., & Lemmink, K.A.P.M. (2007). Gross motor skills and sports participation of children with visual impairments. *Research Quarterly for Exercise and Sport, 78,* 16–23.

Houwen, S., Visscher, C., Lemmink, K.A.P.M., & Hartman, E. (2008). Motor skill performance of school-age children with visual impairments. *Developmental Medicine and Child Neurology, 50,* 139–145.

Jan, J.E., Sykanda, A., & Groenveld, M. (1990). Habilitation and rehabilitation of visually impaired and blind children. *Paediatrician, 17,* 202–207.

Morgan, G.A., Gliner, J.A., & Harmon, R.J. (2001). Measurement validity. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 729–731.

Morgan, P.J., Okely, A.D., Cliff, D.P., Jones, R.A., & Baur, L.A. (2008). Correlates of objectively measured physical activity in obese children. *Obesity (Silver Spring, Md.), 16,* 2634–2641.

Nunally, J.C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Nunally, J.C., & Bernstein, I.H. (1994). *Psychometric theory*. New York: McGraw-Hill.

Okely, A.D., & Booth, M.L. (2004). Mastery of fundamental movement skills among children in New South Wales: Prevalence and sociodemographic distribution. *Journal of Science and Medicine in Sport, 7,* 358–372.

Parkkinen, T., & Rintala, P. (2004). Primary school teachers' and physical education teachers' accuracy in assessing children's gross motor performance. *European Bulletin of Adapted Physical Activity*, 3. Retrieved November 18, 2008, from http://www.eufapa.upol.cz/www/images/stories/files/Journals/Compl_Comm1_vol3_iss1.pdf.

Pollatou, E., Karadimou, K., & Gerodimos, V. (2005). Gender differences in musical aptitude, rhythmic ability and motor performance in preschool children. *Early Child Development and Care, 175,* 361–369.

Portney, L.G., & Watkins, M.P. (1993). *Foundations of clinical research*. Norwalk, CT: Appleton & Lange.

Rankin, G., & Stokes, M. (1998). Reliability of assessment tools in rehabilitation: An illustration of appropriate statistical analysis. *Clinical Rehabilitation, 12,* 187–199.

Reimer, A.M., Smits-Engelsman, B.C.M., & Siemonsma-Boom, M. (1999). Development of an instrument to measure manual dexterity in children with visual impairments aged 6-12. *Journal of Visual Impairment & Blindness, 93,* 643–658.

Saito, Y., Sozu, T., Hamada, C., & Yoshimura, I. (2006). Effective number of subjects and number of raters for inter-rater reliability studies. *Statistics in Medicine, 25,* 1547–1560.

Shapiro, D.R., Lieberman, L.J., & Moffett, A. (2003). Strategies to improve perceived competence in children with visual impairments. *RE:view, 35,* 69–80.

Sherrill, C. (2004). *Adapted physical activity, recreation and sport: Crossdisciplinary and lifespan* (6th ed.). Madison, WI: WCB McGraw-Hill.

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

Smits-Engelsman, B.C.M., Fiers, M.J., Henderson, S.E., & Henderson, L. (2008). Interrater reliability of the Movement Assessment Battery for Children. *Physical Therapy, 88,* 286–294.

Streiner, D.L., & Norman, G.R. (1996). Validity. In D.F. Streiner & G.R. Norman (Eds.), *Health measurement scales* (pp. 144–162). Oxford: Oxford Medical Publications.

Thomas, J.R., & French, K.E. (1985). Gender differences across age in motor performance: A meta-analysis. *Psychological Bulletin, 98,* 260–282.

Ulrich, D.A. (2000). *Test of Gross Motor Development* (2nd ed.). Austin, TX: Pro-Ed.

Whitall, J. (2003). Development of locomotor coordination and control in children. In G. Savelsbergh, K. Davids, J. van der Kamp, & S.J. Bennett (Eds.), *Development of movement coordination in children: Applications in the fields of ergonomics, health sciences and sport* (pp. 251–270). London: Routledge.

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

World Health Organization. (WHO; 2007). *International statistical classification of diseases and related health problems*. Tenth revision. Retrieved April 2, 2009, from http://www.who.int/classifications/apps/icd/icd10online.

Yun, J., & Ulrich, D.A. (2002). Estimating measurement validity: A tutorial. *Adapted Physical Activity Quarterly, 19,* 32–47.