A harmonized and hierarchical method of quantifying upper extremity function post-stroke

Johanne Higgins, B.Sc. (O.T.), M.Sc.
School of Physical and Occupational Therapy
McGill University, Montreal

August 2007

A Thesis Submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Ph.D in Rehabilitation Science

Your file Votre référence
ISBN: 978-0-494-50826-8
Our file Notre référence
ISBN: 978-0-494-50826-8

In dedication to my husband and love Chadwick and to my children Kayleigh, Tyler and Emily who bring joy and balance to my life.

TABLE OF CONTENTS

INDEX OF TABLES iv

INDEX OF FIGURES vi

ABSTRACT vii

ABRÉGÉ viii

PREFACE x

ORGANIZATION OF THESIS XI
CONTRIBUTION OF CO-AUTHORS XIII
STATEMENT OF ORIGINALITY XIV

ACKNOWLEDGMENTS xvi

INTRODUCTION 1

CHAPTER 1 The Impact of Stroke and Upper Extremity Function 3

1.1 THE EPIDEMIOLOGY OF STROKE 3
1.2 THE IMPACT OF STROKE: IMPAIRMENTS, ACTIVITIES, AND PARTICIPATION 3
1.3 REHABILITATION INTERVENTIONS FOR THE UPPER EXTREMITY 4
1.4 THE FIELD OF MEASUREMENT IN REHABILITATION 4
1.5 THE ASSESSMENT OF THE UPPER EXTREMITY AFTER STROKE 5
1.6 DEFINING THE CONSTRUCT OF UPPER EXTREMITY FUNCTION 7

CHAPTER 2 Manuscript 1: The effect of a task-oriented intervention on arm function in persons with stroke: a randomized controlled trial 18

PREFACE TO MANUSCRIPT 1 18
TITLE PAGE 20
ABSTRACT 21
INTRODUCTION 22
MATERIALS AND METHODS 24
DESIGN 25
STATISTICAL METHODS 29
RESULTS 30
DISCUSSION 32
LIMITATIONS 35
CONCLUSION 35
CLINICAL MESSAGE 36

CHAPTER 3. The Science of Measurement **44**

3.1 Classical Test Theory (CTT) 45
3.2 Psychometric properties in CTT 46
3.2.1 Reliability 46
3.2.2 Validity 47
3.3 Scores vs. Measures 48
3.4 Moving from Indices to Measures: The Rasch Measurement Model 53
3.5 The analysis of Item Fit 62
3.6 The Analysis of Person Fit 63
3.7 The analysis of Category Misfit or Ordered thresholds. 64
3.8 The Ascertainment of Unidimensionality 66
3.9 Targeting 67
3.10 Detecting Differential Item Functioning 68
3.11 Item Banking 69
3.12. Linking 70
3.13 Psychometric properties in Rasch measurement 70
3.14 Computer Adaptive Testing 72
3.15 The Rasch Measurement Model vs Item Response Theory (IRT) 73

CHAPTER 4 Rationale and Study Objectives *(Manuscripts 2, 3 and 4)* **74**

CHAPTER 5 Manuscript 2 The use of Rasch Analysis and Item Response Theory in Rehabilitation: A Review of the Literature **75**

Preface to Manuscript 2 75
Abstract 78
Introduction 79
Objective 80
Methods 80
Results 81
Conclusion 85

CHAPTER 6 Manuscript 3 Development and Initial Psychometric Evaluation of an Item Bank created to Measure Upper Extremity Function in persons with Stroke139

Preface to Manuscript 3 139
Abstract 143
Introduction 144
Objective 145
Methods 145
Analysis 147
Results 153
Conclusion 158
Limitations 160
Future work 161

CHAPTER 7 Manuscript 4 .... **187**

PREFACE TO MANUSCRIPT 4 187
DEVELOPMENT AND PSYCHOMETRIC EVALUATION OF A "PAPER-CAT" MEASURE OF UPPER EXTREMITY FUNCTION POST-STROKE 187
INTRODUCTION 191
METHODS 192
ANALYSIS 193
RESULTS 198
DISCUSSION 204
FUTURE WORK 205

CHAPTER 8 Summary and Conclusions **223**

REFERENCES **227**

APPENDICES **263**

APPENDIX 1 Patient consent form **264**

Title of the Study:The Effectiveness of Rehabilitation Therapy in Stroke **264**

APPENDIX 2 – Certificate of Ethical Approval - Institutional Review Board of McGill University.- MANUSCRIPT 1 .... **267**

APPENDIX 3: Agreement of Reproduction – MANUSCRIPT 1 .... **269**

APPENDIX 4: List of Study Measures **271**

APPENDIX 5: Model Quality Criteria for a Rasch Measure **329**

APPENDIX 6: Additional statistical considerations **338**

INDEX OF TABLES

Table 2.1 Demographic and Clinical Characteristics of Study Participants. 37
Table 2.2 Change in Score on the Primary Outcome Measure: the Box and Block Test..39
Table 2.3 Scores on Measures of Arm Impairment and Activity Limitation. 40
Table 2.4 Comparison of the 2 groups on Proportions of Persons who Improved on Indices of Arm Activity Limitation (capacity). 42
Table 3.1 Different Types of Validity...48
Table 3.2 Permissible Statistics 50
Table 3.3 Interpretation of Fit Statistics. 63
Table 5.1 Measures Developed for Stroke, or Used in a Stroke Population. 91
Table 5.2 Generic Measures Developed or Psychometrically Assessed in Stroke along with Other Conditions. 101
Table 5.3 Measures Developed for Brain Injury, or Used in a Brain Injury Population. 106
Table 5.4 Measures Developed for Cerebral Palsy, or Used in a Population with Cerebral Palsy. 112
Table 5.5 Measures Developed for Multiple Sclerosis, or Used in a Multiple Sclerosis Population. 115
Table 5.6 Measures Developed for Parkinson's disease, or Used in a Population with Parkinson's Disease (PD)...116
Table 5.7 Measures Developed for Vision impairments, or Used in a Population with Vision impairments. 117
Table 5.8 Measures Developed for Back Pain or Back Impairments, or Used in a Population with Back Pain or Back Impairments. 122
Table 5.9 Measures Developed for Arthritis and other Inflammatory Diseases, or Used in a Population with Arthritis and other Inflammatory Diseases. 127
Table 5.10 Studies Developing New Measure(s) using Rasch or Item Response Theory 132
Table 5.11 Studies Using Previously Rasch-developed Measures. 137
Table 6.1 Structure of the Combined Data Det Based on Eight Studies.. 163
Table 6.2 Items Chosen to Represent Upper Extremity Function. 165
Table 6.3 Categorization of Influencing (Personal) Factors. 170
Table 6.4 Demographic and Clinical Characteristics of Study Participants. 171
Table 6.5 Items Loading on a Separate Factor for Each of the Studies. 172
Table 6.6 Deleted Items and Reason for Deletion. 174
Table 6.7 Summary of Global Fit Statistics for the Item Bank for Sample No.1. 175
Table 6.8 Summary of Global Fit Statistics for the Item Bank for Sample No.2. 176
Table 6.9 Characteristics of the Items in the Upper Extremity Function Item Bank. Sample No.1 177
Table 6.10 Demographic and Clinical Characteristics of Persons. 183
Table 7.1 Schedule of Evaluation Times Post-stroke for Each of the Eight Studies. 208
Table 7.2. Item bank of Upper Extremity Function Items. 209
Table 7.3 Categorization of Influencing (Personal) Factors. 211
Table 7.4. Demographic and Clinical Characteristics of Study Participants (n = 4058). 212
Table 7.5 Deleted Items and Reason for Deletion. 214
Table 7.6 Summary of Global Fit Statistics for the Item Bank......................................215
Table 7.7 The 35 Items and their Measurement Properties by Location Order. 216

Table 7.8 Final 35-Item Adaptive Measure of Upper Extremity Function Post-stroke...219

INDEX OF FIGURES

Figure 1.1 The ICF Framework to Assess Function 8
Figure 1.2 The ICF Framework to Assess Upper Extremity Function 10
Figure 3.1 Ordinal Scale with Unknown Distances. 49
Figure 3.9 Example Non-Parallel Curves Demonstrating DIF. 69
Figure 2.1 Flow of Subjects Through the Trial..38
Figure 2 2 Effect Sizes in Relation to Samples Sizes of 12 Randomized Controlled Trials 43
Figure 3.1. Ordinal Scale with Unknown Distances...49
Figure 3.2 Level of Spread of Item Difficulties...51
Figure 3.3 Unidimensionality Represented by an Arrow. 52
Figure 3.4 Relationship Between Item and Person. 54
Figure 3.5 The Item Characteristic Curve (adapted from Best Test Design.[138] 55
Figure 3.6 Distribution of Items and Persons on the Same Ruler......................................57
Figure 3.7 Example of Adequate Categorization of Rating Scale. 65
Figure 3.8 Example of Inadequate Categorization of Rating Scale. 65
Figure 3.9 Example Non-Parallel Curves Demonstrating DIF. 69
Figure 3.10 Example of Parallel Curves Demonstrating the Absence of DIF. 69
Figure 5.1 Number of Articles by Year of Publication...90
Figure 6.1 Simplified Rasch Model Overview 184
Figure 6.2 Final Threshold Ordering 185
Figure 6.3 Person and Item Distribution for Sample No.1. 186
Figure 6.4 Upper Extremity Function and Corresponding Subjects' Distribution 187
Figure 7.1 Person-Item Threshold Distribution for the Final 35-Item Measure..............218
Figure 7.2 Test Information Function for the Paper Adaptive Test of Upper Extremity Function 223
Figure 7.1a Example of Disordered Thresholds – The Box and Block Test. 213
Figure 7.1b Example of Ordered Thresholds – The Box and Block Test After Collapsing of the Categories. 213
Figure 7.2 Person-Item Threshold Distribution for the Final 35-Item Measure..............218
Figure 7.3 Test Information Function for the Paper Adaptive Test of Upper Extremity Function 223

