Construct Validity of the Test of Gross Motor Development: A Cross-Validation Approach

Christina Evaggelinou
Aristotle University of Thessaloniki, Greece

Nikolaos Tsigilis
University of Thessaly, Greece

Areti Papa
Aristotle University of Thessaloniki, Greece

This study was designed to examine the underlying structure of the Test of Gross Motor Development (TGMD) in Ulrich (1985). The TGMD was administered to 644 children who were randomly divided into two groups (calibration group and validation group). The calibration group ($n = 324$) included 150 boys and 174 girls, and the validation group included 160 boys and 160 girls, ranging from 3 to 10 years. A two-factor model was postulated and supported. According to the model, seven variables measuring children’s ability for moving into space loaded on one factor (locomotor skills), while five variables measuring children’s ability for controlling objects loaded on the other factor (object control skills). In addition, the proposed model was found to be invariant across the two groups. Good cross-generalizability of the TGMD appears to support its validity. Physical educators working with young children may use it with confidence when assessing and planning physical education programs involving locomotor and object control skills.

Physical education programs in the elementary school are an integral part of the total school curriculum. Physical educators working with young children with or without disabilities must constantly make decisions about the quality of services to be provided. Systematic assessment and use of reliable and valid instruments is an essential aspect of any planning process and can lead to a more effective and efficient program of services (Bailey & Wolery, 1989).

Gross motor development is a major component of most preschool and elementary education programs, including those in special education (Ulrich, 1985).

Christina Evaggelinou and Areti Papa are with the Department of Physical Education and Sports Sciences, Aristotle University of Thessaloniki, Department of Physical Education and Sports Sciences, 62100, Serres, Greece. E-mail: <Evaggeli@phed-sr.auth.gr>. Nikolaos Tsigilis is with the Department of Physical Education and Sports Science at the University of Thessaly, Greece.
Bouchard, McPherson, and Taylor (1991) point out the need for a strong emphasis on developing fundamental motor skills in the early years with a gradual introduction of physical fitness activities during later years. Gross motor development includes skills that are used to move the body through space (locomotor skills), to maintain balance against the force of gravity (stability), and to give and receive force from an object (manipulation; Gallahue & Ozmun, 1998). Today, authorities agree that children’s gross motor skill development involves a continuous change through the life cycle brought about by interaction among the requirements of the task, the biology of the individual, and the conditions of the environment (Gallahue & Ozmun, 1998). Failure to develop fundamental motor skills during the critical years may constrain their ability for learning more advanced motor skills (Gallahue, 1996; Roberton, 1982; Seefeldt & Haubenstricker, 1982).

Although assessment of fundamental motor skill development is an important process, it is a complex one made difficult by current limitations in measurement instruments. Questions about the validity of the tests are of ultimate importance to physical educators because they ask whether an instrument fulfills the functions for which it was intended.

The Test of Gross Motor Development (TGMD) is a norm and criterion-referenced test and is considered one of the most frequently used tests in regular and adapted physical education research conducted in the United States (Burton & Miller, 1998; Sherrill, 1998). The test measures 12 gross motor skills that are grouped into two subtests. Each subtest is assumed to assess a distinct facet of gross motor development (Ulrich, 1985). The locomotor subtest is comprised of seven skills: run, gallop, hop, leap, horizontal jump, skip, and slide. These skills are assumed to represent different aspects of the child’s ability to transfer his or her center of gravity from one point to another. The object control subtest (or manipulative skills) measures five skills: two-hand strike, stationary bounce, catch, kick, and overhand throw. These skills are assumed to assess the child’s ability for projecting and receiving objects.

Previous studies examined various facets of the validity of the TGMD. Particularly, in the study by Ulrich (1983), three experts in motor development were asked to judge whether the test’s items (a) cover a representative sample of behaviors in the fundamental motor domain and (b) are frequently taught in preschool and early elementary grades. Results showed excellent content validity. In another study, Ulrich and Ulrich (1984) demonstrated that the test is sensitive to measuring change in performance after a 10-week formal instruction program in fundamental motor skills in preschoolers. Finally, Ulrich (1985) used factor analysis followed by varimax rotation to examine the underlying structure of the TGMD. Three factors emerged that explained 75% of the variability. However, the pattern of the items loading conceptually and statistically was not clear. There were four items (hop, leap, jump, and skip) that had similar loadings on the first two factors. For example, hop correlated .59 with the first factor and .57 with the second. In addition, only one item (run) loaded on the third factor. As a result, factor analysis of the performances on the TGMD did not succeed in identifying a theoretically defensible factor structure.

