COMPARISON OF RESPONSIVENESS OF THE FUGL-MEYER ASSESSMENT AND THE WOLF MOTOR FUNCTION TEST

By

PEY-SHAN WEN

A THESIS PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF HEALTH SCIENCE

UNIVERSITY OF FLORIDA

2007

To my family and friends

ACKNOWLEDGMENTS

This material is the result of work supported with resources and the use of facilities at the North Florida/South Georgia Veterans Health System.

TABLE OF CONTENTS

LIST OF TABLES

Abstract of Thesis Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Master of Health Science

COMPARISON OF RESPONSIVENESS OF THE FUGL-MEYER ASSESSMENT AND THE WOLF MOTOR FUNCTION TEST

By

Pey-Shan Wen

December, 2007

Chair: Craig Velozo
Major: Occupational Therapy

To determine the effectiveness of a clinical intervention, an important property for outcome measurement is to capture change over time. Responsiveness, defined as the ability to detect change over time, is becoming a new standard for establishing the soundness of outcome instruments. The Fugl-Meyer Assessment (FMA) has been reported as a reliable and valid measurement for stroke population; however, little evidence has been reported to support its sensitivity in measuring change. The Wolf Motor Function Test (WMFT) has been commonly used in the constraint-induced movement therapy (CIMT) studies; however, no study reports its ability in detecting change. The purpose of the study is to compare responsiveness of the FMA and the WMFT.

The data were collected from three studies related to the CIMT. Eighty-seven participants were assessed by the FMA and the WMFT before and after intervention. For detecting responsiveness three methods were used: 1) the paired t-test; 2) the standard response mean; 3) the effect size. This study was approved by the Institutional Review Board of the University of Florida.

The paired t-test results for the WMFT and the FMA were 5.76 and 5.13 respectively. The standard response means for the WMFT and the FMA were 0.73 and 0.65 respectively. The effect sizes for the WMFT and the FMA were 0.17 and 0.32 respectively.

In conclusion, while the results are mixed, the WMFT had a tendency to show better responsiveness than the FMA, especially with the more sophisticated responsiveness designs. Future studies will be necessary to determine the replicability of these findings.

CHAPTER 1
INTRODUCTION

Stroke, which affects four out of five American families, is one of the leading causes of disability in the United States (www.stroke.org). Two-thirds of the four million stroke survivors continue to have moderate or severe disabilities at the end of treatment (www.stroke.org). Several studies demonstrate that only 5% to 20 % of the patients with stroke regain completely functional recovery (Heller et al., 1987; Moskowitz, Lightbody, & Freitag, 1972; Nakayama, Jorgensen, Raaschou, & Olsen, 1994; Parker, Wade, & Langton Hewer, 1986). Approximately half of the stroke survivors cannot functionally use their affected upper extremity (Broeks, Lankhorst, Rumping, & Prevo, 1999). In addition, 67% of participants with stroke claimed that loss of the arm function is their major problem even four years after their first stroke onset (Broeks et al., 1999).

The persistence of functional loss and the limit functional recovery have made selecting treatments for chronic stroke survivors extremely challenging. Historically, the recovery of function in stroke patients has been thought to take place only in the first six months following stroke onset (Langton Hewer, 1990) Furthermore, upper extremity recovery may be limited to the first four months after stroke onset (Broeks et al., 1999; Nakayama et al., 1994). Conventional interventions such as proprioceptive neuromuscular facilitation (PNF) and neurodevelopmental therapy (NDT) have been widely used in rehabilitation; however, limited evidence supports their efficacy (de Pedro-Cuesta, Widen-Holmqvist, & Bach-y-Rita, 1992; Duncan, 1997). In contrast to conventional interventions, novel treatments such as constraint-induced movement therapy (CIMT) and robot treatment have demonstrated the treatment effect for stroke population (Bonifer & Anderson, 2003; Kopp et al., 1999; Kunkel et al., 1999; Liepert et al., 2000; Liepert et al., 1998; Miltner, Bauder, Sommer, Dettmers, & Taub, 1999; Tarkka,

Pitkanen, & Sivenius, 2005; Taub et al., 1993; van der Lee, Beckerman, Lankhorst, & Bouter, 1999; van der Lee, Wagenaar et al., 1999; Wolf, Lecraw, Barton, & Jann, 1989). While more and more novel treatments are proposed, to determine the most effective treatment outcome measure has to be sensitive to change overtime.