ABSTRACT

Only a small proportion of persons with stroke fully recover the use of their more affected upper extremity. The consequences of this are devastating as use of the upper extremity is of the utmost importance to the performance of daily activities. Rehabilitation post-stroke is now being offered in a variety of different settings including acute and rehabilitation hospitals, out-patient rehabilitation centers, at home through home-care and in the community. Evaluating improvement in upper extremity function is difficult because there is no agreed upon measure or set of measures. The proliferation of tests targeting a particular type of client (severe, high level), need (goal setting, clinical practice, research) or setting (hospital/clinic or home) has contributed to the difficulty in creating a harmonized view this construct. A crucial feature of any method for measurement is that true change is detected, the whole spectrum of ability is captured (most basic to most complex tasks), and that comparisons across people and over time are meaningful and not dependent on the units of measurement. It is not unusual to assess persons with stroke with three or four different tests and several questionnaires designed to capture actual use of the arm in daily activities. The measurement protocol currently requires all items of all tests be administered even though there are known redundancies. Summarizing recovery across this measurement spectrum is difficult and hinders communication across disciplines and rehabilitation settings. It has been argued that rehabilitation professionals need to move toward a common language of functional assessment.[1] Using Rasch analysis, a single construct, in this case, upper extremity function, is measured using a ruler-like scale with items hierarchically ordered along its length. This measurement property permits evaluation of relative gains in function (a person has doubled their level function). When computer-assisted, items are chosen for administration automatically depending on ability and on the response to previous items. This yields an accurate estimate of recovery level using the fewest number of items and a minimum amount of time.

ABRÉGÉ

Seule une infime portion des personnes ayant survécu à un accident vasculaire cérébral (AVC) récupère la fonction du membre supérieur le plus affecté. Les conséquences sont dévastatrices car les membres supérieurs sont d'une importance capitale dans l'accomplissement des activités de la vie quotidienne. La thérapie de réadaptation post-AVC est maintenant offerte dans divers milieux tels qu'en institutions de soins aigus, dans les centres de réadaptation ainsi qu'à la maison à travers les services de maintien à domicile et dans la communauté. L'évaluation des gains de la fonction du membre supérieur est ardue car, à ce jour, il n'existe aucun test ou ensemble de tests qui font l'unanimité parmi les professionnels de la réadaptation. La prolifération des tests ciblés à des clientèles, des besoins ou des milieux particuliers a contribue à la difficulté de créer une vue harmonisée de la fonction du membre supérieur. Un aspect crucial d'une méthode de mesure est que les changements réels soient détectés, le spectre complet des habilités soit représenté (des tâches les plus simples aux tâches les plus complexes), et que des comparaisons entre les personnes et dans le temps soient possibles et significatives et ne dépendent pas des unités de mesure des tests utilisés. Il n'est pas rare qu'on évalue des clients avec trois ou quatre tests et questionnaires différents afin de saisir l'utilisation réelle du membre supérieurs dans la vie quotidienne. Les protocoles de mesure actuels exigent que l'on administre tous les items contenus dans un test ou un questionnaire même s'il y a des redondances évidentes. Avoir une vue d'ensemble de la récupération du membre supérieur est complexe et empêche la communication entre les professionnels de la réadaptation et les différents milieux. Il a été argumenté que les professionnels de la réadaptation devraient adopter un langage commun dans le cadre des évaluations fonctionnelles. Les modèles de Rasch permettent de mesurer un seul aspect ou trait, dans notre cas la récupération du membre supérieur en utilisant une échelle construite telle une règle avec les items et les personnes places de façon hiérarchique le long de la règle. Cette propriété permet la mesure des gains relatifs (une personne a doublé son niveau de récupération). Lorsqu'un ordinateur est utilisé afin de présenter les items au client et d'entrer les résultats simultanément, les items peuvent êtres choisis automatiquement à l'aide d'un algorithme basé sur la réponse ou la performance sur

l'item précédent. Ceci produit un estimé juste et précis du la fonction du membre supérieur en utilisant un minimum d'items et de temps.

PREFACE

This thesis presents studies aimed at evaluating the function of the upper extremity in persons having survived a stroke. An impaired upper extremity has a major negative impact on a person's ability to perform their daily activities and pursue their life roles. A first step toward identifying effective interventions to improve upper extremity function is to quantify upper extremity function and to measure true change. There is no agreement amongst rehabilitation professionals or researchers as to how to assess upper extremity function. In order to capture the full realm of function, from the most basic movements to fine manual dexterity and, finally, to the impact of upper extremity impairments on a person's life, both clinicians and researchers have relied on a myriad of upper extremity tests and indices. The first disadvantage of this traditional way is that it places a tremendous burden on patients. Having to perform an endless number of different tasks and having to answer several questionnaires, that are occasionally redundant and not necessarily well targeted to each individual person can be a tiring and frustrating experience. As a researcher or a clinician, it is difficult to interpret and summarize results emanating from several different tests and indices that are scored differently and whose summary score does not have inherent meaning. Over the past several years, Rasch analysis has been increasingly used in the study and development of health care measurement. Rasch models support the development of adaptable instruments for assessing upper extremity function and recovery. Rasch analysis can be used to select items that are targeted to a particular ability level. This is termed *adaptive testing*. When these items are administered with a computer, it is called *Computer Adaptive Testing* or CAT. Test items administered to different samples of subjects or patients at different points in time can also be placed onto a common scale. After items are placed onto a common scale, scores become comparable regardless of the set of items administered. It is possible to produce the "ideal" upper extremity function measure that will contain items spanning the whole range of difficulties and that will be "custom tailored" to each individual patient so that a relatively small number of items will yield maximum information about the patient's level of upper extremity function. This process would avoid having the more recovered patients perform items that are too easy for them and get bored. In addition, this process would avoid the severely affected patients having to try too many items that are too difficult for them and get discouraged by the repetitive failures.

Organization of Thesis

This thesis is comprised of four manuscripts, each representing a step in the evolution of the development of the main goal of the overall project.

After a brief introduction, *Chapter 1* describes the impact of stroke on upper extremity function and reviews the treatment and measurement of upper extremity function as they are being performed at present in the field of rehabilitation.

The first manuscript *(Manuscript 1)* is presented in Chapter 2 and is entitled: "The effect of a task-oriented intervention on arm function in persons with stroke: a randomized controlled trial. In this manuscript, the results of a randomized controlled trial aimed at evaluating the effectiveness of a task-oriented intervention in enhancing upper extremity function in persons with stroke are presented. Within the manuscript, a meta-analysis of different randomized controlled trials aimed at improving upper extremity function and the different tests and indices used in each of the trials is presented.

The second part of this thesis presents work conducted to create a bank of items designed to evaluate the function of the upper extremity after stroke. The recovery of the upper extremity has received an increasing amount of attention over the past few years but is still in its infancy stage when compared to the lower extremity. In order to identify effective interventions to treat the upper extremity of stroke survivors, true measures that quantify upper extremity function and can detect the magnitude of true change need to be developed. Item Response Theory and Rasch models have been used extensively in education and psychology, fields in which abstract concepts that cannot be measured directly by an instrument abound. Health care professionals have now started to adopt these models in order to measure abstract concepts related to health conditions, treatments and recovery. This thesis consists of the exploration of these measurement models in rehabilitation and the application of such models to develop a unique, true measure of upper extremity function in stroke.

Background information on Rasch analysis and Item Response Theory is presented in Chapter 3. First an overview of Classical Test Theory, the way in which most existing tests and indices have been developed is detailed along with definitions of classical psychometric properties. This is followed by an introduction to 'modern measurement models': Rasch analysis and Item Response Theory. The basic concepts are presented and a comparison is made with Classical Test Theory.

Chapter 4 describes the rationale and study objectives answered in the second part of this thesis.

Chapter 5, (*Manuscript 2*) is entitled: "The use of Rasch Analysis and Item Response Theory in Rehabilitation: A Review of the Literature." In this manuscript, an extensive review of the literature on the use of Rasch analysis and Item Response Theory in the field of rehabilitation is presented. This is the first time that such a review is presented and provides the basis and rationale for analyses conducted in Manuscripts 3 and 4.

The third manuscript *(Manuscript 3)* is presented in chapter 6. It is entitled: "Development and Initial Psychometric Evaluation of an Item Bank created to Measure Upper Extremity Function on Persons with Stroke" and outlines the steps undertaken in order to create this unique bank of items that incorporate both tests of capacity and participation, following the International Classification of Functioning, Disability and Health (ICF).

Finally, the last manuscript *(Manuscript 5)* brings it all together and describes a final 'paper version' of an adaptive test of upper extremity function that can easily be used in a clinical setting or incorporated as part of a research project. This new measure of upper extremity function overcomes the disadvantages of the tests and indices that were developed under Classical Test Theory.

Chapter 8, 'Summary and Conclusions' summarizes the findings and presents considerations and suggestions for future work.

Tables and figures are presented at the end of each Manuscript. References for all chapters and manuscripts can be found at the end of the thesis. Additional information, including a description of the tests and indices used in the studies and additional statistical considerations is presented in the appendices.