Construct validity is the most important property in the development of an instrument. It involves testing the adequacy of theoretically derived relationships. Ulrich’s (1985) explanation of the structure of the TGMD relied on exploratory factor analysis. The construct validity of the TGMD should also be tested and
confirmed using a theoretical model that has been specified a priori. Confirmatory factor analysis (CFA) is a frequently employed technique to evaluate the factorial validity of multidimensional instruments. To date, no attempt has been made to assess the underlying structure of the TGMD using CFA. Therefore, the purpose of the present study was to examine further the construct validity of the TGMD using confirmatory factor analysis. A cross-validation procedure was employed because it is an important step in building confidence in the validity of the TGMD. Cross-validation requires two distinct samples from the same population. The first sample is used to established a baseline model with an acceptable fit. This model is then cross-validated on the other sample using various structural equation modeling procedures.

**Method**

**Participants**

The total sample consisted of 644 school children attending school in 10 different prefectures in Greece. Their age ranged from 3 to 10 years. Participants were randomly divided into two subsamples. The first subsample \( (n = 324) \) consisted of 150 males (\( M \) age = 7.47 years, \( SD = 1.59 \)) and 174 females (\( M \) age = 7.72 years, \( SD = 1.69 \)) and was used as the calibration sample. Participants in the second subsample \( (n = 320) \) were 160 males (\( M \) age = 7.68 years, \( SD = 1.60 \)) and 160 females (\( M \) age = 7.68 years, \( SD = 1.51 \)) and they served as the validation sample. Preliminary examination showed that there were no differences between the two samples regarding age (\( t_{642} = 530, p = .596 \)), sex (\( \chi^2 = .885, p = .347 \)), community (\( \chi^2 = 1.157, p = .282 \)), and performance on the 12 items comprising the TGMD (Wilks Lamda = .991, \( F_{(12, 627)} = .494, p = .919 \)).

**Instrument**

The TGMD, developed by Ulrich (1985), evaluates the gross motor functioning of children 3 to 10 years of age. The test measures 12 gross motor skills frequently taught to children in preschool and early elementary classes. The TGMD includes two subscales. Seven skills comprise the locomotor subscale (run, gallop, hop, leap, jump, skip, and slide), with five skills in the object control subscale (two-hand strike, stationary bounce, catch, kick, and overhand throw) the object control subscale. The skill performance is evaluated by the sum of the points earned on each skill component (three or four components, depending on the skill). One point is awarded for each skill component, which is correctly performed. A perfect score on the locomotor subscale is 26 points and on the object control subscale, 19 points. The primary uses of this test are to (a) identify children who are significantly behind their peers in gross motor skill development and should be eligible for special education services in physical education, (b) derive information needed to plan an instructional program in gross motor skill development, (c) assess individual child progress in gross motor skill development, (d) evaluate the gross motor program, and (e) serve as a measurement instrument in research involving gross motor development. Previous studies seem to support the TGMD’s content validity (Ulrich, 1983) and its sensitivity to measuring changes in fundamental motor skills (Ulrich & Ulrich, 1984).
Procedure

One physical education teacher, who was trained and familiar with the testing and scoring procedures, administered the TGMD to 644 children. Each child was assessed individually, and the testing time varied from 15 to 30 min according to the age of the child. Encouragement was given in order to maximize the child’s efforts. The testing took place during physical education class in the school’s indoor facilities. The same equipment was used for all children. Children were asked to wear soled shoes and sport clothes. A video was used to record all trials and assist the examiner with the evaluation of the three or four performance criteria for each of the 12 fundamental gross motor skills. The analysis of the data was made by one of the researchers in the adapted physical education laboratory of the Aristotle University. The examiner received training in the assessment procedures of the TGMD by the author of this test. In addition, a second physical education teacher assessed the TGMD scores of a random sample of 90 children. The computed interrater agreement was very high, yielding a value of .98.

