

Inter-rater Reliability of the K-GMFM-88 and the GMPM for Children with Cerebral Palsy

Jooyeon Ko, P.T., Ph.D., Minyoung Kim, M.D.

Department of Rehabilitation Medicine, CHA Bundang Medical Center, CHA University, Seongnam 463-712, Korea

Objective To examine inter-rater reliability of the Korean version Gross Motor Function Measure (K-GMFM-88) and the Gross Motor Performance Measure (GMPM) based on the video clips.

Method We considered a sample of 39 children (28 boys and 11 girls; the mean age=3.50±1.23 years) with cerebral palsy (CP). Two pediatric physical therapists assessed the children based on video recordings.

Results For the K-GMFM-88, the intraclass correlation coefficient ($ICC_{3,1}$) ranged from .978 to .995, and Spearman's correlation coefficient ranged from .916 to .997. For the GMPM, $ICC_{3,1}$ ranged from .863 to .929, and Spearman's correlation coefficient ranged from .812 to .885. With the gross motor function classification system classified according to the functional level (GMFCS I-II vs. III-V), the ICCs were .982 and .994 for the K-GMFM-88 total score and .815 and .913 for the GMPM total score. There were good or high correlations between the subscales of the two measures ($r=.762-.884$).

Conclusion The K-GMFM-88 and GMPM are reliable tools for assessing the motor function of children with CP. These two methods are highly correlated, which adds more reliability on them. Thus, it is advisable to use K-GMFM-88 and GMPM for children with CP to assess gross motor function.

Key Words Cerebral palsy, Gross motor function, K-GMFM-88, GMPM, Reliability

INTRODUCTION

Cerebral palsy (CP) is a group of disorders affecting the development of movements and postures and causing activity limitations that are attributable to non-

progressive disturbances in the developing fetal or infant brain.¹ Among the functional domains of impairments in children with CP, the gross motor function matters the most for their general activities in life. The gross motor function in children with CP has been conceptualized as having two main features: function and performance.² The "function" means ability to accomplish certain motor activity and it does not necessarily implicate quality of motor control. On the other hand, the "performance" refers to the quality of motor activity or how well the child performs a certain activity. For example, when a child can stand independently for 10 seconds, this description designates the gross motor "function" of the child, whereas the "performance" description refers to the degree of stability during his or her standing.³

Received May 16, 2011; Accepted December 6, 2011

Corresponding author: Minyoung Kim

Department of Rehabilitation Medicine, CHA Bundang Medical Center, CHA University, 351, Yatap-dong, Bundang-gu, Seongnam 463-712, Korea

Tel: +82-31-780-6281, Fax: +82-31-780-6206, E-mail: kmin@cha.ac.kr

© This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2012 by Korean Academy of Rehabilitation Medicine

Previous studies attempting to reveal the effectiveness of some treatment modalities for children with CP have been limited because of a lack of valid and reliable assessment tools capable of quantifying functional changes following interventions.⁴ To address this deficiency, the Gross Motor function Measures Group of researchers and therapists in Ontario have developed two assessment instruments to measure subtle but meaningful changes in the motor function and performance of children with CP. As a result, they proposed the gross motor function measure (GMFM) and the gross motor performance measure (GMPM), which were designed to be used together.⁵ The reliability of the GMFM and the GMPM has been documented in the West.⁶⁻⁸ Russell et al.⁴ reported that the inter-rater reliability of the GMFM ranged from .87 to .99 across five dimensions and .99 for the total score. Thomas et al.⁹ found that the inter-observer reliability of the GMPM ranged from .78 to .86 for the five attributes.

To the authors' knowledge, no study has determined the reliability of the GMFM and the GMPM for Korean children with CP. When an observational assessment tool is used to measure clinical outcomes, it is important to establish the reliability of that tool.¹⁰ Several types of reliability tests are necessary for determining the stability, consistency, and dependability of scores for a specific instrument, and particularly inter-rater, intra-rater, and test-retest reliabilities are basic ones. Although all types of reliability are important, this study focused on determining the inter-rater reliability of the K-GMFM-88 and the GMPM.