Ethical approval for the first study: '*The effect of a task-oriented intervention on arm function in persons with stroke: a randomized controlled trial*' was obtained from the Institutional Review Board (IRB) of McGill University. A copy of the consent form and ethical approval are included in the appendices. All other research projects had ethical approval and there is no additional approval required for secondary analyses of the data.

McGill University requires the presentation of literature review and a final conclusion in addition to those presented in each one of the manuscripts. This results in the presence of repetitions in certain sections throughout the thesis.

Contribution of Co-Authors

The candidate coordinated one of the studies, participated in the recruitment and evaluation of subjects in most of the studies included in this thesis. She developed the study questions and study designs and performed the statistical analyses. The candidate was responsible for the interpretation of the findings and writing of the manuscripts. The co-authors functioned as consultants providing feedback on study design, analyses, and final manuscripts.

Dr Nancy Mayo, the primary investigator on the research projects, provided expert guidance for the design of the thesis studies, the statistical analyses and the writing of the manuscripts.

Lois Finch and the candidate learned Rasch analysis together and formed a study group, meeting on a regular basis to discuss important issues and recent developments. Lois provided excellent feedback on some analytical issues and concepts being developed.

Dr. Jacek Kopec also suggested the development of the paper version of the adaptive measure of upper extremity function.

Statement of Originality

My interest in the recovery of the upper extremity after a stroke started with my Master's degree in Rehabilitation Science. The project consisted of measuring upper extremity function and recovery post-stroke and identifying the predictors of upper extremity at five weeks post-stroke. While pusuing the initial stages of my Ph.D., I was working and studying within the department of Clinical Epidemiology at the McGill University Health Center. Several clinical trials are based in Dr. Mayo' s "Performance Enhancement" laboratory. The studies to which I contributed involved the evaluation of stroke patients using several standardized measures of performance and questionnaires a short time after their stroke, and periodically thereafter for a period of one year. I published, as first-author, the results of a randomized controlled trial of rehabilitation for upper extremity in the journal *Clinical Rehabilitation* and is the first manuscript of this thesis. Clinical evaluations consisted of different tests of performance (capacity) as well as questionnaires on individuals' health in general. Thus, I experienced first hand the difficulties in having to administer a plethora of tests and indices and then combine results for interpretation.

My experiences with these difficulties involved in the area of outcome measurement led me to further develop the goal of my doctoral degree into outcome measurement. I developed a stronger understanding of the challenges faced in this area and also the importance to obtain valid estimates of change when attempting to evaluate the effectiveness of an intervention.

The contribution of this thesis has been to develop a method to hierarchically and parsimoniously estimate upper extremity function post-stroke. To attain this goal, I linked, by using Rasch analysis, different measures of upper extremity function in order to create a common bank of items designed to assess upper extremity function. The originality of this thesis also rests on the combination of Rasch analysis, with the conceptual framework of the International Classification of Functioning, Disability and Health (ICF) to define and quantify upper extremity function. One of the accomplishments was an original item bank of that combined, for the first time, both observable tests of capacity or performance and questions aimed at identifying the activity and participation components of the ICF.

Items in a bank are now available for the design of an adaptive test that could easily be used by both clinicians and researchers to accurately and effectively measure upper extremity post-stroke. This item bank can now be used as a platform against which additional, new upper extremity items can continue to be calibrated in the future, expanding the range and improving the psychometric properties of the items designed to assess upper extremity function post-stroke. This item bank also has the potential of developing into a Computer Adaptive Test (CAT) where items are presented and chosen through an algorithm based on the respondent's previous answer. Item banks and CATs have the potential to revolutionize the way outcome measurements are accomplished in rehabilitation and overcome the difficulties that we are facing today. This led the candidate to the second accomplishment: a paper version of an adaptive test of upper extremity function. The selection of a flexi-level format was based on clinical considerations. As computers are not always readily available in today's clinical settings, a 'paper-and-pencil' format will make it easier to use and, except for the direct entry of the data on a computer, offers the same advantages as a CAT in terms of measurement quality and relief of participant burden.

ACKNOWLEDGMENTS

Dr. Nancy Mayo provided me with such an enriching environment where I was able to develop the skills required to conduct research. I started working with Dr. Nancy Mayo in 1995 when I entered the Master's program in Rehabilitation Science at the School of Physical and Occupational Therapy, McGill University. A few years later, in 2001, she encouraged me to start a Ph.D. and I embarked with no hesitation. Her involvement and her originality with each and every project has triggered and nourished a strong passion and motivation to pursue research in this exciting field. I am grateful in every possible way and hope to keep collaborating in the future.

I would like to express my sincere thanks to the members of my committee, Dr. Nicol Korner-Bitensky whose expertise in stroke research was invaluable in guiding the steps towards the completion of this project. I would like to express my gratitude to Dr. Jacek Kopec for his crucial contribution regarding statistical issues relating to Rasch analysis.

A special thanks to my colleague and friend, who I now call Dr. Lois Finch, for her precious help in understanding these difficult concepts. We started this journey together and surmounted obstacles together. I could not have done it without our regular "Rasch meetings" where we exchanged opinions and knowledge about what we had learned.

I extend a special thank you to Susan Scott and Lyne Nadeau for their statistical expertise and most of all their time and their patience. I would also like to extend my gratitude to Dr. Sara Ahmed, Dr. Nancy Salbach, and Annie Tessier and well as all of the many officemates I have had in the past six years. We not only shared an office but also part of the journey. I express my gratitude to Diana Dawes who kindly reviewed my protocol. Thank you to Carolina Moriello for her friendship, her advice and her kindness, as well as for the proofreading of sections of this thesis.

I extend my deepest gratitude to my parents for their unfailing support and encouragements and for all their help with the kids. I would like to thank them for all they have done for me throughout my life as well as during this thesis. To my sister Marie-Josée, thank you most of all for just listening, and to my sister Patricia for our "talks" on the phone. Special thanks to my parents in law for their time and willingness to help with the kids, it was immensely appreciated. I would like to thank my children Kayleigh, Tyler and Emily who keep me balanced, bring me joy and provide meaning to everything I do. Most of all, I would like to thank my husband Chadwick for his love and unwavering support. I could not have done it without him and will be forever grateful. I would like you to know that now is your time to fulfill your own dreams and aspirations.

Finally, I acknowledge the financial support by the Canadian Stroke Network, the McGill University Health Centre Research Foundation and the Fonds de la Recherche en Santé du Québec. Finally, I would like to thank everybody who played a part in the successful realization of thesis.

INTRODUCTION

The upper extremity plays a central role in a person's life from the ability to perform basic activities of daily life such as bathing and dressing to the ability to carry out family and social roles. Stroke has a devastating effect on the upper extremity of most stroke survivors and impairments of the upper extremity have an important impact on the quality of life of these persons.

The recovery of upper extremity function after stroke lies on our ability to deliver effective treatments. Although the upper extremity after stroke has progressively become the focus of an important body of literature in the field of rehabilitation, the debate is ongoing as to which treatments are effective. To date, there has not been an agreement amongst rehabilitation professionals or researchers.

The measurement of upper extremity function is paramount in evaluating the effectiveness of interventions. At this time, there are a number of trials evaluating different treatments, each using their own outcome assessments, making comparisons of change among subjects as well as changes over time for a given subject very difficult. Traditionally, the effectiveness of rehabilitation interventions has been evaluated by administering different tests of capacity where the performance of patients was observed and scored on an ordinal scale. Measurement in the field of rehabilitation has reached a turning point. An increasing number of researchers have recognized the fact that in order to move the field of rehabilitation towards evidence-based practice, new measures, based on standardized metrics have to be used. Using Rasch analysis, it is possible to create a ruler to measure upper extremity function by arranging items along a difficulty continuum and the persons along the ability continuum.[2] The creation of a measure implies that

parametric statistics can be used on the scores obtained on the measure. In addition, with the use of Rasch, it is possible to estimate a person's score without all the items in the measure having been administered. Using psychometrically sound outcome measures to assess the effectiveness of treatments being delivered to persons with stroke is the first step in devising optimal rehabilitation interventions. Although advances have been made in the treatment of the upper extremity of persons with stroke, research is still needed to solve the enigmas still surrounding this domain of rehabilitation. Because finding effective interventions lies in our ability to measure upper extremity function adequately, a first step is to create a true measure of upper extremity function.

Objectives

The overall objective of this thesis was to create a method to parsimoniously and hierarchically quantify upper extremity function in persons with stroke. As this thesis is a compilation of manuscripts, there are several sub-objectives relating to the research presented as indicated below.

Sub-objectives

1) To evaluate the efficacy of a task-oriented intervention in enhancing arm function in persons with stroke;

2) To summarize the use of latent theory models, namely Rasch analysis and Item Response Theory, in rehabilitation;

3) To identify how many unique constructs comprise upper extremity function from existing pools of upper extremity function indices;

4) To use the Rasch measurement model to identify items that are able to discriminate between levels of recovery along the construct(s);

5) To hierarchically align items to create a bank of items and allow the assignment of a recovery level for individuals;

6) To develop a paper version of an adaptive measure of upper extremity function for clinical use of the new measure.