Model Tested

A two-factor model was postulated and tested. Specifically, it was hypothesized a priori that seven measured item variables (run, hop, gallop, skip, jump, leap, and slide) were manifestations of a latent variable “locomotor skills,” while the other five item variables (bounce, catch, kick, strike, throw) were manifestations of a latent variable “object control skills.” Therefore, each observed variable would have a nonzero loading on the factor it was designed to measure (hypothesis 1). It was also hypothesized that the two factors would be intercorrelated (hypothesis 2) and that the measurement error associated with each item variable would be uncorrelated (hypothesis 3). In addition, to provide further evidence that the proposed two factor model can adequately fit the observed data, three other models were also examined: a null model, a single-factor model, and a two factor model with uncorrelated factors.

Model Cross-Validation

A hierarchical approach to testing for invariance (Bollen, 1989; Byrne, 1994; Byrne, Shavelson, & Muthén, 1989; Li, Harmer, Duncan, Duncan, Accok, & Yamamoto, 1998; Markland, Emberton, & Tallon, 1997) was employed in order to examine the cross-validation of the TGMD. According to this procedure, equality constraints are imposed on a particular set of parameters. Next, invariance is tested by simultaneously fitting the proposed model to the data from multiple groups (Arbuckle, 1997; Byrne, 1994). Constraints are imposed in logically ordered and increasingly restrictive fashion. Finally, the model in which a certain set of parameters is forced to be equal across groups is compared with a less restrictive model in which the same parameters are free to yield any value. A nonsignificant χ² difference indicates that the invariance hypothesis may be considered tenable. At each stage, goodness of fit was also assessed. In the present study, the following set of parameters was examined in relation to group invariance: (a) equality of loading paths (hypothesis 1), (b) equality of factor variances/covariances (hypothesis 2), and (c) equality of error variances/covariances (hypothesis 3; Bollen, 1989; Byrne, 1994, Byrne et al., 1989; Li et al., 1998; Markland et al., 1997).
Data Analysis

Confirmatory factor analysis using AMOS 3.6 (Arbuckle, 1997) was employed in order to examine the construct validity of the TGMD. Exploratory data analysis, based on inspection of the univariate skewness and kurtosis values of the measured variables (Table 1) as well as the Kolmogorov-Smirnov test of normality, revealed that data for both samples were not normally distributed. In addition, Mardia’s coefficient of multivariate kurtosis provided by AMOS was statistically significant (Mardia’s coefficient = 77.48, p < .001 for the calibration sample and Mardia’s coefficient = 101.27, p < .001 for the validation sample), indicating that the assumption of multivariate normality was not tenable. Based on these results and because the present data were in ordinal scale, it was decided not to use the frequently applied maximum likelihood (ML) method of estimation. It is well known that ML procedure requires the observed variables to be continuous and normally distributed (Byrne et al., 1989; Hu, Bentler, & Kano, 1992; Li, Harmer, & Acock, 1996). Instead, the asymptotically distribution-free (ADF) estimation procedure was employed. Many previous studies have used a distribution free method of estimation when the nature of the data could not justify the application of the ML (Li et al., 1996; Li & Harmer, 1996; Li et al., 1998).

Assessment of Fit

The overall fit of the data to the examined models was initially based on the chi-square statistic. Nonsignificant values suggest a good fit, since they indicate minor discrepancy between the observed and the estimated covariance matrix. However, the chi-square statistic is very sensitive to sample size and departures from normality (Byrne, 1994; Li et al., 1998). For instance, in large samples, trivial differences between the two matrixes may be declared significant. In an attempt to overcome this problem, various alternative indices of fit have been proposed. Thus, in addition to the conventional $\chi^2$, multiple fit indexes were used. Many researchers have proposed the use of $\chi^2/df$ ratio as a measure of the model fit (e.g., Byrne et al., 1989; Kline, 1998). Values close to one are considered to represent a correct model. However, it is not clear how far from one you should let the ratio get before concluding that a model is unsatisfactory (Arbuckle, 1997). For example, according to Byrne (1989), ratio values greater than 2 represent an inadequate fit, while Kline (1998) recommends a $\chi^2/df$ values less than 3 in large samples and less than 2.5 in small samples represents an acceptable model fit. Goodness of fit index (GFI) and adjusted goodness of fit index (AGFI) belong to the class of absolute fit indexes (Hu, & Bentler, 1995). GFI can be interpreted in a similar fashion to the $R^2$ in multiple regression in that it indicates the proportion of the observed covariances explained by the model estimated covariances (Kline, 1998). Both GFI and AGFI indexes range from 0-1 and values greater than .90 are indicative of a good fitting model. It is known that the estimation (e.g., ML and ADF) influences the performance of $\chi^2$ and fit indexes. Hu and Bentler (1995) argue that “all ADF behave purely except GFI” (p. 96). Root mean square error of approximation (RMSEA) is another absolute fit index, which represents the standardized root mean square residual, corrected for model complexity (MacCallum, Roznowski, & Necowitz, 1992). RMSEA values less than .05 would indicate a close fit of the model while values of about .08 would indicate a reasonable fit (Brown & Cudeck, 1993).
Table 1  Univariate Descriptive Statistics for the Calibration and the Validation Sample