A number of outcome measurements are available for monitoring upper extremity recovery in stroke. One of the most widely used instruments is the Fugl Meyer Assessment (FMA). The FMA, which is based on Brunstrom's six stages of recovery, is considered the gold standard for assessing stroke patients. The FMA consists of 33 upper-extremity items and 17 lower-extremity items. This instrument includes five domains: motor, sensory, balance, joint range of motion, and joint pain(Fugl-Meyer, Jaasko, & Norlin, 1975). Duncan (1983) reported the total motor score of the FMA has excellent intrarater and interrater reliability (0.98 and 0.98)(Duncan, Propst, & Nelson, 1983). Studies also suggest the FMA is a valid measurement for stroke (De Weerdt & Harrison, 1985; Wood-Dauphinee, Williams, & Shapiro, 1990). While the FMA shows good psychometric properties, little evidence supports its ability of measuring change over time. Gladstone and colleagues (2002) reported the FMA might fail to detect subtle changes in mild motor impaired stroke patients due to the ceiling effect (Gladstone, Danells, & Black, 2002). Moreover Wolf and colleagues (2001) stated that the FMA is not useful for evaluating full range of function in patients with mild to moderate impairment (Wolf et al., 2001). These studies imply that although the FMA has good psychometric properties, it might not be sensitive in monitoring clinical improvement.

In an attempt to more precisely measure the abilities of participants who enrolled in CIMT studies, Wolf and colleagues developed the Emory Motor Test, later called Wolf Motor Function Test (WMFT) (Wolf et al., 1989). The WMFT was specifically designed for assessing the upper

extremity ability of people receiving CIMT. The WMFT consists of three parts: (1) time score, recording how long the examinee completes a standardized task; (2) rating scale, assessing the quality of movement while performing the task; (3) strength. This assessment contains 17 tasks: 15 timed/rating tasks and two strength tasks. These tasks were sequenced according to complexity of upper extremity movement (from single joint to multiple joint movement, and from gross motor to fine motor movements).The WMFT shows good reliability of both the time score and rating scale, ranging from .86 to .97 (Morris, Uswatte, Crago, Cook, & Taub, 2001). In addition, the WMFT time score shows good construct and criterion validity (Wolf et al., 2001).

While reliability and validity are important foundational psychometric elements for measurement, reliability and validity provide little information of an instrument's ability to measure change over time(Guyatt, Walter, & Norman, 1987). Responsiveness, defined as the ability to detect subtle change over time, is becoming a new standard test to determine the soundness of an instrument (Kirshner & Guyatt, 1985). While the FMA and the WMFT are broadly used as outcome measures in stroke population, the comparison of their capabilities to detect the subtle change within the participants over time has not been investigated. Responsiveness may be an important means to further validate the FMA and the WMFT as outcome measures for determining treatment effect. This study compared responsiveness of the FMA and the WMFT using four different responsiveness indices.

CHAPTER 2
METHOD

Participants

The existing data were provided by three CIMT related studies. Study 1 (N=48): The participants only received the constraint induced movement therapy (CIMT-alone). Study 2 (N=19): The participants received transcranial magnetic stimulation (TMS) on their brain and the CIMT (CIMT-TMS). Study 3 (N=20): The participants received Donepezil 10mg/day and the CIMT (CIMT-Donepezil). Total 87 participants enrolled in these three studies. All Participants received CIMT 6 hours a day, 5 days a week for two weeks. The participants were assessed before interventions and immediately after 10 days CIMT training. These three studies were approved by the institution review board.

The inclusion and exclusion of these studies are described as follow: The inclusion criteria: (1) the diagnosis of at least one stroke and no more than 3 strokes on the same side of the brain; (2) ability to understand and follow instructions; (3) the ability to sit independently without back or arm support for 5 minutes; (4) the ability to stand with support of a straight cane, quad cane or hemiwalker for 2 minutes; (5) the ability to actively participate 6 hours of therapy without long rest or nap periods; (6) passive range of motion of all upper extremity motions of at least half the normal range. (7) onset at least 6 months prior to participation. The exclusion criteria: (1) Mini-Mental score less than 24; (2) health problems judged by the screen physician to put the client at significant risk of harm during the study; (3) other neurological conditions; (4) pain that is scored greater than 5 on the McGill Pain Scale.