CHAPTER 1 The Impact of Stroke and Upper Extremity Function

1.1 The Epidemiology of Stroke

Stroke is the fourth leading cause of death, affecting between 40,000 and 50,000 Canadians each year [3] and its incidence is not declining.[4] According to the Heart and Stroke Foundation of Canada, the incidence of stroke is expected to increase by 68% within two decades.[3] In 2003, approximately 300,000 survivors were living with the consequences of stroke in this country and survivors are often faced with a multitude of effects ranging from speech and memory deficits to a complete loss of movement in a limb.[3] In fact, loss of arm function is very common, occurring in up to 85% [5,6] of stroke survivors. Many face physical disabilities that greatly diminish their quality of life and as many as 15% of stroke survivors will require long term care services.[7]

As the population ages, the number of persons surviving a stroke and thus living with its sequalae will increase. Rehabilitation of stroke survivors is important as it aims to improve their functional status, facilitate the return to pre-stroke level of function or palliate persons to this loss of function in the arm. The ultimate goal is community reintegration.

1.2 The Impact of Stroke: Impairments, Activities, and Participation

Losing upper extremity function can have devastating consequences as use of the arms is indispensable in the accomplishment of all basic and instrumental activities of daily living. A great number of survivors require assistance to perform basic daily functions, such as dressing and personal hygiene. From home to workplace, the arms are constantly solicited for a multitude of tasks ranging from simple to very complex and highly skilled movements. Limitations in the use of the affected upper extremity can lead to a diminished sense of well-being one year after a stroke. [8,9] Despite rehabilitation efforts, many stroke survivors do not recover the use of their upper extremity. In fact, only about 15% of those suffering from severe stroke recover function in their more affected hand. [10] Even when the motor impairment is mild, its impact on the performance of daily activities can be very frustrating and can lead to discouragement, a discontinuation in the use of the affected limb[11] and ultimately to learned nonuse.

1.3 Rehabilitation Interventions for the Upper Extremity

Different rehabilitation strategies for the affected arm and hand have been studied. None of the studies comparing the effectiveness of these different strategies has consistently proven one approach to be better than the other.[12-19] "Constraint-Induced Movement Therapy (CIMT)", a relatively new approach, [20] is now being studied extensively. Most studies [21-26] evaluating the effects of CIMT, or a modified version of this modality, have demonstrated positive results. To date, four randomized controlled trials [21,23,25,26] have been conducted. The most recent randomized clinical trial, known as the EXCITE (Extremity Constraint Induced Therapy Evaluation) trial showed significant and relevant improvement in upper extremity function lasting for at least one year.

1.4 The Field of Measurement in Rehabilitation

In rehabilitation settings, measuring function is a daily occurrence whether for clinical assessment or for the purposes of research projects. Outcome measures are used for a variety of reasons, from attempting to establish the effectiveness of a particular treatment to the allocation of appropriate health care resources, all in an attempt to move toward evidence-based practice. Many important decisions are based on such outcome evaluations. One of the major challenges faced by health care providers and rehabilitation professionals is ensuring that the tests and indices used are psychometrically sound. They must be valid, reliable and sensitive to change. The different concepts captured by these assessments must also be easily interpretable by the clinicians using them. One of the many responsibilities of the therapists treating persons following a stroke is the choice of an appropriate outcome assessment. It must be adequate for the clientele and also be suitable for the goal for which it is intended (establishment of the effectiveness of a particular treatment or allocation of appropriate health care resources).

Evaluation of impairments and activity limitations is arduous due to the heterogeneity of the stroke population not only in terms of deficits but also in terms of recovery. Many factors, besides rehabilitation, dictate how fast and how completely persons will recover from stroke, such as the location of the stroke and the initial severity of the deficits, unique and different in each individual person.

Andresen[27] has summarized desirable properties of tests and indices. These include the traditional psychometric properties used in Classsical Test Theory: reliability, validity and responsiveness. In addition, such characteristics as conceptual, item/instrument bias, measurement model, respondent and administrative burden as well as alternate and accessible forms are also considered important criteria to consider when choosing a test or index.

1.5 The Assessment of the Upper Extremity after Stroke

For purposes of clarity, definitions of the terms used in this thesis are presented below. The definitions conform to those recommended by Sloan et al.[28]

Construct: intangible, theoretical entity that is operationalized into one or more items.

Item: single question which can be used as a stand alone question, as part of a series of loosely affiliated questions, or as part of a psychometrically sound index.

Index: psychometrically sound collection of items with an underlying theoretical framework that distinguishes between inter-related constructs relevant to a given health condition.

Tests: direct indicator of the attribute (ie. dexterity test).

Scale: the response options or units for an item, an index or a test.

Tool/Instrument: a device, a piece of equipment.

The term "measure" is reserved for "true measures" such as the one developed using the particular model used in this proposal. It is also used as a verb, an action.

According to Lawton,[29] functional assessment is *"any systematic attempt to measure objectively the level at which a person is functioning in a variety of areas"*. Standardized tests or indices of upper extremity function are now being used in a variety of clinical settings, from acute care to rehabilitation, to home health. Monitoring improvement in upper extremity function can be difficult given the paucity of instruments that can accurately reflect function across different stages of recovery. A therapist or a researcher requires a complete arsenal of tests and indices to examine not only the short-term outcomes of stroke but also the long-term outcomes,[30] and there is no consensus as to which test or index is the best. Some are appropriate for assessing the immediate consequences of the stroke, when persons with stroke are still unable to perform activities whereas others, such as those assessing the area of activity and participation, are more appropriate when persons have reached a higher level of functioning. A compilation of twelve randomized controlled trials evaluating the effect of different treatment strategies for the upper extremity shows that twelve different outcome assessments were used with little overlap across studies.[31] Some authors stipulate that their failure to observe an effect of the treatment may have been due, in part, to the outcome assessment they used. It may not have been targeted to the population under study and thus it was not possible to detect changes.[32]

Difficulty in evaluating the upper extremity may be due in part to the great variety of tasks and activities that are accomplished by the upper extremity. While the lower extremity is used primarily for locomotion, the upper extremity has an unlimited number of actions that also differ from one individual to another. Also, the activities performed by the upper extremity often comprise a high level of motor control and fine dexterity. According to Richards and colleagues,[33] an ideal measure of upper extremity function would be comprised of a wide spectrum of items that would span the complete range of item difficulty from the easiest tasks that can be performed by the most severely affected persons to the hardest ones that can only be accomplished by those with near normal upper extremity function. The tasks included in the measure should require both unilateral and bilateral activities and would take into account the quality of the movements performed. It would also have to be valid and reliable in the specific population targeted by the instrument.

Another group of researchers [34] has defined functional assessment as being aimed not only at evaluating an individual's abilities but also at the limitations or barriers that may hinder the person's ability to perform certain activities. These barriers may be pathophysiological, arise from impairment, functional limitations, or from the environment. According to this group, fulfillment is achieved only when there is equilibrium between health and functioning and the barriers.

1.6 Defining the Construct of Upper Extremity Function

According to Rudman and Hannah,[35] therapists must determine what they want to measure prior to choosing tests or indices. Before an attribute can be measured, a clear definition of that attribute is necessary. To facilitate the restoration of function in persons who have survived a stroke is the goal of rehabilitation. In order to accomplish this, therapists need to have a clear view of this concept and how to measure it. There are no agreed upon definitions of function or of upper extremity function. In a paper on the theory of function,[36] the author concludes that theory must drive the measurement of function and that conceptual models need to be identified.

What is upper extremity function? The World Health Organization's (WHO)[37] International Classification of Functioning, Disability and Health (ICF) provides a framework and a biopshychosocial model for function and has classified health and health related domains (Figure 1.1). These domains include body, individual, societal and environmental perspectives. The domains are: (1) body structure and function and (2) activity and participation. *Functioning* includes all body functions, activities, and participation and any alteration in functioning is called *disability*. An *activity* is defined as the performance or execution of a task or action. *Participation* is the involvement of the person in life situations. According to the World Health Organization, the ICF classification can be used towards a variety of goals including:

-The assessment of individuals: "*What is the person's level of functioning?*"

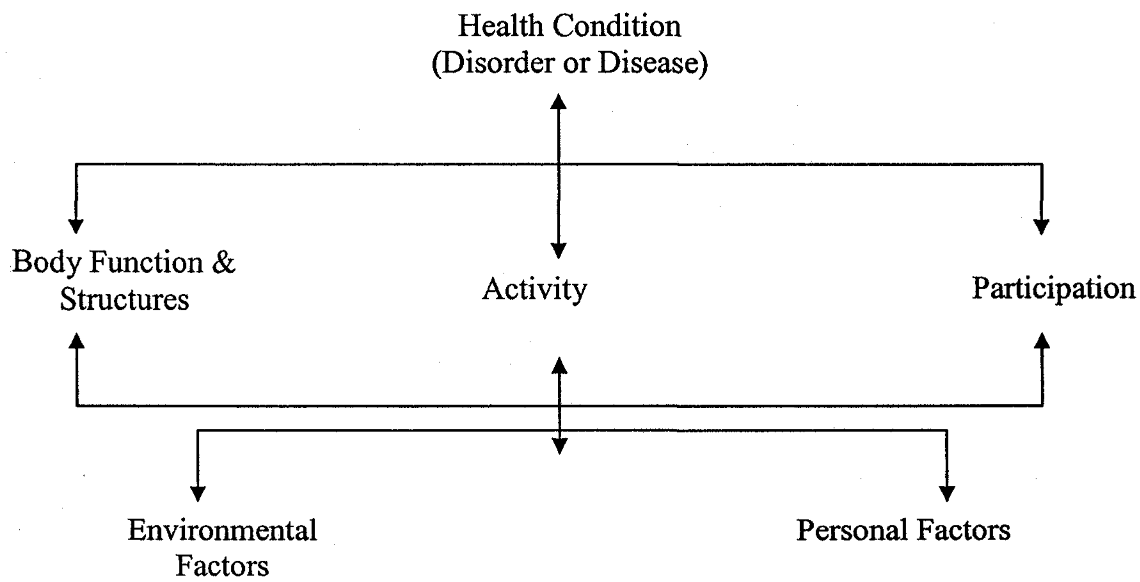
-Individual treatment planning: "*What treatments or interventions can maximize functioning*?"