<table>
<thead>
<tr>
<th>TGMD variables</th>
<th>Calibration</th>
<th>Validation</th>
<th>Skewness $^a$</th>
<th>Kurtosis $^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Run</td>
<td>3.92</td>
<td>.33</td>
<td>3.94</td>
<td>.28</td>
</tr>
<tr>
<td>Gallop</td>
<td>3.19</td>
<td>.90</td>
<td>3.25</td>
<td>.87</td>
</tr>
<tr>
<td>Hop</td>
<td>3.51</td>
<td>.93</td>
<td>3.64</td>
<td>.79</td>
</tr>
<tr>
<td>Leap</td>
<td>2.27</td>
<td>.96</td>
<td>2.26</td>
<td>.94</td>
</tr>
<tr>
<td>Jump</td>
<td>3.14</td>
<td>.90</td>
<td>3.25</td>
<td>1.06</td>
</tr>
<tr>
<td>Skip</td>
<td>2.10</td>
<td>.98</td>
<td>2.13</td>
<td>.99</td>
</tr>
<tr>
<td>Slide</td>
<td>3.85</td>
<td>.55</td>
<td>3.88</td>
<td>.42</td>
</tr>
<tr>
<td>Strike</td>
<td>2.47</td>
<td>1.18</td>
<td>2.48</td>
<td>1.17</td>
</tr>
<tr>
<td>Bounce</td>
<td>2.33</td>
<td>1.05</td>
<td>2.38</td>
<td>.97</td>
</tr>
<tr>
<td>Catch</td>
<td>3.41</td>
<td>.94</td>
<td>3.48</td>
<td>.92</td>
</tr>
<tr>
<td>Kick</td>
<td>2.31</td>
<td>.90</td>
<td>2.32</td>
<td>.94</td>
</tr>
<tr>
<td>Throw</td>
<td>2.70</td>
<td>1.42</td>
<td>2.77</td>
<td>1.34</td>
</tr>
</tbody>
</table>

Notes. $^a$values $>.03$ are significant at $p < .05$.

$^b$values $>1.05$ are significant at $p < .05$ for the calibration group and values $>1.04$ are significant at $p < .05$ for the validation group.

$^c$calibration group.

$^v$validation group.
Results

Table 1 presents univariate descriptive statistics for the 12 test items comprising the TGMD for both calibration and validation samples.

Table 2 displays the goodness of fit measures for the three models tested based on the calibration sample. The first important piece of information is the substantial drop in the overall \( \chi^2 \) value for the single model factor, compared to the null model. It is well known that when models are nested, as in our study, improvement in model fit can be tested. The \( \chi^2 \) difference between the two models (\( \Delta \chi^2 \)) is itself \( \chi^2 \)-distributed, with degrees of freedom equal to the corresponding difference in degrees of freedom. Therefore, it appears that the single-factor model represented a better fit of the data when compared to the null model (\( \Delta \chi^2_{(12)} = 202.745, p < .001 \)).

The derived solution for the two-factor model with uncorrelated factors was not admissible (there were negative error variances) and so excluded from further analysis. Results for the two-factor model with correlated factors revealed a decrement in the chi-square value relative to the single factor model. The difference in \( \chi^2 \) between the single-factor model and the two-factor model with correlated factors was again statistically significant (\( \Delta \chi^2_{(1)} = 4.386, p < .05 \)), indicating a substantial improvement of model fit. Further, alternative measures of fit indices yield acceptable values. Based on the above results the two-factor model was considered the most viable for the calibration sample data.