Instruments

A series of outcome measurements were used in these CIMT related studies. The upper extremity portion of the FMA and the rating scale portion of the WMFT were used in the

responsiveness analyses. The FMA consists of 33 items measuring upper extremity movements rated by 3-point scales (0-2) with the maximum score of 66. Items of the upper extremity portion of the FMA were listed in Appendix A. The WMFT consists of 15 items measuring quality of upper extremity movement rated by 6 point scales (0-5) with the maximum score of 75. Items of the WMFT were listed in Appendix B.

The Stroke Impact Scale (SIS) was developed from patient and caregiver's perspective. The SIS consists of eight domains: mobility, ADL, memory, communication, emotion, participation, physical and hand function. It was rated by 5 point scales (1-5) with converting score from 0 to 100 (Duncan et al., 1999). Hand function domain contains five items. See Appendix C.

Although recent studies suggest that measuring responsiveness is important, there is no general consensus on the methods to assess responsiveness. Several different methods have been proposed to assess responsiveness. These methods can be generally divided into two categories: assessing change in a single population and assessing change in two populations (the population with change and the population without change) (Wallace, Duncan, & Lai, 2002). Since there is no consensus on which method is the best, methods designed for single population and two populations were be compared in this study.

Responsiveness Measure

Three indices were used for evaluating responsiveness in single-population category: the paired t-test, the standardized effect size (SES), and the standardized response mean (SRM).

The paired t-test is commonly used in test statistics for two-time-point measurements in a single population. The paired t-test compares the dependent sample mean of differences in terms of the number of standard errors. Comparing the values of statistic test across instruments over

the same time period, the largest test statistic represents the most responsive instrument. However, the paired t-test has been challenged due to its influence by sample size.

Effect size statistics, quantifying the change scores in terms of certain variations (Cohen, 1977), are independent on sample size, Because of its sample size independency, effect size statistics might be preferred statistics to evaluate responsiveness. Several effect size indices have been proposed. Among these effect size indices, the differences often are the formula of the denominators. The SES and SRM are two common forms of effect size statistics used in evaluating responsiveness. **The SES** is defined as the mean change score divided by the standard deviation of the baseline score (Kazis, Anderson, & Meenan, 1989). **The SRM** is defined as the mean change score divided by the standard deviation of the change (Liang, 1995).

Another effect size index**, the Guyatt effect size (GES),** was conducted for evaluating responsiveness in two-population category (Guyatt et al., 1987). The GES is defined as the minimal meaningful clinical Difference (MMCD) divided by the standard deviation of change in the stable group. In this study design, participants need to be divided into two populations (stable population and improved population) based on an external criterion and then the values of MMCD of instruments to be determined. The external criterion in this study is the SIS hand function score. The improved group was defined as individuals showing a change in the SIS hand function score greater or equal to two points and the stable group was defined as individuals showing a change of less than two points or a negative change of two points (Duncan et al., 1999). Individual's SIS hand function score decreased more than two points were excluded from this analysis. The methods of determining MMCD values are controversial through out the literature. In this study, two methods were used to obtain the MMCD for the FMA and the WMFT; when the amount of change in the FMA or WMFT reached: (1) 0.8 specificity compared

to the level of no change in SIS (Wallace et al., 2002) and (2) 10% of the total score of the FMA or WMFT (Gladstone et al., 2002; van der Lee, Wagenaar et al., 1999).

CHAPTER 3
RESULTS

The first responsiveness comparisons for single-population category were based on 62 out of 87 participants from combining dataset. Twenty-five participants were not analyzed because of the missing values on either FMA or WMFT. The characteristics of 62 participants included in responsiveness analysis based on single-population category presented in Table 3-1. Moreover, the responsiveness comparisons for two-population category (Guyatt effect size) were based on 57 participants. Five additional participants were excluded because of either missing scores on SIS hand function score (external criterion) or decreasing more than 2 points on their SIS hand function scores. Finally, since it is unusual to combine multiple treatment studies in a single analysis, the responsive comparisons for single-population category from the above analyses were also calculated based on only 37 participants from Study 1. The means and standard deviations of pretest and post test for stable and improved group presented in Table 3-2.