MATERIALS AND METHODS

We employed a sample of 39 children with CP (28 boys and 11 girls). These children were admitted to the CHA Bundang Medical Center in Korea for intensive rehabilitation (Table 1). Inclusion criteria were children who were diagnosed as CP by medical doctors specializing in pediatric rehabilitation, and the age range was 2 to 7 years, while exclusion criteria were the children who received orthopedic surgery in last six months because their motor performance was not likely to be indicative of their typical motor ability. All the children and their mothers provided their written consent for their participation in this study, and we followed the principles in the Declaration of Helsinki.

Table 1. Summary Statistics for Subjects (N=39)

Variables	
Mean age (years)	3.50±1.23
Gender	
Boys	28 (71.8)
Girls	11 (28.2)
GMFCS	
I-II	12 (30.8)
III-V	27 (69.2)
Origin of motor dysfunction	
Intraventricular hemorrhage	1 (2.6)
Prematurity	25 (64.1)
Vascular anomaly	1 (2.6)
Polymicrogyria	1 (2.6)
Middle cerebral artery (MCA) infarct	1 (2.6)
Postnatal asphyxia	1 (2.6)
Infantile seizure	1 (2.6)
Schizencephaly	1 (2.6)
Kernicterus	1 (2.6)
Unknown	6 (15.4)
MRI findings	
Periventricular leukomalacia (PVL)	19 (48.7)
Ventricular dilator	1 (2.6)
Abnormal diffusion tensor imaging (DTI)	4 (10.2)
Hydrocephalus	3 (7.7)
Normal	3 (7.7)
Others	9 (23.1)

Values are mean±SD or n (%)

GMFCS: Gross motor function classification system

With permission of the copyright owner, we translated the "Administration and Scoring Guidelines for the GMFM-88 and the GMFM-66" in the GMFM User's manual.⁴ The translation procedure followed the forward-backward-forward method. The GMPM comprises 20 items of the GMFM, and thus, we translated it based on the GMFM-88, which is a criterion-referenced observational measure for assessing the gross motor function in children with CP. The GMFM-88 consists of 88 items grouped into five domains: (A) lying & rolling, (B) sitting, (C) crawling/kneeling, (D) standing, and (E) walking/running/jumping. Each item is scored on a four-point Likert-type scale (0-1-2-3). The higher the score, the better the gross motor function is. We converted the raw score for each domain into the percentage of

the maximum per domain. Each domain was equally weighted, and we calculated the total score by summing the percentages of each domain and dividing the result by five. The total score for the GMFM-88 was based on the percentages for the five domains and was obtained when the subject finished this measure.⁴

The GMPM is composed of 20 items selected from the GMFM through a consensus method for assessing the quality of movements in children with CP. The GMPM uses a subset of GMFM items for the following five domains: lying/rolling, crawling/kneeling, sitting, standing, and walking/running/jumping. Three of the 20 items are static (e.g., standing), whereas the remaining 17 are dynamic (e.g., hopping on one foot). For each GMPM item, three out of the five possible attributes (alignment, coordination, dissociated movement, stability, and weight shift) are determined to be assessed. Alignment refers to the adjustment of parts or segments of the body in relation to each other. Coordination is defined as the smooth and controlled use of movements in motor performance and takes into account the timing, velocity, direction, force, and amplitude of movements. Dissociated movements refer to isolated movements (e.g., the extension of the hip with the flexion of the knee). Stability refers to the active maintenance of a body position in the presence of disturbing forces. Finally, weight shift is defined as movement involving the transfer of the body's center of gravity. We assessed each attribute by using a five-point Likert-type scale ranging from "severely abnormal" (1) to "consistently normal" (5) and calculated the percent scores for the attributes and the total score (scale 0-100%). We scored all three attributes for each item simultaneously, based on the average performance in the three trials.¹¹

