A pilot study on the test re-test and the inter-rater reliability of the Melbourne Assessment of Unilateral Upper Limb Function

**Jayaraman P[1];
Puckree T[2]**

[1] Prithi Jayaraman completed a Master of Physiotherapy in the Department of Physiotherapy, University of KwaZulu Natal.
[2] Threethambal Puckree is associate Professor in the Department of Physiotherapy, University of KwaZulu Natal.

ABSTRACT:

Objective: The Melbourne Assessment of Unilateral Upper Limb Function (commonly referred to as the Melbourne Assessment) was identified as a tool to quantify the quality of upper extremity function in children with cerebral palsy aged 5 to 15 years in South Africa. Since the tool was not tested in a South African population before, it became necessary to determine its inter-rater and test-retest reliability.

Methods: Five South African Black children with hemiplegic cerebral palsy served as the test sample. The raters were 2 neurodevelopmental trained physiotherapists with more than 2 and 8 years of experience in pediatric physiotherapy but novice to the use of the Melbourne Assessment. Both therapists acquainted themselves with the tool kit and manual prior to the rating. The entire assessment of each child was videotaped and reassessed a week later by one of the therapists for test-retest reliability.

Results: Ratings of the 2 raters and test-retest scores were correlated using the weighted Kappa due to the small sample size. Kappa scores for individual scores for interrater reliability and test-retest was 0, 75 and that for the total scores were 0, 72 and 0, 82 respectively.

Conclusion: These findings suggest that good inter-tester and test-retest reliability can be achieved for the Melbourne Assessment when used in a group of South African Black children.

KEYWORDS: MELBOURNE ASSESSMENT, REACH, GRASP, FUNCTION,

INTRODUCTION

The Melbourne Assessment of Unilateral Upper Limb Function, abbreviated to the 'Melbourne Assessment' was developed by Johnson et al (1994). Johnson and his coworkers found that there were no reliable and valid measures that quantified the quality of upper extremity function in children with neurological impairment in the age group of 5 to 15 years. Therefore the Melbourne assessment was developed to fill the gap in quantifying outcome measures. The only other test that measures the quality of movement is the QUEST (Quality of Upper Extremity Skill Test). It had been developed for children between 18 months to 8 years. It consists of 36 items grouped in four domains, namely, dissociated movement, grasp, protective extension, and weight bearing. These components are more representative of the components of hand function development that occurs from birth to 18 months

(DeMatteo et al., 1992) and is therefore not suitable for children of school going age. Since the Melbourne assessment is the only tool that measures the quality of upper limb function in a cohort of children of school going age it was chosen for a study on quality of movement in children of school going age.

The Melbourne Assessment comprises of 16 criterion-referenced items that include reach, grasp, release, and manipulation. The child is evaluated when sitting at a table, or if unable to sit independently, sitting in their usual form of support (i.e. wheelchair) with an appropriate tray, or table. The entire assessment is administered using the standardized directions in the kit, and is videotaped for precise scoring at a later time. The Melbourne Assessment is scored on the child's performance as the task is attempted. Components of each test item are measured and make up the criteria for scoring, including range of move-

ment, target accuracy, fluency, grasp, accuracy of release, finger dexterity, and speed depending on the item. Video recording is used for precise observation as is required in research. The score sheet consists of 3-, 4-, or 5-point scales that allocate scores on the 16 items, with 37 subscores, according to success and quality of movement. The sum of the individual scales of the 37 subscores is recorded as a raw score and converted to a percentage score. For the complete test, the total possible score is 122 points. The test takes approximately 30

Correspondence to:
Prof. T. Puckree
Department of Physiotherapy
University of KwaZulu Natal
Private Bag X54001
Durban
4001
E-mail: puckreet ukzn.ac.za

minutes to administer and 30 minutes to score (Johnson et al 1994). It does not require specialized training to administer (Cusick et al 2005).

The initial reliability of the Melbourne Assessment was done by Johnson et al, (1994) on eleven children with cerebral palsy in Australia. It was found that the Melbourne assessment strongly related to the clinical judgment of experts. Upon administration of the assessment to 20 children the inter-rater reliability (0.68) was found to be substantial and intra-rater agreement after two weeks was 0.80.

The original 12 item assessment was reviewed and modified into a 16 item assessment with 37 sub-items. The reliability of the revised tool was tested by Randall et al, (2001) on 20 children in Australia. The results demonstrated that there was high internal consistency of test items (?=0.96), moderate to high agreement both within and between raters for all test items (intra-class correlations of at least 0.7) apart from item 16 (hand to mouth and down), and high inter-rater reliability (0.95) and intra-rater reliability (0.97) for total test scores. Test–retest results revealed moderate to high intra-rater reliability for item totals (mean of 0.83 and 0.79) for each rater and high reliability for test totals (0.98 and 0.97).

The construct validity and correlation of the Melbourne Assessment with the Pediatric Evaluation of Disability Inventory (PEDI) was investigated by Bourke-Taylor, (2003). The pediatric evaluation of disability inventory is a tool that measures functional status in children aged between six months and seven years. The results revealed very high correlation coefficients between the Melbourne Assessment and self-care (0.939) and mobility domains (0.783) of the PEDI and the overall functional skills section of the PEDI (0.718).