Using the combined dataset (N=62), in single population category, paired t-tests were significant; with the t values for the FMA and the WMFT were almost identical at 5.13 and 5.16, respectively. The SES analysis showed the FMA to be more responsive than the WMFT (0.32 and 0.17, respectively) while the SRM analysis showed the WMFT to be more responsive than the FMA (0.73 and 0.65, respectively). See Table 3.

Using the data from Study 1 (N=37) only, again, both the FMA and the WMFT showed significant t-tests, though the t-value for the WMFT was larger than that of the FMA (4.91 and 2.19, respectively). For Study 1, the SES values for the FMA and WMFT were almost identical (0.16 and 0.17, respectively), while the SRM values was greater for the WMFT than the FMA (0.36 and 0.81, respectively). See Table 3.

In two-population category, when MMCD values were established by the definition as the amount of change in the FMA or the WMFT reaching 0.8 specificity, the MMCD for the FMA and the WMFT were 7 and 3 respectively. The GES test for the WMFT was greater than that for the FMA (1.35 and 1.54, respectively). Again, when the MMCD values were established by the definition of MMCD as 10% of the total score of the FMA or the WMFT, the MMCD for the WMFT was greater than that of the FMA (6.6 and 7.5, respectively). Furthermore, the GES tests again showed the WMFT to be more responsive than the FMA (2.88 and 1.11, .88 respectively). See Table 4.

Table 3-1. The characteristics of 62 participants included in responsiveness analysis based on single-population category.

Demography Information (N=62)	
Mean Age ± SD	64.1±12.6
Mean Time after Stroke ± SD	3.8 ± 3.5
Gender	
Female/Male	29 (46%) / 34 (54%)
Hemiparesis Side	
Left/Right	37 (58.7%) / 26 (41.3%)
Concordance	
Yes/No	34 (54%) / 29 (46%)

Table 3-2. The means and standard deviations of pretest and post test for stable and improved group

Instrument		*All (N=62)*	*Stable (N=19)*	*Improved (N=38)*	*Study1 (N=37)*
FMA	Pre	36.53 ± 11.49	31.68 ± 11.54	39.26 ± 10.90	34.35 ± 12.46
	Post	40.16 ± 11.58	35.00 ± 10.72	43.24 ± 11.51	36.29 ± 11.84
	Change	3.63 ± 5.57	3.32 ± 5.94	3.97 ± 5.60	1.95 ± 5.4
WMFT	Pre	34.66 ± 11.54	27.37 ± 8.96	38.16 ± 11.47	30.97 ± 12.38
	Post	36.65 ± 11.18	29.63 ± 9.71	40.24 ± 10.8	33.11 ± 12.11
	Change	1.98 ± 2.71	2.26 ± 2.60	2.08 ± 2.78	2.14 ± 2.65

Table 3-3. Responsiveness comparisons for single-population category

Responsiveness Measure	*Participants*	*FMA*	*WMFT*
Pair t-test	All (N=62)	**5.13	**5.76
	Study 1 (N=37)	**2.19	**4.91
Standardize Effect Size	All (N=62)	0.32	0.17
	Study 1 (N=37)	0.16	0.17
Standardize Response Mean	All (N=62)	0.65	0.73
	Study 1 (N=37)	0.36	0.81

** represent statistically significant.

Table 3-4. Responsiveness comparison for two-population category (N=57).

	Minimal Meaningful Clinical Difference		*Guyatt Effect Size*	
	FMA	*WMFT*	*FMA*	*WMFT*
0.8 Specificity	8	4	8/5.94=1.35	4/2.6=1.54
10% total score range	6.6	7.5	6.6/5.94=1.11	7.5/2.6=2.88

CHAPTER 4
DISCUSSION

Using the combined dataset, inconsistent ranking of the instruments exists across different responsiveness indices. T-tests indicated that both the FMA and WMFT showed significant responsiveness. The standardized effect size (SES) showed the FMA to be more responsive than the WMFT, while the standardize response mean (SRM) and the Guyatt effect size (GES) showed the WMFT is more responsive than the FMA.

These conflicting results are not surprising. Stratford (1996) demonstrated the disagreement by using the SES and SRM to compare responsiveness of two instruments, the Jan Van Breemen Function Questionnaire (JVBF) and Raland-Marris Questionnaire (RMQ). While the SES showed the JVBF is more responsive than the RMQ, the SRM showed the opposite result. This confirms that the study design and the responsiveness indices strongly impact responsiveness results(Stratford, Binkley, & Riddle, 1996). In contrast to the inconsistent ranking from analyzing the combining dataset, analyzing all three responsiveness indices of single-population category using the data from Study 1 showed the WMFT is more responsive than the FMA.