We administered the assessments based on the K-GMFM-88 and the GMPM in a pediatric physical therapy room that was comfortable and familiar to the subjects, and the procedure was videotaped by two pediatric physical therapists. We used video recordings because instant scoring in detail needs long time from frequent break, and it enables more accurate scoring. All children were assessed barefoot, without assistive devices. It took about 40 minutes and 20 minutes to administer the test to record the subject's gross motor function and performance respectively. To assess inter-rater reliability, two pediatric physical therapists (rater

A and B) who were not involved in the video-recording served as raters for the K-GMFM-88 and the GMPM. Both of them had more than six years of experience in the evaluation and treatment of children with CP. The two raters attended a one-week GMFM-88 and GMPM training workshop on the administration and scoring of the K-GMFM-88 and GMPM, which included ten hours of manual education and ten hours of videotapescore sessions. The training session was guided by a senior pediatric physical therapist who was fully proficient in the assessment tools. The two raters viewed video recordings of 39 subjects independently and scored them by using K-GMFM-88 and the GMPM over a oneweek period without discussing the results.

We calculated the means and standard deviations (SDs) for each test. The K-GMFM-88 and the GMPM are ordinal measures presenting a number of response options that "order" the characteristics of interest from better to less skilled performance, and thus, we employed non-parametric statistics.⁴ We used intraclass correlation coefficients ($ICC_{3,1}$) with 95% confidence intervals and Spearman's correlation coefficients to evaluate the inter-rater reliability of domain/attribute scores and total scores for the K-GMFM-88 and the GMPM. In addition, we employed the Wilcoxon signed-rank test for differences between the raters. Further, we determined ICCs through functional classification. Finally, we used Spearman's correlation coefficients to assess the relationship between the K-GMFM-88 and the GMPM. We conducted all the statistical analyses by using the SPSS software package for Window (ver. 12.01). A p-value of less than 0.05 was considered significant.

RESULTS

Table 1 provides the characteristics of the 39 subjects. In terms of the five domains and total scores for K-GMFM-88, ICCs ranged from .978 to .995, and Spearman's correlation coefficients ranged from .916 to .997. In terms of the five attributes and total scores for the GMPM, ICCs ranged from .863 to .929, and Spearman's correlation coefficients ranged from .812 to .885 (Table 2). There was no statistical difference between the two raters for domain and total score of the K-GMFM. For the GMPM, all attributes except for "alignment" showed no difference between raters A and B (Table 3). When

Table 2. Inter-rater Reliability for the K-GMFM-88 and the GMPM

Attribute/Domain	ICC _{3,1}	95% CI	Spearman's rho
K-GMFM-88			
Lying/rolling	.978	.958-.988	.916*
Sitting	.987	.987-.996	.980*
Crawling/kneeling	.988	.988-.997	.983*
Standing	.984	.969-.992	.980*
Walking/running/jumping	.992	.984-.996	.997*
Total score	.995	.991-.998	.995*
GMPM			
Alignment	.905	.819-.905	.872*
Coordination	.907	.823-.951	.818*
Stability	.892	.793-.943	.829*
Weight shift	.923	.854-.960	.812*
Dissociated movement	.863	.739-.938	.847*
Total score	.929	.864-.963	.885*

ICC: Intraclass correlation coefficient, CI: Confidence interval, K-GMFM-88: Korean version Gross motor function measure-88, GMPM: Gross motor performance measure

*p<0.01

stratified by the functional levels GMFCS I and II, ICCs ranged from .898 to .982 and from .757 to .830 for the K-GMFM-88 and the GMPM, respectively. For GMFCS III, IV, and V, ICCs ranged from .974 to .997 and from .808 to .913 for the K-GMFM-88 and the GMPM, respectively. Spearman's correlation coefficients, in the GMFCS I, II, was ranged from .875 to .991 and from .649 to .758 for the K-GMFM and GMPM respectively. Spearman's correlation coefficient of the GMFCS III-V was .925-.998 and .715-.838 individually (Table 4). Table 5 described the correlation between the K-GMFM and the GMPM. All the subscale scores of the measures showed good or high correlation positively (r=.762-.884).

