Reliability of the test of gross motor development second edition (TGMD-2) for Kindergarten children in Myanmar

Thanda Aye, RPT, MS, Khin Saw Oo, MD, PhD, Myo Thuzar Khin, MD, PhD, Tsugumi Kuramoto-Ahuja, RPT, MA, Hitoshi Maruyama, RPT, PhD

Department of Physical Therapy, Graduate School of Health and Welfare Sciences, International University of Health and Welfare: 2600-1 Kitakanemaru, Ota city, Tochigi 324-8501, Japan
Department of Physiotherapy, University of Medical Technology, Yangon, Ministry of Health and Sports, Republic of the Union of Myanmar
Department of Physical Medicine and Rehabilitation, University of Medicine 2, Yangon, Ministry of Health and Sports, Republic of the Union of Myanmar

Abstract. [Purpose] The purpose of this study was to investigate reliability of the test of gross motor development second edition (TGMD-2) for Kindergarten children in Myanmar. [Subjects and Methods] Fifty healthy Kindergarten children (23 males, 27 females) whose parents/guardians had given written consent were participated. The subjects were explained and demonstrated all 12 gross motor skills of TGMD-2 before the assessment. Each subject individually performed two trials for each gross motor skill and the performance was video recorded. Three raters separately watched the video recordings and rated for inter-rater reliability. The second assessment was done one month later with 25 out of 50 subjects for test-rest reliability. The video recordings of 12 subjects were randomly selected from the first 50 recordings for intra-rater reliability six weeks after the first assessment. The agreement on the locomotor and object control raw scores and the gross motor quotient (GMQ) were calculated. [Results] The findings of all the reliability coefficients for the locomotor and object control raw scores and the GMQ were interpreted as good and excellent reliability. [Conclusion] The results represented that TGMD-2 is a highly reliable and appropriate assessment tool for assessing gross motor skill development of Kindergarten children in Myanmar.

KEYWORDS: Reliability, TGMD-2, Kindergarten children

INTRODUCTION

Development of gross motor skills in young children is important for further developmental functions such as perceptual and cognitive abilities1, 2. Early assessment for development of gross motor skills during preschool and elementary school years is particularly important to monitor changes of motor development, identify delays or deficits of development, and to assist Physical Therapist (PT) and other health care practitioners to properly design exercise programs3–5.

There are several assessment tools to measure gross motor skill development of children. The assessment tools can be standardized (set protocol) or non-standardized (therapist selects relevant items) types6. The standardized assessment tools have prescribed guidelines for administration which must be followed to ensure reliability6, 7. The standardized assessment tools have two major types although some may have been designed to capture both elements. Those are norm-referenced type and criterion-referenced type7. Several standardized assessment tools for gross motor skill development of children are being used in clinical, educational and research settings3.
Test of gross motor development second edition (TGMD-2) is one of the most widely used assessment tools to assess gross motor skill development of children\(^2,8–10\). TGMD-2 is a standardized norm- and criterion-referenced test that measures gross motor abilities in children from 3-10 years of age\(^5,6\). The test is used to identify children who are significantly behind their peers in gross motor skill development, to plan an instructional program in gross motor skill development, to assess individual progress in gross motor skill development, to evaluate the success of the gross motor program, and to serve as a measurement instrument in research involving gross motor development. TGMD-2 is a highly reliable and valid assessment tool using normative sample of 1,208 persons residing in 10 states of the United State of America (USA) and three sources of error variance, content sampling, time sampling, and inter-scorer differences, were analyzed in relation to TGMD-2 subtest and quotient scores\(^3\).

There are several evidences for reliability of TGMD-2 as cross-cultural studies in Australia, Belgium, Brazil, Chile, China, Iran, Philippine, South Korea, and many other countries for typically developing children and children with special needs\(^3,4,10–16\).