Van der Lee and colleague compared responsiveness of the Action Research Arm Test and the FMA. Data were from pretest and post test of 22 chronic stroke participants underwent the CIMT. (van der Lee, Beckerman, Lankhorst, & Bouter, 2001) By applying the means and standard deviations obtained from this study to the SES and the SRM indices, the SES and the SRM of the FMA (0.15 and 0.36 respectively) are almost identical to what we obtained from the Study 1 dataset: CIMT with chronic stroke (0.16 and 0.36 respectively). However, the SES and the SRM recalculated from Van der Lee's study do not equal to what we obtained from combining dataset: CIMT and other treatments with chronic stroke (0.32 and 0.65 respectively).

This finding suggests that the SES and the SRM are “treatment sensitive”. This finding is in accordance with Husted’s statement: “The internal responsiveness of a measure will depend on both the particular treatment and the particular outcomes used to determine treatment effect” (Husted, Cook, Farewell, & Gladman, 2000).

There are several limitations to this study. The pair-t test, the SES and the SRM for single-population category assume participants have homogeneous change. The homogeneous change assumption is unlike to hold since the the three studies in the present analysis employed different treatments. Moreover, under the assumption of homogeneous change, these responsiveness indices are considered a weak design since they cannot account for different amounts of change (Husted et al., 2000; Stratford et al., 1996; Stratford & Riddle, 2005). While combining the data from similar intervention studies increased the sample size, it is unclear how differences in study interventions affected our results.

In GES study design, the decision of selecting cut-off point for dividing participants into two groups (stable vs. improved) was arbitrary. Duncan and colleagues (1999) proposed that 10 to 15 points change in Stroke Impact Scale (SIS) represent clinical meaningful change (Duncan et al., 1999). According to this statement, the clinically meaningful change for the hand domain of the SIS was established proportionally as two points. However, applying two points as cut-off point to distinguish participants as stable or improved is not empirical based. The decision of cut-off point could potentially affect the results of the GES (Stratford et al., 1996). In addition, although the SIS is a reliable, valid and responsive measurement (Duncan et al., 1999), whether hand function domain of the SIS is a valid external criterion is uncertain.

Minimal meaningful clinical difference (MMCD) was arbitrarily set by two different criteria: the amount of change in the instruments reached certain specificity in external criterion

(Wallace et al., 2002) and at 10% of the total range of the scale (Gladstone et al., 2002; van der Lee, Wagenaar et al., 1999). However, the minimal meaningful clinical change is population and instrument dependent. Whether applying these criteria to obtain the MMCD is adequate needs to be further evaluated. Empirical evidence to establishing MMCD prior to further investigate responsiveness analysis is required for future study. Due to the limitation of available data from the existing dataset, the stronger study designs were not feasible in this study. Future studies should employ stronger designs.

In conclusion, while the results are mixed, the WMFT had a tendency to show better responsiveness than the FMA, especially with the more sophisticated responsiveness designs. Future studies, will be necessary to determine the replicability of these findings. With responsiveness designs and statistics being relatively new to healthcare outcomes research, studies are needed to determine the relative strengths and weakness of these designs. Furthermore, item response theory models which determine the item-level psychometrics of instruments may be useful in determining why one instrument is more responsive than another.

