

Inter-Rater Reliability of the Gross Motor Performance Measure

Yi Chung-hwi, Ph.D., P.T.

Dept. of Physical Therapy, College of Health Science, Yonsei University
Institute of Health Science, Yonsei University

Park So-yeon, M.Sc., P.T.

Dept. of Rehabilitation Therapy, The Graduate School, Yonsei University

Ko Myung-suk, B.H.Sc., P.T.

Seoul Community Rehabilitation Center

국문요약

대동작 운동 수행능력 측정 도구의 측정자간 신뢰도

이충휘

연세대학교 보건과학대학 물리치료학과 및 보건과학연구소

박소연

연세대학교 대학원 재활학과

고명숙

서울 장애인종합복지관

대동작 운동 수행능력 측정도구(GMPM)는 뇌성마비 아동의 움직임을 질적인 면에서 평가하기 위해 개발된 도구이다. 이 연구의 목적은 대동작 운동 수행능력 측정도구의 측정자간 신뢰도를 알아보는 것이다. 뇌성마비 아동 10명(평균 5.6세, 범위 4~8세)에게 GMPM 평가를 실시하였다. 평가 과정을 비디오로 녹화하여 각 속성 항목별로 3명의 평가자간의 급간내 상관계수로 일치도를 보았다. 전반적으로 측정자간 신뢰도는 '불량~보통' 범주에 속했다. 이 연구의 결과는 충분한 교육을 받지 않고 평가하면 그 결과를 신뢰하기 어렵다는 것을 말해준다. 향후 임상에서 GMPM을 이용하여 평가할 때 측정자간 신뢰도에 어떤 변화가 있는지 알아보는 연구가 필요하다.

핵심단어: 대동작 운동 수행능력 측정도구; 신뢰도.

This research was supported by grants from Yonsei Maji Institute, 2002.

Corresponding author: Yi Chung-hwi yeh0406@dragon.yonsei.ac.kr

Introduction

Researchers at McMaster University, Queen's University, and the Hugh MacMillan Rehabilitation Center, Ontario, initiated the development and validation of two assessment instruments designed to measure the subtle, but meaningful, changes over time in motor function and motor performance found in children with cerebral palsy. The Gross Motor Function Measure (GMFM) and the Gross Motor Performance Measure (GMPM) were designed to be used together (Gowland et al, 1995). The distinction between the measures is that the GMFM measures 'how much' a child can do, whereas the GMPM measures 'how well' a child can perform a subset of the motor tasks (Gowland et al, 1995).

The GMPM consists of 20 items selected from the GMFM to assess the quality of movement in children with cerebral palsy (Boyce et al, 1991;1992; 1993;1995;1998). The GMPM uses a subset of four GMFM items from 5 dimensions: rolling, crawling/kneeling, sitting, standing, and walking/running/jumping. Three of the 20 items are static, such as standing, while the remaining 17 items are dynamic, such as hopping on one foot. For each GMPM item, three of five possible attributes are assessed: alignment, coordination, dissociated movement, stability, and weight shift. Each attribute is assessed using a five point scale with a score of 1 representing 'severely abnor-

mal' and 5 representing 'consistently normal.' All three attributes for each item are scored simultaneously and are based on the average performance of three trials (Boyce et al, 1998).

When an observational assessment tool is being used as a measure of clinical outcome, it is important to establish the reliability of that tool. Several types of reliability testing are necessary in order to determine the stability, consistency, and dependability of the scores for a particular instrument, particularly inter-rater, intra-rater, and test-retest reliability (Portney and Watkins, 2000). Although all types of reliability are important, the purpose of the present study was to determine the inter- and intra-rater reliability of the GMPM in Korea.

Methods

Subjects

A total of 10 children with cerebral palsy participated in the study from September 2003 to November 2003. The sample group comprised 4 children diagnosed with spastic diplegia, 3 with spastic hemiplegia, and 1 with spastic quadriplegia, 1 with athetoid, and 1 with ataxia. The mean age was 5.6 years, ranging from 4 to 8 years.

Procedures

A 1-day GMPM training program was developed. The workshops commenced with a description of the research back-

Table 1. Subject characteristics

(N=10)

	Type	Frequency
Cerebral palsy type	Spastic diplegia	4
	Spastic hemiplegia	3
	Spastic quadriplegia	1
	Ataxia	1
	Athetoid	1
Sex	Male	3
	Female	7
Age (yr)	5.6±1.35	

ground and properties of the GMPM. An overview of general concepts in administering and scoring the test was followed by group discussion on the scoring issues of each GMPM item using videotaped examples. The workshop focused on the specific definitions and criteria used to evaluate each GMFM item using the GMPM attributes.

Three inexperienced physical therapy students were the evaluators for the study. All three evaluators participated in a one-day training workshop on the administration of the GMPM taught by a physical therapist with 6 years of clinical experience in pediatric physical therapy.

Before commencing the test, evaluators were given a GMPM manual and instructed to use the administration and scoring guidelines when scoring the test videotape. Prior to being shown the GMPM item on videotape, the item number and the number of trials they would see the child attempt for that item were identified. The tape was stopped between

items to allow participants time to score and prepare for the next item. No items were replayed. All children were assessed barefoot, without assistive devices. No discussion took place regarding scoring during the assessment.

Reliability testing

Interrater reliability was assessed by comparing the simultaneous independent assessment from the three evaluators using the GMPM. All items were performed three times and scored according to the definitions provided in the manual (Boyce et al, 1998).