DISCUSSION

We evaluated the inter-rater reliability of the K-GMFM-88 and the GMPM. To the authors' knowledge, this study is the first to consider a population of Korean children with CP to determine the reliability of the K-GMFM-88 and the GMPM simultaneously by using the same subjects and video recordings. The results indicates good reliability of K-GMFM-88 and the GMPM.

Although there are several statistical methods for assessing inter-rater reliability, we employed the ICC and

Table 3. Wilcoxon Signed-rank Test for the K-GMFM-88 and the GMPM

Attribute/Domain	Rater A Mean±SD	Rater B Mean±SD	Z	p
K-GMFM-88				
Lying/rolling	75.77±28.66	76.13±28.42	-.962	.336
Sitting	55.85±34.14	55.97±34.00	-.405	.686
Crawling/kneeling	44.36±38.09	44.38±38.11	.000	1.000
Standing	24.90±33.80	24.82±33.72	-.816	.414
Walking/running/jumping	17.54±28.44	17.23±27.35	-.736	.461
Total score	43.64±29.32	43.69±29.12	-.157	.876
GMPM				
Alignment	38.46±20.65	45.98±22.44	-3.225	0.001*
Coordination	37.59±19.05	39.01±22.95	-.844	.399
Stability	40.51±16.59	42.51±20.98	-.876	.381
Weight shift	38.72±19.71	38.07±22.41	-.350	.727
Dissociated movement	38.72±17.74	40.00±28.60	-.683	.494
Total scores	38.80±17.61	41.09±22.73	-1.089	.276

Values are mean±SD

K-GMFM-88: Korean version Gross motor function measure-88, GMPM: Gross motor performance measure

*p<0.05

Table 4. Inter-rater Reliability for the K-GMFM-88 and the GMPM Classified by Functional Level

Attribute/Domain	GMFCS I, II (n=12)			GMFCS III-V (n=27)		
	ICC _{3,1}	95% CI	Spearman's rho	ICC _{3,1}	95% CI	Spearman's rho
K-GMFM-88						
Lying/rolling	.959	.866-.988	.875*	.974	.944-.988	.925*
Sitting	.898	.684-.969	.929*	.984	.964-.993	.972*
Crawling/kneeling	.962	.872-.989	.883*	.985	.968-.993	.984*
Standing	.930	.775-.979	.915*	.997	.992-.998	.948*
Walking/running/jumping	.982	.939-.995	.991*	.992	.983-.996	.998*
Total score	.982	.938-.995	.986*	.994	.987-.999	.989*
GMPM						
Alignment	.813	.351-.941	.758*	.844	.657-.929	.746*
Coordination	.757	.156-.930	.661*	.897	.774-.953	.760*
Stability	.779	.232-.936	.695*	.822	.609-.919	.736*
Weight shift	.830	.411-.951	.685*	.901	.783-.955	.715*
Dissociated movement	.757	.155-.930	.649*	.808	.578-.912	.744*
Total score	.815	.357-.947	.715*	.913	.809-.960	.838*

Values are mean±SD

K-GMFM-88: Korean version Gross motor function measure-88, GMPM: Gross motor performance measure, GMFCS: Gross motor function classification system

*p<0.01

Table 5. Relationship between the K-GMFM-88 and the GMPM

	GMPM					
	Alignment	Coordination	Stability	Weight shift	Dissociated movement	Total score
K-GMFM-88						
Lying/rolling	.809*	.767*	.762*	.784*	.775*	.797*
Sitting	.866*	.800*	.796*	.808*	.855*	.838*
Crawling/kneeling	.884*	.866*	.847*	.834*	.864*	.870*
Standing	.844*	.779*	.786*	.792*	.853*	.818*
Walking/running/jumping	.835*	.779*	.786*	.789*	.849*	.819*
Total score	.878*	.835*	.819*	.819*	.868*	.859*