According to Byrne and colleagues (1989), however, invariance testing for multiple groups requires a well fitting baseline model. Thus, in order to identify possible sources of misfit, the modification indexes provided by AMOS were examined. Results revealed that the model would be significantly improved if an error covariance between the observed variables strike and throw would be included. Inspection of the chi-square statistic for the two-factor model with correlated errors indicated a substantial decrement. The \( \chi^2 \) difference (\( \Delta \chi^2 \)) between the two models (see Table 2) was significant (\( \Delta \chi^2_{(1)} = 23.63, p < .001 \)), revealed a substantial improvement in model fit. Moreover, the fit indexes were substantially

<table>
<thead>
<tr>
<th>Model</th>
<th>( \chi^2 )</th>
<th>df</th>
<th>( \chi^2/df )</th>
<th>GFI</th>
<th>AGFI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null model</td>
<td>326.95*</td>
<td>66</td>
<td>4.95</td>
<td>.88</td>
<td>.80</td>
<td>.11</td>
</tr>
<tr>
<td>Single-factor model</td>
<td>124.21*</td>
<td>54</td>
<td>2.30</td>
<td>.94</td>
<td>.91</td>
<td>.06</td>
</tr>
<tr>
<td>Two-factor model</td>
<td>119.82*</td>
<td>53</td>
<td>2.26</td>
<td>.94</td>
<td>.91</td>
<td>.06</td>
</tr>
<tr>
<td>Two-factor model with correlated errors</td>
<td>96.19*</td>
<td>52</td>
<td>1.85</td>
<td>.95</td>
<td>.93</td>
<td>.05</td>
</tr>
</tbody>
</table>

Note. GFI = goodness of fit index. AGFI = adjusted goodness of fit index. RMSEA = root mean square of error approximation. *denotes significant values.
improved, indicating an acceptable fit. All factor loadings were statistically significant \((p < .05)\) and they ranged from .40 to .81. In addition, internal consistency based on Crobach’s \(\alpha\) was acceptable for both factors (.77 and .75 for the locomotor and object control factors, respectively).

**Cross-Validation of the TGMD**

Results from the cross-validation procedure are presented in Table 3. Validation sample analysis yielded similar results to the calibration sample. Alpha reliability coefficient was in the acceptable range for both factors (.75 and .74 for locomotor and object control factors respectively). Thus, the model derived from the calibration sample could be replicated using the validation sample. The next series of analyses represent the examination of the invariance hypothesis between the two groups.

The chi-square statistic regarding the model with the same factor loadings for the calibration and validation group (model 1) was statistically significant. However, all other fit indexes supported the plausibility of the model fit. Thus, hypothesis 1 about group-invariant factor pattern was supported. The following analysis tested the equality constraints related to the factor variances/covariances, assuming equal factor pattern. Results from model 2 revealed that the factor variances/covariances was the same for both groups. The chi-square difference between the two models was not significant \((\Delta \chi^2(4) = 2.08, p > .05)\). Further, all fit indexes yield similar values to model 1. Thus, hypothesis 2 assuming equal factor

![Table 3 Model Cross-Validation Results of the TGMD](https://i.imgur.com/3Q4Q.png)

**Note.** \(M_v\) = application of the model with the same form derived from the calibration group to the validation group. \(M_1\) = model with same factor loadings for the calibration and validation group. \(M_2\) = model with same factor variances/covariances for the calibration and validation group given equality of factor loadings. \(M_3\) = model with same error variances for the calibration and validation group given equality of factor loadings and factor variances/covariances. GFI = goodness of fit index. AGFI = adjusted goodness of fit index. RMSEA = root mean square of error approximation. \(^*\) denotes significant values.
Construct Validity of the TGMD

Loadings and factor variance/covariance was tenable. Finally, model 3, in which the invariance testing criteria were the most stringent, was examined. Results revealed an increment in chi-square value, which, however, was not significant compared to the previous model ($\Delta \chi^2 = 18.84, p > .05$). In addition, there was a loss in the values of measures of goodness of fit, which was considered trivial. Taken together, the above results seemed to provide strong support for the construct validity of the proposed two-factor structure of the TGMD.

The final solution based on model 3 is depicted in Figure 1. All factor loadings were statistically significant ($p < .05$) yielding moderate positive values. The

Figure 1 — Hypothesized two-factor model of TGMD. Unidirectional arrows represent direct causal influence, bidirectional arrows represent noncausal influence, and er1 to er12 represents observed variables’ errors.
mean loading for the locomotor factor was .50 while for the object control factor was .57. The lowest loading was observed in the run (.27) and the highest in the bounce (.81). In addition, the correlation between the two factors yield a high positive value (.82).