Content validity was established (Johnson et al 1994) by examining relevant literature, reviewing existing clinical upper-limb assessments, and by workshops with clinicians experienced in working with children who have cerebral palsy. In the absence of other reliable and valid test, concurrent-criterion validity was established by comparing the scores of the Melbourne Assessment with expert clinical judgment. Internal consistency of test items was determined and the results indicated that the items correlated significantly with each other and with the total score ((Johnson et al 1994).

The above studies indicate that the Melbourne Assessment is a valid and reliable tool when administered in a group of children in Australia but the ethnic background was not mentioned. Upon a thorough review of literature it was found that no such studies had been done in South Africa. Therefore a pilot study was conducted to investigate the test retest and inter rater reliability of the Melbourne Assessment in South African children.

METHODS
The study was conducted at the Tongaat School for the Severely Mentally Handicapped. Ethical clearance was obtained from the University of KwaZulu Natal Ethics committee following which informed consent was obtained from the guardian of the children.

SUBJECTS
Due to the need to contain variability between subjects, only subjects with hemiplegic cerebral palsy were identified. This criterion limited the number of subjects that were available to undertake the study. Four of the 5 participating children were diagnosed as hemiplegic cerebral palsy and one child had a hemiplegic distribution of symptoms as a result of hydrocephalus. All five children (3 males and 2 females, in the age group of 5 to 15; mean= 9.48) were Black African. Table 1 gives information about the demographics of the children as well as details about the severity of the condition as determined by the child's pediatrician and therapist.

PROCEDURE
The Melbourne Assessment was administered by the researcher following the standardized instruction in chapter 6 of the instruction manual (Randall et al., 1999). The children were assessed randomly (Test 1) and reassessed after a period of 5 hours (Re-test 1). Each child was made to sit on a chair appropriate to his/her size ensuring that the feet rested on the ground. The tools for the subtests were placed on a table to ensure easy access. This table was adjusted such that it was at the chest level of the child and just below the nipple line. The subtest tools were placed on the table at the marked position. The 'marked position' was the exact spot on the table where the test items were placed each time. This position was determined by marking on the table that point which was a comfortable forearm distance from the child's midline. As the assessments had to be videotaped, the camera (Panasonic VHS-C movie camera, model no. NV-RZ1EN/ENC) was mounted on top of a stand. The camera was positioned as per the guidelines in the instructional manual of the Melbourne Assessment (Randall et al., 1999).

Instructions were given to the child in English by the researcher. Also a standardized set of instructions were given in IsiZulu with the help of an assistant. Each child was allowed two test trials before each task was performed for monitoring. The child's performance was videotaped and scored later. The tapes that were used were the Panasonic HD extra 60 minute tape and the JVC 60 minute tape.

Scoring was done as per the instructions in the manual (Randall et al., 1999). For the inter-rater reliability a second rater from the University of KwaZulu Natal was recruited. Both the raters were novice users of the Melbourne Assessment which would not have much bearing on the results as the assessment protocol has been reported to be reliable even when used by novice users (Cusick et al., 2005). Both raters were Neurodevelopmental Therapy (NDT) trained, one with greater than 2 years experience and the other with greater than 8 years experience. Prior to scoring, the researcher and the second rater familiarized themselves with the contents of the manual. A brief discussion was also held to clarify doubts. Following this, each rater scored each child independently at the same time and a discussion was held to clear any ambiguities. Tapes of Test 1 were then scored. The researcher and the rater individually scored each child. For the

Table 1: Demographic profile of the children

Child	Age	Sex	Race	Diagnosis	Side Affected	Severity
1	10 yrs 11 mnths	Female	African	Spastic Hemiplegic	Right	Severe
2	9 yrs 9 mnths	Male	African	Spastic Hemiplegic	Right	Moderate
3	9 yrs 4 mnths	Female	African	Spastic Hemiplegic	Left	Mild
4	9 yrs 10 mnths	Male	African	Hydrocephalous (Spastic Hemiplegic)	Left	Moderate
5	10 yrs 7 mnths	Male	African	Spastic Hemiplegic	Right	Mild

Table 2: Melbourne Assessment: Total test score %

Child number	Inter rater scores		Test –retest scores	
	Rater 1	Rater 2	Test 1	Test 2
1	34	35	34	35
2	72	72	72	72
3	84	83	84	84
4	47	48	47	48
5	82	80	82	82

Table 3: Statistics for the inter rater and test retest scores for individual items and total scores for the Melbourne Assessment

Tests	Agreement	Expected agreement	Kappa	Standard error	Z scores	p
Individual items						
Inter rater	91,43	65,14	0,754	0,251	3,00	0,001
Test-retest	91,43	65,14	0,754	0,251	3,00	0,001
Total Scores						
Inter rater	90,00	64,00	0,722	0,250	2,89	0,002
Test retest	93,33	62,67	0,821	0,274	3,00	0,001

test-retest reliability the researcher scored the re-test 1 tapes two week after scoring the Test 1 tapes.