APPENDIX A
FUGL-MEYER ASSESSMENT OF UPPER EXTREMITY FUNCTION

Item number and label	Description of items
Reflex Activity	
Item 1: Biceps Reflex	Biceps reflex elicited with reflex hammer
Item 2: Triceps Reflex	Triceps reflex elicited with reflex hammer
Flex Synergy	
Item 3: Flexor Synergy 1	Scapular elevation
Item 4: Flexor Synergy 2	Scapular retraction
Item 5: Flexor Synergy 3	Shoulder abduction
Item 6: Flexor Synergy 4	Shoulder external rotation
Item 7: Flexor Synergy 5	Elbow flexion
Item 8: Flexor Synergy 6	Forearm supination
Extensor Synergy	
Item 9: Extensor Synergy 1	Shoulder adduction with internal rotation
Item 10: Extensor Synergy 2	Elbow extension
Item 11: Extensor Synergy 3	Forearm pronation
Combination of Synergy	
Item 12: Combo of Synergy 1	Hand to lumbar spine
Item 13: Combo of Synergy 2	Shoulder flexion to 90 degrees with elbow extended
Item 14: Combo of Synergy 3	Pronation –supination of forearm with elbow extended
Movements out of Synergy	
Item 12: Out of Synergy 1	Shoulder abduction to 90 with elbow extended
Item 13: Out of Synergy 2	Shoulder flexion to 90-180 with elbow extended
Item 14: Out of Synergy 3	Pronation-supination of forearm with elbow extended
Item 18: Normal Reflex Activity	Normal reflex activity
Wrist	
Item 19: Wrist 1	Wrist stable with elbow at 90
Item 20: Wrist 2	Wrist flexion with elbow at 90
Item 21: Wrist 3	Wrist stable with elbow extended and shoulder at 30
Item 22: Wrist 4	Wrist flexion with elbow extended and shoulder at 30
Item 23: Wrist 5	Wrist circumduction
Hand	
Item 24: Hand 1	Finger mass flexion
Item 25: Hand 2	Finger mass extension
Item 26: Hand 3	Hook grasp
Item 27: Hand 4	Lateral prehension
Item 28: Hand 5	Palmar pinch
Item 29: Hand 6	Cylindrical grasp
Item 30: Hand 7	Spherical grasp
Coordination	
Item 31: Coordination 1	Tremor
Item 32: Coordination 2	Dysmetria
Item 33: Coordination 3	Speed

APPENDIX B
WOLF MOTOR FUNCTION TEST

1. Forearm to table (side)
2. Forearm from table to 25.4-cm box (side)
3. Extend elbow 28cm on table top (side)
4. Extend elbow 28cm on table top (1-lb weight)
5. Hand to table (front)
6. Hand to box (front)
7. Retrieve .45-kg weight from 28-cm line on table top by elbow flexion
8. Lift can to mouth
9. Lift pencil from table
10. Lift pencil clip from table
11. Stack 3 checkers
12. Flip 3 cards
13. Turn key in lock: clockwise to 180 degree, counterclockwise to 180 degree, and back to the starting position
14. Fold face towel
15. Lift basket with 1.35-kg weight from desk (29” high) to bedside table that placed on the desk

Rating scale
0-did not attempt
1-unable to perform
2-performed very slowly or with difficulty, needed greater than 2 attempts, needed assistance from stronger arm, or task modified
3-performed slowly or with synergy
4-almost normal, just not as fast or accurate
5-appeared normal

APPENDIX C
STROKE IMPACT SCALE

The purpose of this questionnaire is to evaluate how stroke has impacted your health and life. We want to know from **your point of view** how stroke has affected you. We will ask you questions about impairments and disabilities caused by your stroke, as well as how stroke has affected your quality of life. Finally, we will ask you to rate how much you think you have recovered from your stroke.

The following questions are about your ability to use your hand that was MOST AFFECTED by your stroke.

7. In the past 2 weeks, how difficult was it to use your hand that was most affected by your stroke to…	Not difficult at all	A little difficult	Somewhat difficult	Very difficult	Extremely difficult
a. Carry heavy objects (e.g. bag of groceries)?	5	4	3	2	1
b. Turn a doorknob?	5	4	3	2	1
c. Open a can or jar?	5	4	3	2	1
d. Tie a shoe lace?	5	4	3	2	1
e. Pick up a dime?	5	4	3	2	1

LIST OF REFERENCES

Bonifer, N., & Anderson, K. M. (2003). Application of constraint-induced movement therapy for an individual with severe chronic upper-extremity hemiplegia. *Physical Therapy, 83*(4), 384-398.

Broeks, J. G., Lankhorst, G. J., Rumping, K., & Prevo, A. J. (1999). The long-term outcome of arm function after stroke: Results of a follow-up study. *Disability and Rehabilitation, 21*(8), 357-364.

Cohen, J. (1977). *Statistical power analysis for behavior science*. New York: Acdemic Press.

de Pedro-Cuesta, J., Widen-Holmqvist, L., & Bach-y-Rita, P. (1992). Evaluation of stroke rehabilitation by randomized controlled studies: A review. *Acta Neurological Scandinavica, 86*(5), 433-439.