**Discriminant Validity**

Given the relatively high association between the two factors of the TGMD, it seemed appropriate to further examine their discriminant validity. To address this issue, the procedure proposed by Anderson and Gerbing (1988) was followed. The estimated correlation parameter between the locomotor and object control factors was constrained to unity. Then, a chi-square difference test was conducted on the values obtained for the two models (unconstrained and constrained). Results indicated that the chi-square difference test was statistically significant ($\Delta \chi^2(1) = 282.81$, $p < .001$). The model with the unconstrained correlation parameter (see model 3 in Table 3) yielded significantly better fit indexes than the constrained model (GFI = .79, AGFI = .70, RMSEA = .142). The above results suggest that the examined discriminant validity of the locomotor and object control constructs of the TGMD is tenable.

**Discussion**

The TGMD has been widely used in assessing children’s motor development. However, only one study examined its underlying structure. The research reported in the present study was designed to further investigate the factorial validity of the TGMD. A two-factor model was initially hypothesized and supported. Cross-validation procedures were employed in order to further examine the integrity of the postulated model. Results demonstrated that factor loadings, factor variances/covariances and error variances/covariances are invariant across the calibration and validation groups. Thus, the TGMD showed good cross-generalizability. Practitioners may be more confident that they measure two distinct constructs, namely, locomotor skills and object control skills, when using the TGMD in clinical or educational practice.

The relatively high association between the two latent constructs suggested that a single factor model might be plausible. However, chi-square differences as well as fit indexes for the one-factor and two-factor model suggested that the latter was the closest approximation of the hypothesized TGMD structure to the empirical data. Moreover, discriminant analysis clearly indicated that the locomotor and object control factors are not perfectly correlated and represent distinct constructs. This high correlation between the two factors may explain the clouded picture that emerged in the factor analysis conducted by Ulrich (1985), based on the national norm data.

Generally, item loadings yield acceptable values. The only item that had a low loading was the run (.27), which was statistically significant. This low value was not surprising. Previous research has consistently shown that the run item has the lowest reliability of the 12 fundamental motor skill items (Ulrich & Wise, 1984; Ulrich, Ulrich, & Branta, 1988). Further, in the study by Ulrich (1985), the run item formed a distinct factor. Clearly, additional research is needed to examine the factors that are responsible for the low reproducibility on the run item.
One hypothesis postulated was that the measurement error associated with each observed variable would be uncorrelated. However, the two-factor model was significantly improved by adding an error covariance between the observed variables of strike and throw. Cross validation analysis seemed to provide encouraging support for this modification. Although it is early for final conclusions, three possible explanations can be offered. First, several authors agree that these two skills have biomechanical similarities in pattern (Broer, 1973; Cooper, Adrian, & Glassow, 1982; Langendorfer, 1987). Indeed, both skills are fundamental manipulative skills; they are characterized by propulsive movement, in which an object is moved away from the body, and have similar rotation schema (Langendorfer, 1987). Furthermore, both skills combine two or more movements and are generally used in conjunction with other forms of movement (i.e., stepping, turning, swinging, stretching). Finally, both skills are acquired at about the same approximate age (Gallahue & Ozmun, 1998). Second, many performance criteria of the TGMD are similar in both object control skills (two-hand strike and overhand throw). These are hip rotation, weight transfer by stepping forward, and nondominant side of the body faces the tosser or the target. Therefore, correlated variance errors between the two items is not surprising. A third interpretation might be related to cultural differences. Traditional games and sports like softball and baseball are not widely practiced in Greece. Greek children might use the throw movement pattern to execute the striking skill. Therefore, the modification should be cautiously interpreted, since it may reflect idiosyncrasies of the sample used.

Although the present study provided important results regarding the factorial validity of the TGMD, the findings should be further replicated. The sample used in the present study consisted of children without disability. It would be valuable to extend the examination of the TGMD factorial validity for children with mental, sensory-motor, and other disabilities.

Physical educators must constantly assess the child’s motor development and plan experiences and instructional strategies that will help the child establish mature patterns of fundamental motor skills. The TGMD was developed in response to these requirements, and today it is considered one of the few standardized instruments that can be used to assess the effectiveness of instructional programming in gross motor development. The results of the present study support the construct validity of the TGMD proposed by Ulrich (1985) encouraging practitioners to use it as a valid assessment tool in physical education.

References


---

**Acknowledgments**

The authors would like to thank Dr. D. Ulrich, University of Michigan, for his valuable suggestions and comments on an earlier draft of this paper.