Although there are evidences of reliability of TGMD-2 all over the world, it should be considered the sociocultural differences in children in Myanmar. Although several assessment tools are being used in clinical and school settings in Myanmar, the documented evidence of reliable assessment tools for gross motor skill development is limited. The government of the Republic of the Union of Myanmar has implemented a new National Education Strategic Plan (NESP) during the period 2016–2021 including a key reform focus on the provision of quality, healthy, play-centered pre-school and primary education for all children aged 3 to 6 years\(^7\). The quality Kindergarten (KG) program is one of the strategies of NESP and which has been implemented in all public schools, private schools, monastic schools and other types of schools across Myanmar by the Ministry of Education (MOE) since June 2016\(^8\). The national education law of the Republic of the Union of Myanmar (2014) states that KG is education that promotes holistic development using appropriate methods for five year-olds to ease their transition to first grade and will be regarded as the base level of primary education\(^9\). This study was a preliminary study of assessing gross motor skill development of KG children in Myanmar. Thus, the purpose of this study was to investigate reliability of the TGMD-2 for KG children in Myanmar.

**SUBJECTS AND METHODS**

This study was conducted with 50 healthy KG children (23 males, 27 females) who were attending at one public school for 2016–2017 academic year in Yangon city area of Myanmar. The characteristics of the subjects were age: 5.4 ± 0.3 years, height: 106.1 ± 7.0 cm, weight: 17.2 ± 2.5 kg, BMI: 15.3 ± 1.7 kg/m\(^2\). The exclusion criteria were children with known developmental disability (e.g. Cerebral palsy, Down’s syndrome, Autism Spectrum Disorder, Attention Deficit Hyperactive Disorder), obvious deformity (e.g. scoliosis, bow leg), and orthopedic injury in both upper and lower extremities within six months. Information on this study was provided to the principals, the teachers, the parents or guardians and the children themselves before their voluntary participation. They were provided with written information letter and verbal explanation. Prior to the participation, the parent or guardian of a child had to sign a written informed consent form to allow the child as a subject and the child him or herself was asked verbally to get his/her informed assent. They were also explained that they had the right to refuse participation and withdraw this study at any time. This study was approved by the Ethics Review Committee (ERC) of the University of Medical Technology, Yangon, Myanmar (approval number: 3/2016) and endorsed by the ERC of the department of Medical Research, Ministry of Health and Sports (MOHS), Myanmar. The official permission letter for the study areas was obtained from the department of basic education of the MOE, Myanmar.

The main equipment used was TGMD-2 (examiner record forms). Other materials used were two 8- to 10-inch playground balls, one basketball, one soccer ball, one 4-inch light weight ball, one tennis ball, one softball, one 4- to 5-inch square beanbag, color tapes, two traffic cones, one plastic bat, one batting tee, and a video camera (Sony HD, HDR-PJ410). TGMD-2 consists of 12 gross motor skills which are divided into locomotor and object control subtests. Locomotor subtest includes run, gallop, hop, leap, horizontal jump, and slide. There are four performance criteria for each skill of run, gallop, horizontal jump, and slide, while hop has five and leap has three criteria, and total 24 criteria for locomotor subtest. Object Control subtest includes striking a stationary ball, stationary dribble, catch, kick, overhand throw, and underhand roll. There are five performance criteria for striking a stationary ball, three for catch, while all the rest four items have four for each, and also total 24 for object control subtest. Each skill is performed two trials for assessment and each criterion is given a score of 1 or 0 for the pass and fail attempt respectively. The scores of two trials are added up to get total criterion score, the total criterion scores for the performance criteria are added up to get skill score, the six-skill scores are added up to get subtest raw scores (0–48). The subtest raw scores are converted into standard scores (1–20) and percentiles (<=1–>99) depending on age and gender according to the normed tables in TGMD-2 manual. The standard scores of locomotor and object control subtests are added up and converted into gross motor quotient (GMQ) (46–160). Finally, seven descriptive ratings: very poor, poor, below average, average, above average, superior, and very superior are given for the subtest standard scores and the GMQ for evaluation\(^11\).

Three raters were responsible for this study. Rater A is a PT, Rater B and C are physiatrists, and all the three raters work in medical universities and university affiliated hospitals under the MOHS of Myanmar. They all have more than 15 years of experience teaching, observing and evaluating children’s gross motor development. The Rater A and B have been introduced TGMD-2 two years earlier and more practice time than the Rater C. None of the three raters had any prior experience in
administration of TGMD-2 but they all have already finished training for TGMD-2.