Data analysis

The intraclass correlation coefficient [ICC (2,1)] was used to assess the degree of correspondence and agreement among ratings (Portney and Wakins, 2000). Overall agreement score of ICCs were determined by dimension and attribute. For this study, ICCs below .75 were considered 'poor to moderate', those above .75

were considered 'good', and above .90 'excellent' (Portney and Wakins, 2000).

Results

The reliability indexes are shown in Table 2. There was great variability in the performance for judgements of attribute, with ICC(2,1) values ranging from .13 to .53. ICCs revealed that weight shift was in the 'poor' category while the attributes of dissociated movement, coordination, alignment, and stability were found to be in the 'moderate' category (Table 2).

Table 2. Reliability measures by attribute for the total group (N=10)

Attribute	ICC(2,1)
Dissociated movement	.52
Coordination	.46
Alignment	.53
Weight shift	.13
Stability	.38
Total	.44

Discussion

Knowledge of inter-rater reliability and agreement provides the clinician with insight into the standardization of the test, and the raters with the ability to use the criteria correctly in scoring behaviors (Harris et al, 1984; Tinsley and Weiss, 1975). An important consideration in all reliability studies is the need to sample

the range of performance across the range of items (Russell et al, 1994). Previous studies have used either a descriptive or correlative statistic. These statistics are not appropriate for reliability testing of an ordinal scale. The intraclass correlation coefficient statistic is preferred because it measures agreement while accounting for chance agreement (Portney and Wakins, 2000).

Primary purveyors of measures usually spend a great deal of time developing and validating a new instrument, and collecting normative data. Generally, a much smaller amount of effort is directed toward issues of training. Although clinicians have a responsibility to acquire the necessary training before using a new measure, it is often not clear what the necessary training is, or how to acquire it in a systematic and effective manner. The time and cost associated with setting up a training package have likely been deterrents to its development (Russell et al, 1994).

There are a number of disadvantages and advantages to the use of videotapes as a medium for training and evaluating new users of a test such as the GMPM. One of the main disadvantages of using criterion videotapes to assess reliability is that this method is only testing the participant's ability to score the videotaped test reliably and is not providing an indication of the assessor's ability to administer and score the test in a clinical situation (Russell et al, 1994).

Another problem with using videotapes is the quality of videotaping, in particular, the ability to capture on videotape, from the best possible camera angle, the movement the therapist is trying to test. Experience has shown it may be more difficult to judge whether a child is "initiating" a movement from videotape or from real life.

There are, however, a number of advantages to using videotapes as a method of assessing reliability. First, it is possible to evaluate the effects of an intervention (such as a training workshop) in a standardized manner. Second, the use of videotapes allows an efficient means of assessing several patients of varying diagnostic and functional levels while eliminating the issue of patient compliance. This advantage is particularly appealing when dealing with children. Videotapes can be edited to ensure they are capturing different training issues and covering an appropriate spectrum of function. Third, by having a criterion testing videotape with the "correct" score, as determined by experts, the therapist can ensure that response are not only reliable, but valid.

When the therapists then assess inter-rater reliability, it may be high because everyone agrees on how to score, but the score is not the correct (valid) one. Finally, another use for criterion testing videotapes is to have an easy method of assessing ongoing levels of competency. Tests can be completed at regular intervals to ensure that high levels of reli-

ability are maintained over time. Gross (1991) and Gross and Conrad (1991) offer further discussion of the advantages and disadvantages of using videotape to capture observational data.

The results from this pilot study indicate that it is questionable to assess the reliability of the quality of movement in children with CP using a videotape. Further research is needed to determine whether this reliability is improved in a clinical situation.

Conclusion

Our purpose was to examine the inter-rater reliability of inexperienced physical therapy students' scores using the GMPM from videotaped performances. The results from this study indicate that it is questionable to assess the reliability of the quality of movement in children with CP using GMPM. Further work is needed to determine whether the reliability of GMPM is improved in a clinical situation.

References

- Boyce W, Gowland C, Hardy S, et al. Development of a quality-of-movement measure for children with cerebral palsy. *Phys Ther.* 1991;71:820-832.
- Boyce W, Gowland C, Rosenbaum P, et al. *Gross Motor Performance Measure Manual.* Kingston, Queen's University, 1998.

- Boyce W, Gowland C, Rosenbaum P, et al. The Gross Motor Performance Measure: Validity and responsiveness of a measure of quality of movement. *Phys Ther.* 1995;75:603-613.
- Boyce W, Gowland C, Rosenbaum P, et al. Gross Motor Performance Measure for children with cerebral palsy: Study design and preliminary findings. *Can J Public Health.* 1992;suppl2: S34-40.
- Boyce W, Gowland C, Russell D, et al. Consensus methodology in the development and content validation of a Gross Motor Performance Measure. *Physiotherapy Canada.* 1993;45:94-100.
- Gowland C, Boyce W, Wright V, et al. Reliability of the Gross Motor Performance Measure. *Phys Ther.* 1995;75:597-602.
- Gross D. Issues related to validity of videotaped observational data. *West J Nurs Res.* 1991;13:658-663.
- Gross D, Conrad B. Issues related to reliability of videotaped observational data. *West J Nurs Res.* 1991;13:798-803.
- Harris SR, Haley SM, Tada WL, et al. Reliability of observational measures of the movement assessment of infants. *Phys Ther.* 1984;64:471-476.
- Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice.* 2nd ed. New Jersey, Prentice Hall Health, 2000.
- Russell DJ, Rosenbaum PL, Lane M, et al. Training users in the Gross Motor Function Measure: Methodological and practical issues. *Phys Ther.* 1994; 74(7):630-636.
- Tinsley HEA, Weiss DJ. Interrater reliability and agreement of subjective judgements. *J of Counselling Psychology.* 1975;22:358-376.