Values are Spearman's rho

K-GMFM-88: Korean version Gross motor function measure-88, GMPM: Gross motor performance measure

*p<0.01

Spearman's correlation coefficient to provide an in-depth analysis of the inter-rater reliability of the K-GMFM-88 and the GMPM. We used the ICC to assess the degree of correspondence and agreement between the rates,¹² and Spearman's correlation coefficient to measure the correlation. Spearman's correlation coefficient provides information on the level of the correlation as well as the direction of the correlation. Watkins and Portney¹² reported that an ICC≥.90 indicates a high reliability, .75-

.90 indicates good reliability, .50-.75 indicates moderate reliability, and ≤.50 indicates poor reliability. Meyer¹³ reported that a correlation coefficient $r \geq .8$ indicates a high correlation, $r = .6-.8$ indicates a good correlation, $r = .4-.6$ indicates a moderate correlation, and $r \leq .4$ indicates a poor correlation. In this study, the ICCs and Spearman's correlation coefficient for the K-GMFM-88 were high for all the domains and total scores. It was reported that the reliability of the GMFM-88 from 317

children with CP aged 1 and 15 years was very good.¹⁴ The total ICC was reported as .96 (.69-.98)⁵ and .78-.86 by attributes of the GMPM.⁹ In the present study, the ICCs and Spearman's correlation coefficients for total scores and attributes for the GMPM were from good to high. We also assessed the difference in each score between the two raters, and found some difference in the "alignment" attribute of the GMPM through the Wilcoxon signed-rank test, although the ICCs were high. Assessing the "alignment" involves the observation of more than one body segment,⁵ and requires a decision from the multiple features at a time, which might bring greater chance of error or discordance. Thus, raters should be familiar with the manual and scoring practices to minimize the inter-rater inconsistencies.

In the present study, inter-rater reliability did not vary according to the functional level of the subject for the five domains and total score of the K-GMFM-88. However, the total scores for the GMPM were lower for GMFCS I and II than for GMFCS III, IV, and V. Sorsdahl et al.¹⁵ examined the inter-reliability of the GMPM by using video recordings of the children with CP and reported that the ICC for the total score of the GMPM was lower in the independent ambulators who often move at high speed, which is consistent with the present study results. We attempted to slow the subjects down, but the observation was limited by their innate movement patterns.

The reliability of a test refers its ability to provide consistent results. A number of sources of variations may influence the reliability of results obtained from a measure. These sources include problems with the test itself, such as unclear administration guidelines or imprecise scoring systems. A lack of training can lead to variations in raters. In this study, we provided the raters with intensive training to minimize such variations. In addition, structured video recordings and video scoring offered many advantages. In this study, we allowed the raters to watch video recordings several times and to stop the recordings to review the scoring guidelines in the manuals. Thus, video observations may be better when the scoring is performed by trained and skilled raters.¹⁵

The correlation between the K-GMFM-88 and the GMPM ranged from .762 to .884, which suggested a considerable overlap between their constructs. This indicates that to examine the gross motor function of children with CP,

we need to consider the level as well as performance quality of their gross motor function.¹⁶ This study's main limitation was that we evaluated only the ICC. A high ICC indicates high relative reliability, however it does not necessarily imply high absolute reliability. Relative reliability refers the consistent ranking of scores for an individual in a group by repeated measurements, and small measurement errors are needed for absolute reliability. Thus, a more appropriate way to investigate the reliability of an instrument intended for use in a clinical setting may be to examine absolute reliability. In addition, high relative and absolute reliability may require intensive efforts under practical guidelines. Further research to examine the inter-rater reliability of K-GMFM and GMPM should be continued by pediatric therapists and clinicians with a wide range of clinical experience across multi-center.