The test venue and equipment were set up according to TGMD-2 requirements in the assembly hall of the school. The researcher thoroughly explained and demonstrated correct performance of all 12 gross motor skills of TGMD-2 before the assessment. After that, each child started to perform each gross motor skill under the supervision of the researcher and the KG class teachers. The child was allowed at least one test trial for each gross motor skill. The child had to perform two trials for each of all 12 gross motor skills (rest period was provided between two consecutive gross motor skill tests). The performance of every child was video-recorded. The video camera was fixed in the proper position and angle to record the whole performance of each motor skill, except recording of run, gallop and slide when the angle of the video camera was changed to record the whole performance. Total duration of the assessment of all 12 gross motor skills for each child lasted about 10–15 minutes (including rest periods). All children were assessed with their barefooted performance of all the skills. The assessment procedures were done according to the standardized guidelines of TGMD-2 and finished within three consecutive days for all 50 children. The video recordings were assessed and rated separately by three raters for the inter-rater reliability. For the test-retest reliability, out of 50 children already assessed a month before, 25 were randomly selected to be asked to perform all 12 motor skills for the second occasion. The performance of each child in the second occasion was also video recorded and assessed. The assessment for the second occasion was finished within two consecutive days. For the intra-rater reliability, after six weeks of the first assessment, the Rater A watched the same video recordings of 12 out of 50 children once and assessed again.

The agreement on raw scores for both locomotor and object control skills were calculated. The agreement on GMQ was also calculated. The inter-rater reliability was calculated by the use of Cronbach’s alpha, intra-class correlation coefficients (ICC) (two-way random, average measures, absolute agreement), Pearson correlation coefficients and Spearman correlation coefficients. The Spearman correlation coefficients were calculated only for the agreement on seven categories of descriptive ratings for overall skills converted from the GMQ by all the raters. The test-retest reliability and intra-rater reliability were also calculated by the use of Cronbach’s alpha, ICC (two-way mixed, average measures, consistency) and Pearson correlation coefficients. The software used for data analysis was IBM SPSS statistic version 22.0 for Windows.

RESULTS

The results of mean values for the locomotor and object control raw scores and the GMQ by three raters, Cronbach’s alpha, and ICC for the inter-rater reliability are shown in Table 1. All the values of Cronbach’s alpha and ICC showed excellent for all the measures in the inter-rater reliability. Table 2 shows the values of Pearson correlation coefficients for the inter-rater agreement. The results revealed strong positive to very strong positive degrees of correlation among all the three raters. The Spearman correlation coefficients for inter-rater agreement on seven categories of descriptive ratings for overall skill ranks converted from the GMQ are shown in Table 3. The results also indicated strong positive to very strong positive degrees of correlation among all the three raters. Table 4 and 5 show the results of mean values, Cronbach’s alpha, ICC and Pearson correlation coefficients for the test-rest reliability and the intra-rater reliability respectively. The results of the test-rest reliability presented acceptable to good alpha values for Cronbach’s alpha, good to excellent agreement values for ICC and strong positive to very strong positive degrees of correlation for Pearson correlation coefficients between Day 1 and Day 2 assessments. The results of the intra-rater reliability revealed excellent values for all the measures between the first and second ratings.