-The evaluation of treatment and other interventions: "*What are the outcomes of the treatment? How useful were the interventions?*"

-Communication among physicians, nurses, physiotherapists, occupational therapists and other health workers, social service workers and community agencies

-Self-evaluation by consumers: *"How would I rate my capacity in mobility or communication?"*

Figure 1.1 The ICF Framework to Assess Function



One of the aims of the ICF is to provide a common language to improve communication across users: health care professionals, researchers, policy makers as well as clients [37] and thus will be used to define upper extremity function and to classify the existing upper extremity outcome tests and indices. The impact of a stroke on a person's upper extremity function can be illustrated using the ICF framework.

Impact on Function

A stroke is an interruption of blood flow to a part of the brain that causes neurons in the affected area to die. Depending on which area was affected, it can have an impact on upper extremity function. Upper extremity impairments include for example, decreased strength and decreased range of motion.

Impact on Activities

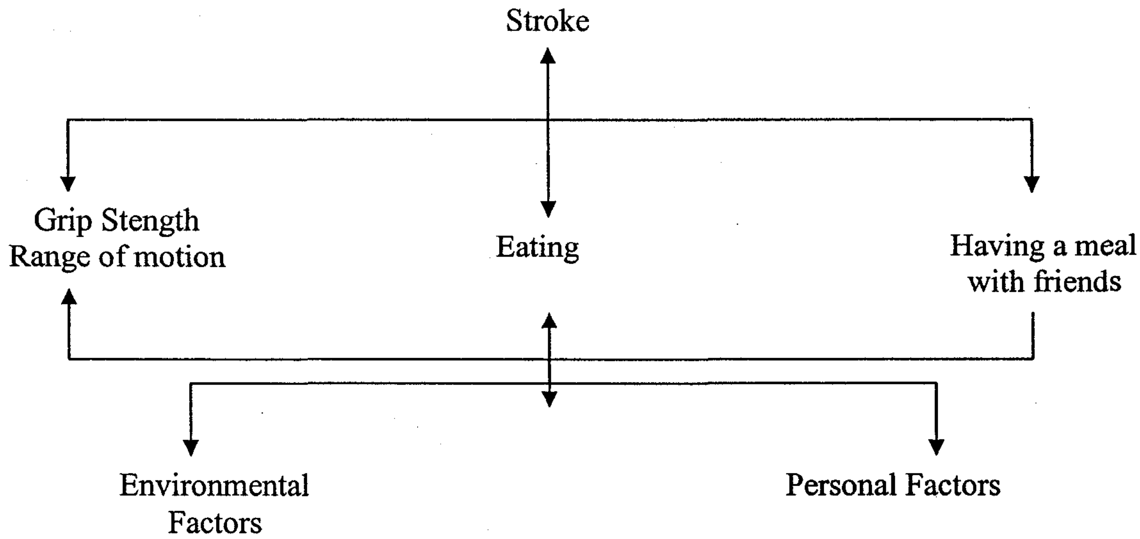
When a stroke affects a person's upper extremity, it can produce important limitations on a person's ability to perform specific activities of daily living such as eating, bathing and dressing.

Impact on Participation

A stroke may prevent a person from fulfilling their roles and participating in meaningful activities that are regarded as normal. Social roles may be different from person to person. They can be being a bread winner for a family, being a parent or a grand-parent, taking care of another person, participating in leisure activities, etc…

Upper extremity function is the result of the interrelationships between these components. Because the goal of rehabilitation interventions is to improve upper extremity function, each one of the components has to be considered and evaluated. Figure 1.2 illustrates the impact of a stroke on upper extremity function.

Figure 1.2 The ICF Framework to Assess Upper Extremity Function



Barreca and colleagues[38] also developed their own definition of upper extremity function that also included the completion of various activities of living, work and leisure. Typically, recovery of the upper extremity following a stroke progresses through a series of stages. Usually the recovery occurs from proximal to distal, shoulder movements appearing first and the hand recovering last.[39] It is believed that motion occurs first in the proximal and then in the more distal portions of the arm.[40,41] This pattern of recovery is similar to the normal acquisition of motor skills in young children. The flexor synergy of the upper limb is the first movement pattern to recover after the flaccidity stage immediately following the acute episode. Then, the spasticity increases and synergy patterns or some of their components can be preformed voluntarily. At later stages, the spasticity declines, movements that deviate from synergies become possible and isolated joint movements can finally be performed with ease.[42]

Therefore, at the level of impairment, shoulder movements should be the easiest to accomplish whereas fine movements of the hand should be the hardest. Likewise, at the activity level, activities requiring the use of fine movements of the hands are the hardest. Once the movements of the arm and hand are coordinated and performed with ease, persons are able to use their affected arm in daily activities and finally, improve their health-related quality of life by fulfilling their roles in society.

Also, according to Patrick,[43] capturing function and its relationship with perceived quality of life is of the utmost importance when providing healthcare because it may help identify factors modifiable through interventions, including rehabilitation. Because the association between disability and perceived quality of life is not always in the expected direction, other factors such as individual characteristics as well as environmental factors must be considered in the delivery of care.[43] Shumaker and Naughton [44] define health related quality of life (HRQL) as *'people's subjective evaluations of the influences of their current health status, health care, and health promoting activities on their ability to achieve and maintain a level of overall functioning that allows them to pursue valued life goals and that is reflected in their general well-being'.*

Therefore, the ultimate goal of rehabilitation interventions is to improve or "recover" a person's quality of life. To this end, many investigators have moved away from relying only on clinical tests and indices and have started to incorporate assessments that target the individual's own perception of their health status in order to capture "true" recovery. Thus, indices of impairment, indices of performance of tasks or activities, indicators of use of the affected upper extremity and indices of health related quality of life would cover the entire spectrum of "upper extremity function", and define, by the same token the concept of upper extremity function. In other words, *upper extremity function* is the ability to move the upper extremity in a way that allows the accomplishment of daily life activities in a fulfilling and satisfying manner.

At this time, separate indices of upper extremity function are used, each one capturing a narrow level of upper extremity function, there is no link between the tests/indices, and it is difficult to extract and communicate meaningful information. The rehabilitation field is predominated by tests and indices that are scored differently, each having their own scaling and whose total scores do not have any inherent meaning. Another noteworthy disadvantage is the burden to clients and to participants to research projects as the assessment can be lengthy and demanding.

In accordance with the framework provided by the ICF, outcomes related to upper extremity function used in different research projects within the McGill stroke studies as well as the Canadian Stroke Registry are classified in this thesis as:

Tests and Indices of Impairment comprise outcomes which target deficits at the impairment level (e.g. the STroke REhabilitation Assessment of Movement (STREAM) in which the items are basic movements or grip strength).

Tests and Indices of Activity Limitation consist of outcomes evaluating individual activities performed by the patient (e.g. pouring a glass of water or tying a scarf).

Indices of Upper Extremity Use is used to describe questionnaires aimed at assessing a person's actual use of their arm in daily activities.

Indices of Health-related Quality of Life consist of outcomes such as the EuroQoL-5D (EQ-5D) Index,[45] The Health Utility Index[46,47] and the Medical Outcomes Study 36-Item Short form Questionnaire (SF-36).[48]

Tests and Indices of Impairment

The Box and Block Test (BBT) [49] evaluates gross unilateral manual dexterity. It consists of moving one by one the maximum number of blocks from one compartment of a box to another in one minute. This test has been shown to have test-retest reliability greater than 0.9.[49] Desrosiers and associates [50] verified construct validity of this instrument in an elderly population. Significant correlations were found between the BBT, an upper limb performance measure, and a measure of functional independence (the SMAF [51] ($r = 0.42$ to 0.54).[50] Age and gender specific norms have been established.[52] The score reported is the number of blocks transferred in one minute.

The ***Canadian Neurological Scale*** **(CNS)** is an index of stroke severity. Internal consistency was determined (Cronbach's alpha = 0.792) [53] and cutoff values have been established predicting mortality and the occurrence of a second vascular event within six months of stroke. Content validity has been demonstrated. An evaluation of concurrent validity compared the CNS with a standard neurologic evaluation, resulting in Spearman rank correlation coefficients ranging from 0.574 to 0.775. ($p<0.001$).[54] Two items evaluate the upper extremity with a maximum score of 1.5 for each.

The ***Chedoke-McMaster Stroke Assessment Scale***[55] is divided into two parts, the Impairment Inventory and the Activity Inventory and was specifically designed for stroke survivors.[55,56] It is a performance based assessment. The Impairment Inventory includes shoulder pain, postural control, the arm, the hand, the leg, and the foot. The arm and hand items were shown to be reliable (ICC = 0.95 and 0.93 respectively for intra-rater, 0.93 and 0.85 for inter-rater, and 0.84 and 0.85 for test-retest).[55,56] Each item is scored on a seven-point scale where the maximum score is seven.

Grip Strength : The Jamar™ dynamometer is a tool used to assess grip strength. Good inter-rater (Pearson product-moment correlation coefficient = 0.97 or above) and test-retest reliability (Pearson product-moment correlation coefficient = 0.80 or above) have been observed using standard procedures [57] and gender and age-specific normative data are available. [58] Grip strength is reported in kilograms of force.