De Weerdt, W., & Harrison, M. A. (1985). Measuring recovery of arm hand function in stroke patients: A comparison of the Brunnstrom-Fugl-Meyer test and Action Research Arm test. *Physiotherapy Canada, 37*, 65-70.

Duncan, P. W. (1997). Evidence based medicine. *Physiotherapy Research International, 2*(4), 271-272.

Duncan, P. W., Propst, M., & Nelson, S. G. (1983). Reliability of the Fugl-Meyer assessment of sensorimotor recovery following cerebrovascular accident. *Physical Therapy, 63*(10), 1606-1610.

Duncan, P. W., Wallace, D., Lai, S. M., Johnson, D., Embretson, S., & Laster, L. J. (1999). The stroke impact scale version 2.0. Evaluation of reliability, validity, and sensitivity to change. *Stroke, 30*(10), 2131-2140.

Fugl-Meyer, A. R., Jaasko, L., & Norlin, V. (1975). The post-stroke hemiplegic patient. II. Incidence, mortality, and vocational return in Goteborg, Sweden with a review of the literature. *Scandinavian Journal of Rehabilitation Medicine, 7*(2), 73-83.

Gladstone, D. J., Danells, C. J., & Black, S. E. (2002). The Fugl-Meyer assessment of motor recovery after stroke: A critical review of its measurement properties. *Neurorehabilitation and Neural Repair, 16*(3), 232-240.

Guyatt, G., Walter, S., & Norman, G. (1987). Measuring change over time: Assessing the usefulness of evaluative instruments. *Journal of Chronic Diseases, 40*(2), 171-178.

Heller, A., Wade, D. T., Wood, V. A., Sunderland, A., Hewer, R. L., & Ward, E. (1987). Arm function after stroke: Measurement and recovery over the first three months. *Journal of Neurology, Neurosurgery and Psychiatry, 50*(6), 714-719.

Husted, J. A., Cook, R. J., Farewell, V. T., & Gladman, D. D. (2000). Methods for assessing responsiveness: A critical review and recommendations. *Journal of Clinical Epidemiology, 53*(5), 459-468.

Kazis, L. E., Anderson, J. J., & Meenan, R. F. (1989). Effect sizes for interpreting changes in health status. *Medical Care, 27*(3 Suppl), S178-189.

Kirshner, B., & Guyatt, G. (1985). A methodological framework for assessing health indices. *Journal of Chronic Diseases, 38*(1), 27-36.

Kopp, B., Kunkel, A., Muhlnickel, W., Villringer, K., Taub, E., & Flor, H. (1999). Plasticity in the motor system related to therapy-induced improvement of movement after stroke. *Neuroreport, 10*(4), 807-810.

Kunkel, A., Kopp, B., Muller, G., Villringer, K., Villringer, A., Taub, E., et al. (1999). Constraint-induced movement therapy for motor recovery in chronic stroke patients. *Archives of Physical Medicine and Rehabilitation, 80*(6), 624-628.

Langton Hewer, R. (1990). Rehabilitation after stroke. *The Quarterly Journal of Medicine, 76*(279), 659-674.

Liang, M. H. (1995). Evaluating measurement responsiveness. *The Journal of Rheumatology, 22*(6), 1191-1192.

Liepert, J., Bauder, H., Wolfgang, H. R., Miltner, W. H., Taub, E., & Weiller, C. (2000). Treatment-induced cortical reorganization after stroke in humans. *Stroke, 31*(6), 1210-1216.

Liepert, J., Miltner, W. H., Bauder, H., Sommer, M., Dettmers, C., Taub, E., et al. (1998). Motor cortex plasticity during constraint-induced movement therapy in stroke patients. *Neuroscience Letters, 250*(1), 5-8.

Miltner, W. H., Bauder, H., Sommer, M., Dettmers, C., & Taub, E. (1999). Effects of constraint-induced movement therapy on patients with chronic motor deficits after stroke: A replication. *Stroke, 30*(3), 586-592.

Morris, D. M., Uswatte, G., Crago, J. E., Cook, E. W., 3rd, & Taub, E. (2001). The reliability of the wolf motor function test for assessing upper extremity function after stroke. *Archives of Physical Medicine and Rehabilitation, 82*(6), 750-755.