CONCLUSION

The inter-rater reliability of K-GMFM-88 and the GMPM was highly satisfactory in terms of total scores and sub-scores indicating that they are reliable methods for assessing the gross motor functional ability as well as the quality of movement in children with CP.

REFERENCES

1. Rosenbaum P, Paneth N, Leviton A, Goldstein M, Bax M, Damiano D, Dan B, Jacobsson B. A report: the definition and classification of cerebral palsy April 2006. *Dev Med Child Neurol Suppl* 2007; 109: 8-14
2. Palisano RJ, Gracely EJ, Rosenbaum PL. Gross motor capability and performance of mobility in children with cerebral palsy: a comparison across home, school, and outdoor/community settings. *Phys Ther* 2004; 84: 419-429
3. Boyce WF, Gowland C, Rosenbaum PL, Lane M, Plews N, Goldsmith CH, Russell DJ, Wright V, Potter S, Harding D. The Gross motor performance measure: validity and responsiveness of a measure of quality of movement. *Phys Ther* 1995; 75: 603-613
4. Russell DJ, Rosenbaum PL, Avery LM, Lane M. Gross motor function measure (GMFM-66 & GMFM-88) user's manual, 2nd ed, Hamilton: Gross Motor Measure Groups, 2002, 3-4

5. Gowland C, Boyce WF, Wright V, Russell DJ, Goldsmith CH, Rosenbaum PL. Reliability of the Gross Motor Performance Measure. *Phys Ther* 1995; 75: 597-602
6. Bower E, McLellan DL, Arney J, Campbell MJ. A randomized controlled trial of different intensities of physiotherapy and different goal-setting procedures in 44 children with cerebral palsy. *Dev Med Child Neurol* 1996; 38: 226-237
7. Guyatt GH, Water S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987; 40: 171-178
8. Mclaughlin JF, Bjornson KF, Astley SJ, Hays RM, Hoffinger SA, Armantrout EA, Roberts TS. The role of selective dorsal rhizotomy in cerebral palsy: critical evaluation of a prospective clinical series. *Dev Med Child Neurol* 1994; 36: 755-769
9. Thomas SS, Buckon CE, Phillips DS, Aiona MD, Sussman MD. Interobserver reliability of the gross motor performance measure: preliminary results. *Dev Med Child Neurol* 2001; 43: 97-102
10. Harris SR, Haley SM, Tana WL, Swanson MW. Reliability of observational measures of the Movement Assessment of Infants. *Phys Ther* 1984; 64: 471-477
11. Boyce W, Gowland C, Rosenbaum P, Hardy S, Lane M, Plews N, Goldsmith C, Russell D, Wright V, Potter S, et al. Gross motor performance measure manual, Kingston: Queen's University, 1998, 6
12. Watkins MP, Portney LG. Foundations of clinical research: applications to practice, 1st ed, East Norwalk: Conn Appleton and Lange, 1993, 53-67
13. Meyer CR. Measurement in physical education, 1st ed, New York: Ronald Press Co, 1979
14. Beckung E, Carlsson G, Carlsdotter S, Uvebrant P. The natural history of gross motor development in children with cerebral palsy aged 1 to 15 years. *Dev Med Child Neurol* 2007; 49: 751-756
15. Sorsdahl AB, Moe-Nilssen R, Strand LI. Observer reliability of the Gross Motor Performance Measure and Quality of Upper Extremity Skills Test, based on video recordings. *Dev Med Child Neurol* 2008; 50: 146-151
16. Palisano RJ, Hanna SE, Rosenbaum PL, Russell DJ, Walter SD, Wood EP, Raina PS, Galuppi BE. Validation of a model of gross motor function for children with cerebral palsy. *Phys Ther* 2000; 80: 974-985