The Nine-Hole Peg Test (NHPT)[59] is a test of fine manual dexterity consisting of placing nine dowels into nine holes and removing them in the shortest amount of time possible. It evaluates fine hand motor skills. A study by Mathiowetz et al. [60] demonstrated high interrater reliability (right: r = 0.97, left: r = 0.99) and a moderate to high test-retest reliability (right: r= 0.43, left: r = 0.43). Also, clinical norms for adults 20 to 75+ years of age for both males and females were established [60] and was used in a study in the acute phase post-stroke.[61] It is reported in seconds taken to complete the task.

The Rankin Index[62] is a global functional health index that focuses on physical disability although some authors mention that mental and physical adaptations to neurological deficits are incorporated. Interrater reliability was found to be satisfactory with $\kappa = 0.56$ and the weighted $\kappa = 0.91$.[63] The index consists of 6 grades, from 0 to 5, with 0 corresponding to no symptoms and 5 corresponding to severe disability.

The STroke REhabilitation Assessment of Movement (STREAM)[64] assesses voluntary motor ability of the upper and lower limbs and basic mobility post stroke. It includes 30 items, 10 of which assess voluntary motor ability of the upper extremity. Excellent interrater and interrater reliability were reported with generalizability correlation coefficients of 0.98-0.995 for total STREAM score and for subscales. [65] Correlation coefficients for test-retest reliability was 0.96 for the upper extremity subscale.[65] Each item of these ten items is scored on a three-point ordinal scale to which letters a, b, and c are added to assess the quality of the movement for a total of 30 points for the upper extremity subscale.

Tests and Indices of Activity Limitation

The Frenchay Arm Test (FAT)[66] assesses recovery in arm function after stroke and has been tested in a chronic, subacute and acute stroke population. It consists of five pass or fail tasks and makes use of the familiar objects. The interrater reliability of the test has been demonstrated (Spearman's rho: 0.68 to 0.90).[67] The test-retest reliability has also been established (Spearman's rho: 0.75to 0.99).[67] The maximum possible score is 5. The time required to perform the task is not recorded and the quality of the movements is not graded.[66]

The Test Évaluant la Performance des Membres Supérieurs des Personnes Agées (TEMPA)[68] consists of 9 tasks, both unilateral and bilateral. The objects used in the administration of the TEMPA are familiar objects that most people use in everyday life. Interrater and test-retest reliabilities were 0.75 to 1.0 and 0.70 to 1.0 respectively (Cohen's weighted Kappas and ICCs). [68] Correlations between the TEMPA and the Action Research Arm Test[69] and the Box and Block Test [49] were: $r = 0.90$ to 0.95 and 0.73 to 0.78, respectively, depending on the task. Norms are only available for the time to execute the tasks and not the ratings. [70] It has been used in a clinical trial involving stroke patients. [71] The performances are timed and the quality of the movements is rated on a four-point ordinal scale from 0, the task is successfully completed to -3, the task could not be completed.

Indices of the Amount of Use of the Upper Extremity

The Barthel Index[72] assesses performance in self-care and mobility. The interrater reliability of the Barthel Index in a mixed neurological population using the Pearson product moment correlation was 0.99 ($p<0.001$) for total scores.[73] Test-retest reliability was demonstrated with a kappa score of 0.98.[74] The four items requiring use of the arm include feeding, personal hygiene, bathing and dressing/undressing. Two items are scored out of 5 and two are scored out of 10 resulting in a maximum summative score of

30.

*The **Older Americans Resources and Services Scale-Instrumental Activities of Daily Living (OARS-IADL)*** [75] is designed to provide a profile of functioning and the need for services for older individuals living with some degree of impairment in the community. Interrerater reliability for selfcare was 0.86 (ICC)[75] and test-retest reliability between two IADL assessments (five weeks apart) was 0.72.[75] It has been shown to correlate with the SF-36 Physical Function subscale (0.36), with the Katz Index (0.33) and with the ADL subscale of the Functional Status Questionnaire (0.70).[76] It is an index in which each item is scored on an ordinal scale from zero to two, a higher score indicating a higher level of function. Two items are considered to relate to the use of the arms (meal preparation and housework). They are scored from 0 to 2 with 2 representing the highest score.

*The **Preference-Based Stroke Index (PBSI)***[77] includes 10 items; walking, climbing stairs, physical activities/sports, recreational activities, work, driving, speech, memory, coping and self-esteem. Content validity and preliminary evidence of construct validity has been demonstrated.[77] Moderate correlations were found between the physical function ($r = 0.78$), vitality ($r = 0.67$), social functioning ($r = 0.32$). It was also shown that the PBSI can differentiate stroke patients by severity ($p<0.05$).[77] Each item has a 3-point response scale. Recreational activities, Work/Activities and Driving were considered to be related to the upper extremity.

*The **Reintegration to Normal Living Index (RNL)***[78] is an index composed of eleven statements including the following domains: indoor, community, and distance mobility; self-care; daily activity (work and school); recreational and social activities; general coping skill; family role(s); personal relationships; and presentation of visual analogue scale. It has been tested in stroke. Cronbach's alpha for internal consistency were 0.90, 0.92 and 0.95 for patients, significant others and health professionals respectively.[79] Interrater reliability ranged from $r = 0.39$ to 0.69[79] and test-retest reliability for older individuals living in the community was $r = 0.83$.[80] Also with community dwelling elders, the RNL was related to the Canadian Occupational Performance Measure and to the Satisfaction with Performance Scaled Questionnaire ($r = 0.72$).[81] The total score can be converted to 100. Higher scores represent a higher level of reintegration. Two items, self-care needs and recreational activities require use of the upper extremities.

Indices of Health-Related Quality of Life

The EuroQoL-5D (EQ-5D) Index[45] comprises five items divided into two sections, devised to assess health-related quality of life and has been tested with stroke survivors. Kappa scores for inter-rater agreement ranged from 0.05 to 0.64 between patient and proxy when the questionnaire is interview based. [82] The ICC for test-retest reliability was 0.83 for stroke persons at 3 weeks. [83] Pearson correlation coefficients between the EQ-5D and subscales of the Medical Outcomes Study 36-Item Short form Questionnaire (SF-36) in the stroke populations varied between 0.17 and 0.64.[84,85] Two items are related to the use of the upper extremity: self-care and usual activities. They are scored using a 3-point scale (0 to 2).

The Health Utility Index Mark3 (HUI-III)[46,47] was developed from two previous versions. The HUI-III has been modified to be used in the general population health surveys to monitor the health of the general population and provide a summary score of health-related quality of life. Assessment of the different attributes in the HUI-III is based on capacity rather than performance. Inter-rater reliability in the stroke population ranges from ICC = 0.39 to 0.78 depending on the domain.[86] Responsiveness estimates were obtained in an injured worker population (standardized response mean = 0.57 and standardized effect size = 0.40).[87] Each attribute contains 4 to 6 levels of ability. Each possible combination of response choice describes a health state. Using a scoring algorithm, each health state is then assigned a utility value that ranges from 0 (worst possible health state) to 1.0 (best possible health state). Seven items are related to the use of the upper extremity.

The Medical Outcomes Study 36-Item Short form Questionnaire (SF-36)[48] is a health-related quality of life measure. Interrater reliability between proxy and subject in a sample of stroke survivors demonstrated ICCs varying between 0.15 for the emotional problems subscale to 0.67 for the physical functioning subscale.[88] Test-retest reliability in stroke populations varied from 0.30 to 0.81 depending on the subscale.[83] Correlations between subscales of the SF-36 and the EuroQoL in a stroke population demonstrated that comparable dimensions on each of the measures correlated more highly than between less comparable dimensions.[84] There are two arm-related questions: grocery carrying and bathing/dressing. These items are scored on an ordinal scale from one to three where higher scores indicate better functioning. Two items are related to the use of the upper extremity: Pick-up/lift grocery bags and take a bath/dress.

The Stroke Impact Scale (SIS)[89] is a 59-item index that evaluates the impact of stroke on a variety of domains: strength, hand function, mobility, activities of daily living, emotion, memory, communication, and social participation. Reliability estimates for internal consistency (Cronbach's alpha) range between 0.93 and 1.00.[90] Test-retest reliability (ICC's) vary between 0.70 and 0.92 except for the emotion domain (ICC = 0.57). When correlated with other measures: the Folstein Mini-Mental State Examination, the Barthel Index, Lawton IADL, and Motricity scale (Rankin), Pearson correlation coefficients were all above 0.67, except for the proxy SIS memory domain (0.37). Total scores for each domain are calculated from 0 to 100, 100 being the best outcome. Thirteen items assess tasks related to the use of the upper extremity.

CHAPTER 2 Manuscript 1: The effect of a task-oriented intervention on arm function in persons with stroke: a randomized controlled trial

Preface to Manuscript 1

In manuscript 1, the results of a two-group, parallel-groups, stratified, block-randomized controlled trial aimed at evaluating the efficacy of a task-oriented intervention in enhancing upper extremity function in persons with stroke are presented. The experimental intervention involved practice of functional, unilateral and bilateral tasks that were designed to improve gross and fine manual dexterity whereas the control intervention was comprised of walking tasks. The consent forms used in this trial are presented in Appendix 1. The Certificate of Ethical Approval provided by the Institutional Review Board of McGill University can be found in Appendix 2. The tests and indices used in this trial are presented in Appendix 4 along with all the tests and indices used throughout this thesis.