Moskowitz, E., Lightbody, F. E., & Freitag, N. S. (1972). Long-term follow-up of the poststroke patient. *Archives of Physical Medicine and Rehabilitation, 53*(4), 167-172.

Nakayama, H., Jorgensen, H. S., Raaschou, H. O., & Olsen, T. S. (1994). Recovery of upper extremity function in stroke patients: The Copenhagen Stroke Study. *Archives of Physical Medicine and Rehabilitation, 75*(4), 394-398.

Parker, V. M., Wade, D. T., & Langton Hewer, R. (1986). Loss of arm function after stroke: Measurement, frequency, and recovery. *International Rehabilitation Medicine, 8*(2), 69-73.

Stratford, P. W., Binkley, F. M., & Riddle, D. L. (1996). Health status measures: Strategies and analytic methods for assessing change scores. *Physical Therapy, 76*(10), 1109-1123.

Stratford, P. W., & Riddle, D. L. (2005). Assessing sensitivity to change: Choosing the appropriate change coefficient. *Health and Quality of Life Outcomes, 3*, 23.

Tarkka, I. M., Pitkanen, K., & Sivenius, J. (2005). Paretic hand rehabilitation with constraint-induced movement therapy after stroke. *American Journal of Physical Medicine and Rehabilitation, 84*(7), 501-505.

Taub, E., Miller, N. E., Novack, T. A., Cook, E. W., III, Fleming, W. C., Nepomuceno, C. S., et al. (1993). Technique to improve chronic motor deficit after stroke. *Archives of Physical Medicine and Rehabilitation, 74*(4), 347-354.

van der Lee, J. H., Beckerman, H., Lankhorst, G. J., & Bouter, L. M. (1999). Constraint-induced movement therapy. *Archives of Physical Medicine and Rehabilitation, 80*(12), 1606-1607.

van der Lee, J. H., Beckerman, H., Lankhorst, G. J., & Bouter, L. M. (2001). The responsiveness of the Action Research Arm test and the Fugl-Meyer Assessment scale in chronic stroke patients. *Journal of Rehabilitation Medicine, 33*(3), 110-113.

van der Lee, J. H., Wagenaar, R. C., Lankhorst, G. J., Vogelaar, T. W., Deville, W. L., & Bouter, L. M. (1999). Forced use of the upper extremity in chronic stroke patients: Results from a single-blind randomized clinical trial. *Stroke, 30*(11), 2369-2375.

Wallace, D., Duncan, P. W., & Lai, S. M. (2002). Comparison of the responsiveness of the Barthel Index and the motor component of the Functional Independence Measure in stroke: the impact of using different methods for measuring responsiveness. *Journal of Clinical Epidemiology, 55*(9), 922-928.

Wolf, S. L., Catlin, P. A., Ellis, M., Archer, A. L., Morgan, B., & Piacentino, A. (2001). Assessing Wolf Motor Function Test as outcome measure for research in patients after stroke. *Stroke, 32*(7), 1635-1639.

Wolf, S. L., Lecraw, D. E., Barton, L. A., & Jann, B. B. (1989). Forced use of hemiplegic upper extremities to reverse the effect of learned nonuse among chronic stroke and head-injured patients. *Experimental Neurology, 104*(2), 125-132.

Wood-Dauphinee, S. L., Williams, J. I., & Shapiro, S. H. (1990). Examining outcome measures in a clinical study of stroke. *Stroke, 21*(5), 731-739.

BIOGRAPHICAL SKETCH

Pey-Shan Wen was born on June 22, 1974 in Kaohsiung, Taiwan. She is the middle child with an older brother and a younger sister. She graduated from Kaohsiung Municipal Girl's Senior High School in 1992. She was awarded B.S. degree in occupational therapy in National Cheng Kung University in Tainan, Taiwan in 1996. She enrolled in the advanced master's program in the Occupational Therapy department of the University of Florida in 2003. In 2004, she started the rehabilitation science doctoral program in University of Florida.

Pey-Shan Wen had worked as a licensed occupational therapist for five years before she came to the United States for her master's degree. She worked in Taipei Medical University Hospital from 1996 to 1998. Between 1999 and 2002, she worked as a chief occupational therapist in hospital and worked part-time as a school-occupational therapist in school system.

In the fall 2003, Pey-Shan Wen worked as teaching assistant in Occupational Therapy department. Since 2004, she has worked as a research assistant for Dr. Craig Velozo.