DISCUSSION

The findings of all the reliability coefficients for the locomotor and object control subtests and the GMQ are interpreted as good and excellent[20,21]. Portney & Watkin reported that the reliability coefficients ICC below 0.50 represent poor reliability, from 0.50 to 0.75 suggest moderate reliability, and values above 0.75 indicate high reliability[20]. Cicchetti indicated that the reliability coefficients ICC below 0.40 represent poor reliability, coefficients from 0.40 to 0.59 suggest fair reliability, from 0.60 to 0.74 represent good reliability and values above 0.75 indicate excellent reliability[21]. The inter-rater reliability ICC for the locomotor subtest, the object control subtest, and the GMQ were higher than 0.80. This value can be interpreted as excellent agreement among all the raters. Kim et al. reported that inter-rater reliability ICC for the locomotor raw scores was 0.94 and the sum of standard scores was 0.96 in 40 South Korean children who were assessed by three raters[14]. The test-retest reliability ICC for all the gross motor tests in this study were more than 0.75 which indicated high reliability. Farrokhi et al. reported that ICC for the locomotor subtest was 0.65, the object control subtest was 0.85 and the GMQ was 0.81 in the test-retest reliability testing of 63 children in Iran[13]. The test-retest ICC for the locomotor subtest of this study was higher but the object control subtest and the GMQ were lower than their study. It would be attributed to time interval differences between the first and second assessments and age differences of participants between two studies. The time interval between the test and retest in Farrokhi et al. was two weeks and the subjects were 3–10 years of age[13]. The intra-rater ICC for all the gross motor tests in this study were higher than 0.95 which represented high or excellent reliability. The similar findings for the intra-rater ICC were reported by Farrokhi et al. in Iran[13]. The subjects in the present study and the previous studies by Farrokhi et al. and
Kim et al. were typically developing children\textsuperscript{13, 14}. The similar findings were also found when the results of this study were compared to the study by Houwen et al. for children with visual impairments in the Netherlands\textsuperscript{15}. They had found that the inter-rater ICC for the locomotor subtest, the object control subtest and the GMQ were 0.82, 0.93 and 0.89 respectively, the intra-rater ICC for the locomotor subtest was 0.85, the object control subtest was 0.96 and the GMQ was 0.95, and the test-retest ICC for the locomotor subtest, the object control subtest and the GMQ were 0.86, 0.87 and 0.92 respectively. The time interval between the test and retest in Houwen et al. was also two weeks and the subjects were 6–12 years of age\textsuperscript{15}.

DeVellis reported that the alpha coefficients from 0.70 to 0.80 suggest acceptable, from 0.80 to 0.90 indicate good and above 0.90 represent excellent\textsuperscript{22}. The alpha coefficient values of the inter-rater reliability for the locomotor raw scores, the object control raw scores, and the GMQ were higher than 0.90. The values were not differ from the original TGMD-2 which showed 0.98 of alpha coefficient values for the locomotor raw scores, the object control raw scores and the GMQ in a set of

### Table 1. Inter-rater reliability (Cronbach’s alpha and ICC)

| TGMD-2 Scores          | Rater A (Mean ± SD) | Rater B (Mean ± SD) | Rater C (Mean ± SD) | Cronbach’s alpha | ICC   |
|------------------------|---------------------|---------------------|---------------------|------------------|-------|
| Locomotor raw scores   | 34.5 ± 7.95         | 35.8 ± 7.37         | 38.3 ± 6.12         | 0.98             | 0.95***|
| Object control raw scores | 26.9 ± 6.60       | 28.6 ± 5.71         | 32.7 ± 5.32         | 0.96             | 0.88***|
| Gross motor quotient   | 98.9 ± 13.5         | 101.9 ± 11.9        | 110.3 ± 11.1        | 0.96             | 0.89***|

Mean ± SD, ICC: Intra-class correlation coefficient, ***p<0.001

### Table 2. Inter-rater reliability (Pearson correlation coefficients)

| TGMD-2 scores          | Pearson Correlation Coefficients |
|------------------------|----------------------------------|
|                        | Rater A × Rater B | Rater A × Rater C | Rater B × Rater C |
| Locomotor raw scores   | 0.97***             | 0.94***             | 0.93***             |
| Object control raw scores | 0.96***             | 0.85***             | 0.96***             |
| Gross motor quotient   | 0.97***             | 0.87***             | 0.85***             |

***p<0.001

### Table 3. Inter-rater reliability (Spearman correlation coefficients)

| TGMD-2 scores          | Spearman Correlation Coefficients |
|------------------------|----------------------------------|
|                        | Rater A × Rater B | Rater A × Rater C | Rater B × Rater C |
| Descriptive rating of overall skills | 0.89*** | 0.67*** | 0.72*** |

***p<0.001

### Table 4. Test-retest reliability (Cronbach’s alpha, ICC and Pearson correlation coefficients)

| TGMD-2 scores          | Day 1 (Mean ± SD) | Day 2 (Mean ± SD) | Cronbach’s alpha | ICC   | Pearson correlation coefficients |
|------------------------|-------------------|-------------------|------------------|-------|----------------------------------|
| Locomotor raw scores   | 36.4 ± 7.49       | 42.1 ± 6.84       | 0.82             | 0.82***| 0.69*** |
| Object control raw scores | 26.2 ± 6.15       | 32.5 ± 7.12       | 0.79             | 0.79***| 0.67*** |
| Gross motor quotient   | 102.3 ± 13.2      | 118.9 ± 16.2      | 0.76             | 0.76***| 0.62*** |