DATA ANALYSIS

The inter-rater and test-retest scores were statistically analyzed using the kappa statistics (weighted by the size of disagreement) to show the extent of agreement. For this sample the Kappa was considered to be the best statistical test to determine reliability due to the small sample size and the use of just 2 raters.

RESULTS

There were no dropouts of subjects from the study. All data was usable.

Inter-rater reliability and test-retest reliability

The total score is the percentage of the sum of the scores in all the sub-items. The scores for the two raters and test-retest scores for the one rater are given in Table 2. The test-re-test scores were the same in three children and there was only 1% difference in the other two children. The variation between the scores of rater 1 and 2 and test-retest scores are minimal namely 0 -2%.

The individual scores for each of the 16 items by the two raters are not shown. A statistical summary is provided in Table 3. There were minor variations between the two raters in one child for four items. In two children the scores

varied in 3 items. There was perfect agreement between the two raters for all the children for items 4, 7, 12, 13 and 16.

The individual item test-retest scores are not shown. In one child there was a difference in scores in 5 items. The test-retest scores were exactly the same for all the children in items 4, 5, 7, 8, 9, 11,12,13 and 14. The inter-rater and test re-test scores were the same in items 4, 5, 12 and 13.

For inter-rater reliability, the kappa value for the total score was 0.72 and the item score was 0.75 (Table 3) which indicates that there was a substantial agreement between the raters.

For test-retest reliability, the kappa

value for the total score is 0.82 which indicates an almost perfect agreement. The kappa value for the sub-items was 0.75 which indicates a substantial agreement.

DISCUSSION

The results of the study indicate that the Melbourne Assessment produced reliable outcomes on the study sample which was heterogeneous in terms of the degree and extent of severity of the condition. The modification that had to be done to administer the protocol to a South African population was that the instructions had to be delivered in IsiZulu which is the regional language. Nevertheless the results of this study confirm the work of Johnson et al., (1994) who also found that the inter-rater agreement was substantial. Randall et al, (2001) too found moderate to high agreement with respect to inter- and intra-rater reliability which is also consistent with the results of this study. This lends support to the fact that the protocol can be used for children originating from different ethnic and geographic backgrounds.

The kappa agreement was higher for the sub-item score than for the total test scores in inter-rater reliability. This indicates that individual items have been constructed in such a fashion that it does not lead to too much variations as might occur due to differences in personal interpretation during assessment.

Studies by Cusick et al., (2005) have shown that the tool is also reliable for novice users as well. However there have been no studies done that have correlated differences in the past experience in treating children with neurological problems to the reliability in scores.

While the manual states that the scorers have to have more than two years experience in the field of pediatric neurology, the question arises whether a large difference in the clinical experience really does have a bearing on the results and further studies are required to confirm this.

The fact that the test-retest reliability had a perfect agreement may lead to the argument that this could have been due to memory of the previous assessment being retained as the same rater scored both tests 1 and 2. This can not be ruled out but the fact that the reassessment was done after a fortnight belies this effect to a large extent. However further studies with multiple raters and a larger sample size is recommended.

The Melbourne Assessment had not been used in South Africa prior to this study. Limited literature on the tool suggests that it was developed and used in Australia. As such it has not gained popularity in the rest of the world as a tool that could be used to quantify upper extremity function like grasp, reach and manipulation in children with cerebral palsy. The researcher finds the Melbourne Assessment an 'easy to administer' tool, which does not take up much of the therapists time.

In conclusion, irrespective of the fact that this study had a small sample size and the inter- rater agreement was only tested using two raters, one can come to the conclusion that the Melbourne Assessment suggests that good inter-rater and test re-test reliability can be achieved with this tool. This indicates that this tool can be used to assess the quality of upper limb function in Black South African children with hemiplegia.

ACKNOWLEDGEMENTS

REFERENCES

Bourke-Taylor H 2003 Melbourne Assessment of Unilateral Upper Limb Function: construct validity and correlation with the Pediatric Evaluation of Disability Inventory. Dev Med Child Neurol. 45(2):92-6.

Cusick A, Vasquez M, Knowles L, Wallen M 2005 Effect of rater training on reliability of Melbourne Assessment of Unilateral Upper Limb Function scores. Developmental Medicine & Child Neurology 47: 39–45.

DeMatteo C, Law M, Russel D, Pollock N, Rosenbaum P 1992 QUEST: Quality of upper extremity Skills Test, Hamilton, ON: McMaster University, Neurodevelopmental Clinical Research Unit.

Johnson LM, Randall MJ, Reddihough DS, Oke LE, Byrt TA, Bach TM 1994 Development of a clinical assessment of quality of movement for unilateral upper-limb function. Dev Med Child Neurol. Nov; 36(11):965-73.

Randall M, Carlin JB, Chondros P, Reddihough D 2001 Reliability of the Melbourne Assessment of Unilateral Upper Limb Function. Developmental Medicine & Child Neurology 43: 761–767.

Randall M, Johnson L, Reddihough D 1999 The Melbourne Assessment Unilateral Upper Limb Function Test Administration Manual, Melbourne: Arena Printing, Royal children's hospital.