As part of my training, I participated in the coordination of this trial. This entailed overseeing recruitment of participants, the randomization procedure, the training and management of personnel, delivery of the interventions, quality assurance and data entry. I conducted all the analyses presented in this manuscript.

The primary test of arm function used in this study was the Box and Block Test. Secondary tests included the Nine-hole Peg Test, maximal grip strength, the Test d'Évaluation des Membres Supérieurs des Personnes Agées and the Stroke REhabilitation Assessment of Movement. As well, several tests and questionnaires to evaluate the use of the affected arm and also health-related quality of life were used. This was done in an attempt to capture the breadth of upper extremity function, from the inability to perform the most basic movement to limitations in the participation of the individual in life roles brought about by the impairments in the affected arm. Data were analyzed on the basis of intention to treat. Standard statistical approaches were used to evaluate treatment effects: group comparisons were made using a *t*-test for independent samples for variables measured on a continuous scale and the Wilcoxon Rank Sum Test for variables measured on an ordinal scale. T-tests and Wilcoxon rank sum test were also done on the change scores between groups for each one of the tests and indices used. The finding of positive results would enhance the present body of knowledge concerning not only the treatment of upper extremity impairment after stroke but also its assessment. Finding tests and indices that are easy to administer, have good psychometric properties and are responsive to changes in upper extremity function is not an easy task as there are no 'gold standards'.

The effect sizes of several studies reporting on an intervention aimed at improving upper extremity function were calculated and presented. It was observed that in the twelve randomized controlled trials presented, twelve different tests or indices of upper extremity function were used. Not only is there no agreement amongst rehabilitation professionals as to the appropriate rehabilitation interventions to deliver to improve function in the affected arm but there are just as many outcome evaluations used as there are studies. This makes the process of comparing the studies very difficult as no standard way of measuring upper extremity function exists.

Title Page

The effect of a task-oriented intervention on arm function in persons with stroke: a randomized controlled trial

Published in Clinical Rehabilitation 2006; 20:296-310.

Authors: Johanne Higgins, MSc, O.T., Nancy M. Salbach, MSc, P.T., Sharon Wood-Dauphinee, Ph.D., P.T., Carol L. Richards, Ph.D., P.T., Robert Côté, M.D., Nancy E. Mayo, Ph.D., P.T.

Posts of the Authors at the Time of the Study:

Higgins J: PhD candidate, School of Physical and Occupational Therapy, Faculty of Medicine, McGill University.

Salbach NM: PhD candidate, Department of Epidemiology and Biostatistics and Occupational Health, Faculty of Medicine, McGill University.

Currently: Post-doctoral fellow, Department of Physical Therapy, University of Toronto

Wood-Dauphinee S: Professor, School of Physical and Occupational Therapy, Faculty of Medicine, McGill University, and Department of Epidemiology and Biostatistics and Occupational Health, Faculty of Medicine, McGill University

Richards CL: Professor, Department of Rehabilitation, Faculty of Medicine, Laval University

Côté R: Department of Neurology, Montreal General Hospital, McGill University Health Centre

Mayo NE: Associate Professor, School of Physical and Occupational Therapy, Faculty of Medicine, McGill University, Associate member, Department of Epidemiology and Biostatistics and Occupational Health, Faculty of Medicine, McGill University

Work conducted in: Montreal, Quebec, Canada, and Quebec City, Quebec, Canada

Address correspondence to: Nancy Mayo, Royal Victoria Hospital, Division of Clinical Epidemiology, 687 Pine Avenue West, Ross 4.29, Montreal, Quebec, H3A 1A1 CANADA, Telephone: (514) 934-1934 ext. 36922, Fax: (514) 843-1493, e-mail: nancy.mayo@mcgill.ca

Abstract

Objective: To evaluate the efficacy of a task-oriented intervention in enhancing arm function in persons with stroke.

Design: Two-center, observer-blinded, stratified, block-randomized controlled trial.

Setting: General community.

Patients: Ninety-one individuals within one year of a first or recurrent stroke consented to participate between May, 2000 and February, 2003.

Interventions: The experimental intervention involved practice of functional, unilateral and bilateral tasks that were designed to improve gross and fine manual dexterity whereas the control intervention was composed of walking tasks. Members in both groups participated in 3 sessions a week for 6 weeks.

Main Outcome Measure(s): The primary test of arm function was the Box and Block Test. Secondary tests included the Nine-hole Peg Test, maximal grip strength, the Test d'Évaluation des Membres Supérieurs des Personnes Agées and the Stroke Rehabilitation Assessment of Movement.

Results: Results are for the more affected arm. Baseline performance on the Box and Block Test was an average of 26 blocks (standard deviation (SD)=16) in the experimental group (n=47) and 26 blocks (SD=18) in the control group (n=44). These values represent approximately 40% of age predicted values. Values for the postintervention evaluation were an average of 28 (SD=17) and 28 (SD=19) blocks for the experimental and control group respectively. No meaningful change on other measures of arm function was observed.

Conclusions: A task-oriented intervention did not improve voluntary movement or manual dexterity of the affected arm in persons with chronic stroke.

Key words: stroke, cerebrovascular accident, randomized controlled trial, arm, upper extremity.

Introduction

Stroke affects 15 million persons in the world each year and approximately one third will live with the sequalae of this disease.[91] Stroke commonly leads to paresis of an arm. As many as 85% of stroke survivors initially present with an impaired arm[6,92] and in most patients admitted with severe stroke, the more affected arm never becomes useful.[93] Because use of the arms is necessary for the performance of activities of daily living (ADL) and instrumental activities of daily living (IADL), this lack of recovery can be devastating to a person's health-related quality of life, particularly as more stroke survivors are returning to live at home.[7] The rehabilitation of the affected arm and hand remains a challenge. Although motor recovery of the arm has been shown to be similar to that of the lower extremity,[94] observed improvements are not necessarily translated into increased performance of daily activities as these tasks are more highly complex than functional activities of the lower limbs. Many different therapeutic approaches have been used in clinical settings to rehabilitate the affected arm. None of the studies comparing the effectiveness of these different approaches has consistently proven one approach to be superior to the other.[12-16,18,95,96] This early research led some authors to conclude that for patients with severe initial arm paresis, rehabilitation efforts should be geared more towards the teaching of compensatory techniques using the less affected arm.[93] More recently, a robotic intervention has demonstrated benefits in decreasing motor impairments in persons with chronic stroke but functional gains were small.[97] Also, sensorimotor stimulation was shown to improve motor recovery in the arm. Notably, this intervention was more effective in patients with a severe motor deficit. Unlike the previous study, this intervention was administered in the acute phase post-stroke,[98] when the most and the fastest recovery is known to take place. Jang and associates[99] demonstrated that a 4-week task-oriented training programme consisting of six arm tasks performed for 40 minutes/day, 4 days/week for 4 weeks lead to functional recovery in the chronic phase post-stroke. To be included in this study, however, subjects had to meet specific criteria regarding minimal residual movement and no severe spasticity or tremor in their affected arm.

Several randomized controlled trials have also been performed testing the effects of robot-assisted movement,[100] neuromuscular stimulation,[101-103] functional task practice, strength training,[104] and arm ability training.[71] Most of them included only a small number of subjects, decreasing their power to detect any real changes.

As well, a new and very promising treatment modality called 'Constraint-Induced Movement Intervention (CIMT)'[20] has received considerable attention over the past few years. Although most studies [21-23,25,105] evaluating the effects of CIMT or a modified version have demonstrated positive results,[15] only three randomized controlled trials [21,23,25] have been conducted. Furthermore, the intensity of this program in its original format (6 h of therapy per day for 14 days during 90% of waking hours while the less affected arm is constrained[106]) is probably beyond the stamina of the vast majority of stroke patients.

In studies with similar interventions to the present study such as the one by Kwakkel et al[107], which was a randomized controlled trial comparing treatment with emphasis on arm training versus leg training versus control programme, authors found that individuals in the arm rehabilitation training had a small but significant effect on the functional recovery of dexterity of the paretic arm. Subjects in this study were treated in the acute phase after their stroke, within fourteen days.

Dean and colleagues,[32] aimed at evaluating the effects of a training programme on the performance of locomotor tasks in chronic stroke (experimental group) and in which the control group received arm training for 1 h, three times a week for four weeks. They found no significant difference between the experimental and the control groups in grip strength or dexterity. The authors speculated that failure to observe improvement in arm function may be due to their small sample size, subject inclusion criteria, and the measures used. Indeed, the Perdue Pegboard is a very high level evaluation that may have demonstrated floor effects among subjects with lower ability level in their affected arm. The authors suggest the Test d'Evaluation des Membres supérieurs des Personnes Agées (TEMPA) may have been more appropriate and sensitive in detecting changes among persons with varying levels of arm function.

In their study, Duncan and co-workers[108] examined the effect of therapeutic exercises in subacute phase after stroke. The exercise programme aimed to improve strength, balance and endurance and also to encourage persons to use their affected arms more in their activities of daily life. Balance, endurance and mobility improved but improvement in arm activity performance was only observed in patients who entered the study with better arm performance.

The findings emerging from the mobility part of the current randomized controlled trial[109] support the efficacy of task oriented practice in enhancing functional walking capacity and walking speed within the first year following stroke.