Mean ± SD, ICC: Intra-class correlation coefficient, ***p<0.001, **p<0.01

### Table 5. Intra-rater reliability (Cronbach’s alpha, ICC and Pearson correlation coefficients)

| TGMD-2 scores          | Assessment 1 (Mean ± SD) | Assessment 2 (Mean ± SD) | Cronbach’s alpha | ICC   | Pearson correlation coefficients |
|------------------------|--------------------------|--------------------------|------------------|-------|----------------------------------|
| Locomotor raw scores   | 36.9 ± 8.72              | 38.3 ± 7.76              | 0.98             | 0.98***| 0.96*** |
| Object control raw scores | 27.8 ± 6.62              | 28.8 ± 7.36              | 0.95             | 0.95***| 0.91*** |
| Gross motor quotient   | 105.3 ± 17.1             | 108.0 ± 17.2             | 0.97             | 0.97***| 0.94*** |

Mean ± SD, ICC: Intra-class correlation coefficient, ***p<0.001
positive and more than 0.80 indicate very strong positive correlations. Valentini also informed that the alpha coefficient values of inter-rater reliability for the locomotor raw scores and the object control raw scores were 0.88 and 0.89 respectively in typically developing Brazilian children. Simons et al. reported that the internal consistency Cronbach’s alpha for the locomotor scale was 0.82, the object control scale was 0.86 and the GMQ was 0.90 in 7–10 year old Flemish children with intellectual disability.

Chowdhuru et al. reported that Pearson and Spearman correlation coefficients values from 0.60 to 0.79 represent strong positive and more than 0.80 indicate very strong positive correlations. The Pearson and Spearman correlation coefficients for the inter-rater reliability in this study showed higher than 0.75. Most of correlation coefficients between the Rater A and B is higher when compared to the correlation coefficients between the Rater A and C as well as between the Rater B and C. It might be ascribed to the different duration of familiarity with TGMD-2 among three raters. The Pearson test-retest reliability results showed more than 0.60 for the locomotor raw scores, the object control raw scores and the GMQ. The original TGMD-2 manual stated that the Pearson test-retest reliability results for the locomotor raw scores, the object control raw scores and the GMQ were 0.88, 0.93 and 0.96 respectively. Valentini also indicated that the Pearson test-retest reliability results for the locomotor raw scores was 0.83 and the object control raw scores was 0.91. The values in this study were lower than that of Ulrich and Valentini. It might also be attributed to time interval differences between the first and second assessments of three studies and cultural differences of Myanmar, USA and Brazil. The time interval between the test and retest was four weeks in this study which was not similar to the previous studies. The time interval of most of the previous studies was from 7–10 days to two weeks. The subjects in this study were one month older at the second assessment and their performances on all gross motor skills were better than the first assessment. TGMD-2 was first developed in the US and the skill for “Striking a stationary ball” is considered an important motor skill for American children. It is quite different for Myanmar children that the motor skills not only “Striking a stationary ball” but also some of the object control skills are not widely used in games and daily play activities. It was their first experience to play with a bat and batting tee for the majority of subjects in this study.

The limitations of this study were: 1) the subjects were only KG students, which meant only one age group, and 2) the subjects were only from Yangon city area. Further studies are still needed to find out the reliability of TGMD-2 using larger sample of Myanmar children, with age range from 3–10 years, from different regions and states.

In conclusion, the results in this study revealed high or excellent reliability of the tests. Thus, TGMD-2 is a highly reliable and appropriate assessment tool for assessing gross motor skill development of KG children in Myanmar.

Conflict of interest
No conflict of interest was declared.

REFERENCES

1) Pick JP, Dawson L, Smith LM, et al.: The role of early fine and gross motor development on later motor and cognitive ability. Hum Mov Sci, 2008, 27: 668–681. [Medline] [CrossRef]
2) Veldman SL, Jones RA, Okely AD: Efficacy of gross motor skill interventions in young children: an updated systematic review. BMJ Open Sport Excerc Med, 2016, 2: e000067. [Medline] [CrossRef]
3) Simons J, Daly D, Theodorou F, et al.: Validity and reliability of the TGMD-2 in 7–10 year-old Flemish children with intellectual disability. Adapt Phys Activ Q, 2008, 25: 71–82. [Medline] [CrossRef]
4) Valentini NC: Validity and reliability of the TGMD-2 for Brazilian children. J Mot Behav, 2012, 44: 275–280. [Medline] [CrossRef]
5) Ulrich DA: Test of gross motor development: examiner’s manual, 2nd ed. Austin: Pro-Ed publisher, 2000, pp 1–60.
6) Human Kinetics. Test Instruments Used in Adapted Physical Education. http://www.humankinetics.com/excerpts/excerpts/test-instruments-used-in-adapted-physical-education (Accessed Jul. 3, 2017).
7) Connolly BH: Examination and evaluation: tests and administration. In: Therapeutic exercises in developmental disabilities, 3rd ed. SLACK Incorporated, 2005, pp 21–79.
8) De Mediros P, Zequino MA, Fronza FC, et al.: Motor assessment instruments and psychometric procedures: a systematic review. Motricidade, 2016, 12: 64–75 http://revistas.rcaap.pt/motricidade/article/view/6397 (Accessed Jun. 20, 2017)
9) Cools W, Martelaer KD, Samaey C, et al.: Movement skill assessment of typically developing preschool children: a review of seven movement skill assessment tools. J Sports Sci Med, 2009, 8: 154–168. [Medline]
10) Kim CI, Han DW, Park II: Reliability and validity of the test of gross motor development-II in Korean preschool children: applying AHP. Res Dev Disabil, 2014, 35: 800–807. [Medline] [CrossRef]
11) Barnett LM, Minto C, Lander N, et al.: Interrater reliability assessment using the Test of Gross Motor Development-2. J Sci Med Sport, 2014, 17: 667–670. [Medline] [CrossRef]
12) Cane-Cappellacci M, Leyton FA, Carreho JD: Content validity and reliability of test of gross motor development in Chilean children. Rev Saude Publica, 2015, 49: 97. [Medline] [CrossRef]
13) Farrokhi A, Zareh ZM, Karimi AL, et al.: Reliability and validity of test of gross motor development-2 (Ulrich, 2000) among 3–10 aged children of Tehran city. J Physic Educ Sport Manag, 2014, 5: 18–28.
14) Kim S, Kim MJ, Valentini NC, et al.: Validity and reliability of the TGMD-2 for South Korean children. J Mot Behav, 2014, 46: 351–356. [Medline] [CrossRef]
15) Houwen S, Hartman E, Jonker L, et al.: Reliability and validity of the TGMD-2 in primary-school-age children with visual impairments. Adapt Phys Activ Q, 2010, 27: 143–159. [Medline] [CrossRef]
16) Lopes VP, Saravia L, Rodrigues LP: Reliability and construct validity of the test of gross motor development-2 in Portuguese children. International Journal of Sport and Exercise Psychology, 2016. https://www.researchgate.net/publication/307896008_Reliability_and_construct_validity_of_the_test_of_gross_motor_development-2_in_Portuguese_children (Accessed Apr. 12, 2017)
17) Aung San Suu Kyi: Foreword. In: National education strategic plan 2016–21. The government of the Republic of the Union of Myanmar, Ministry of Education, 2016, p 5. http://www.moe.gov.mm (Accessed Apr. 10, 2017)
18) National education strategic plan 2016–21. The government of the Republic of the Union of Myanmar, Ministry of Education, 2016, pp 70–83. http://www.moe.gov.mm (Accessed Apr. 10, 2017)
19) National Education Law: Myanmar, 2014 (41/2014). http://www.ilo.org/dyn/natlex/natlex4.detail?p_lang=en&p_isn=100493&p_count=3&p_classification=09 (Accessed Apr. 10, 2017)
20) Portney LG, Watkins MP: Foundation of clinical research: application to practice, 3rd ed. Philadelphia: F. A. Davis, 2015, pp 77–96.
21) Cicchetti DV: Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess, 1994, 6: 284–290. [CrossRef]
22) DeVellis RF: Scale development: Theory and applications. Los Angeles: Sage, 2012, pp 109–110.
23) Chowdhury AK, Debsarkar A, Chakrabarty S: Novel methods for assessing Urban air quality: combined air and noise pollution approach. J Atmos Pollut, 2015, 3: 1–8.