The present study, therefore, focused on a less intensive task-oriented intervention (90-min sessions, three times a week for six weeks) that required a more realistic time commitment. The study included subjects in the chronic phase after stroke with no minimal criteria for movement in their affected arm and used measures of impairment as well as measures of activity limitation such as the TEMPA. The objective of the study was to evaluate the efficacy of task-oriented training in enhancing arm function post stroke. The hypothesis tested was that people who received a six-week programme of arm training would improve their arm function to a greater extent than people who received a walking intervention.

Materials and Methods

The results of the present study emerge from a randomized controlled trial designed to evaluate the efficacy of a task-oriented intervention in improving walking competency in people with stroke (walking group).[109] A parallel objective of the study was to evaluate the efficacy of task-oriented training in enhancing arm function post stroke (arm exercise group). The study was designed in parallel and it was powered accordingly, taking into account the two primary hypotheses related to the effect of task-oriented training on arm and walking function.

Subjects

A total of 91 subjects entered the trial between May 2000 and February 2003. Subjects were recruited from 9 hospitals and 2 rehabilitation centers in Montreal or Quebec City. Randomization within each site was separate. Patients included in the study met the following criteria: (1) clinical diagnosis of a first or recurrent stroke, (2) residual walking deficit, (3) a minimum score of 14 out of 22 on the Telephone Version of the Mini-Mental State Examination,[110] (4) ability to walk 10 m independently, with or without supervision or aid, (5) sufficient language ability to follow testing procedures (6) living in the community, (7) discharged from physical rehabilitation, and (8) less than 1 year post-stroke at the time of recruitment. Patients were excluded if they had: (1) neurological deficit related to metastatic disease, (2) recovery of functional walking capacity defined by age- and gender-specific norms[111] on the Six Minute Walk Test [112] (SMWT), (3) discharge to a long-term care facility, or (4) comorbid conditions that precluded participation in arm or walking training. All participants provided voluntary, written consent to take part in this study. This study was approved by the Institutional Review Board of McGill University as well as by the ethics committees of individual hospitals and rehabilitations centers from which patients were recruited.

Design

Forty-seven subjects were randomized to receive arm training and forty-four were randomized to receive mobility training. Subjects were stratified at baseline as having a mild, moderate or severe walking deficit based on their comfortable gait speed in order to proceed with a permuted block randomization to avoid imbalance in the number of subject randomized to each group. The allocation sequence was randomly ordered in block sizes of two and four within each stratum. People not involved in the study placed the treatment group allocations in envelopes and sealed them. The evaluator opened the appropriate envelope only once the baseline evaluation had been completed.
Evaluations were conducted at baseline and on completion of the intervention by trained evaluators. Over the course of this three-year study, a total of eight physical, occupational, or exercise therapists, served as evaluators. To be an evaluator, these therapists underwent a 3-h training session in the administration of the study measures

employed and were provided with an instruction manual. As well, review sessions were conducted every six months. When conducting postintrvention evaluations, raters were unaware of the subjects' group assignment. Participants were advised not to mention their group assignment to the evaluator.

Measurement - Measures of arm activity limitation (capacity)

The Box and Block Test (BBT).[49] A test of gross manual dexterity was used as the main outcome in the study. The measurement scale (quasi-continuous) is the number of blocks a subject can move, from one compartment of a box to another within 1 min. A normal value for people in the age group of the present sample is approximately 67 blocks.[50] The Box and Block Test has been shown to have good test-retest reliability,[49] and test performance correlated highly with performance on a similar test of dexterity.[50] Test-retest reliability and construct validity of this instrument in an elderly population with arm impairment has been demonstrated.[50] Furthermore, the Box and Block Test is a significant predictor of physical health as measured by the SF-36 (Medical Outcomes Study 36-Item Short from Questionnaire) with a difference of seven blocks associated with a difference in physical health of two units; for a difference of five units on the Physical Component Summary score of the SF-36, the corresponding clinically meaningful difference was 17.5 blocks.[113]

The Nine-Hole Peg Test (NHPT).[59] The Nine-Hole Peg Test was used to measure fine manual dexterity. The time for a subject to place nine dowels into nine holes on a board, and remove them is recorded. High interrater reliability and moderate test-retest reliability have been demonstrated and norms for adults up to 75 years of age and above for both genders were established.[60] The time to complete the test was recoded into four ordered categories defined by the number of SD units they were from age-and gender-specific norms. Scores within 1 SD of the normal value were assigned a value of 3, the value decreasing by 1 point for each additional SD away from the normal value. Scores that were 4 SD's or more away from the normal value were assigned a value of 0.

The TEMPA (Test d'Evaluation du Membre Supérieur des Personnes Agées). This test was developed by Desrosiers and co-workers[68] to evaluate activity performance of the arms in individuals over the age of 60. It contains four unilateral and five bilateral functional tasks. Normative data have been published for this population.[70] Both speed of execution and quality of the movement were analyzed for this study. Scores for the timed tasks of the TEMPA were recoded into four ordered categories, defined by the number of SD units they were from age- and gender-specific norms.[70] Scores within 1 SD of the normal value were assigned a value of 3, the value decreasing by 1 point for each additional SD away from the normal value. Scores that were 4 SDs or more away from the normal value were assigned a value of 0. Scores on the Functional Rating scale of the TEMPA that reflect movement quality were transformed into an ordinal scale from 0 to 3, with 0 representing lowest quality.

Measures of arm impairment

Grip Strength. Three grip strength measures of each hand were taken using the Jamar™ dynamometer (Sammons Preston Rolyan, Bolingbrook, IL, USA) with standardized positioning and instruction.[57] The highest score was retained. The measurement scale (kilograms of force) is continuous. Good inter-rater and test-retest reliability have been observed using these procedures.[58]

The Upper Extremity Subscale of The STroke REhabilitation Assessment of Movement (STREAM).[114] The STroke REhabilitation Assessment of Movement consists of 30 items, equally divided into three sections: voluntary movement of the arm, voluntary movement of the lower extremity, and basic mobility. Only the arm subscale was used in this investigation. The total score was transformed to a percentage, making it a quasi-continuous scale. A study by Daley and colleagues reported content validity and excellent inter-rater and intrarater reliability.[115]

Measurement – Indices of arm activity limitation (capacity)

The Barthel Index.[72] This is a weighted scale that assesses performance in self-care and mobility. Only responses for items requiring use of the arm (feeding, personal hygiene, bathing and dressing/undressing) were analyzed. Items are scored on an ordinal scale. Two items are scored out of 5 and two are scored out of 10 resulting in a maximum summative score of 30.

The Older Americans Resources and Services Scale-Instrumental Activities of Daily Living (OARS-IADL)[75] Each item is scored on an ordinal scale from 0 to 2, a higher score indicating a higher level of performance. Only responses to two items relating to the use of the arm were analyzed (meal preparation and housework).

The Medical Outcomes Study 36-Item Short form Questionnaire (SF-36).[48] This is a commonly used health-related quality of life measure. The two arm-related questions that were analyzed (grocery carrying and bathing/dressing) are scored on an ordinal scale from 1 to 3 where higher scores indicate better functioning.

The Geriatric Depression Scale[116] was used to classify individuals as having no (0-9 points), mild (10-19 points) or severe (20-30 points) depressive symptoms. Socio-demographic and clinical information was obtained from the medical chart.

Interventions

Arm Intervention

Subjects in both groups participated in 18 practice sessions three times a week for six weeks and were supervised by either a licensed physical or occupational therapist. The intervention took place in a research area within a rehabilitation setting. Therapy was administered to a subject on a one to one basis with the therapist. Each session lasted approximately 90 min. At the start of the intervention, subjects were asked to identify daily activities that were difficult to perform and that they wanted to improve. Providing patients had sufficient movement in their more affected arm to attempt the functional tasks, they were practiced. Examples of such tasks included manipulating playing cards, clothes pins as well as writing exercises. For three subjects who did not have sufficient movement in their more affected arm to practice such tasks, the therapist assisted the person by guiding the limb through the tasks while applying other modalities such as vibration and passive range of movement to facilitate mobility and decrease spasticity. When subjects had maximized their performance, tasks were changed or their level of difficulty was heightened at the discretion of the therapist. Both the duration and level of difficulty achieved in each task were recorded at every session. All subjects were given a home program to be done for a minimum of 15 min per day for the period of the intervention. The home program consisted mainly of similar tasks to those practiced during the intervention. Most of the therapy material was common objects found in most homes.

Walking Intervention

The walking intervention consisted of 10 functional tasks designed to strengthen the lower extremities and enhance walking balance, speed as well as distance.[109]

Statistical methods

Data were analyzed on the basis of intention to treat. In the primary analysis, the chi-square test was used to compare between groups the proportion of subjects who deteriorated, remained the same, improved between one and six blocks, or improved more than six blocks on the Box and Block Test. The effect of arm training on the remaining tests of arm impairment, activity limitation and performance was also evaluated. Group comparisons were made using a *t*-test for independent samples with associated 95% confidence intervals (CIs) for variables measured on a continuous scale and the Wilcoxon Rank Sum Test for variables measured on an ordinal scale. T-tests and Wilcoxon rank sum test were also done on the change scores between groups. Transformations of scores on the Nine-Hole Peg Test and TEMPA were performed as these tests are scored on an ordinal scale (TEMPA) or on a continuous scale without a natural zero (TEMPA, Nine-Hole Peg Test). Scores for these tests were transformed into four ordered categories defined by the number of standard deviations away from the normal values (see Measurement Section).

Multiple linear regression was then used to identify and adjust for prognostic variables to enhance the accuracy of estimation of the arm training effect on change in Box and Block Test scores. With the indicator variable for group in the model, the effect of adjusting for age, sex, level of depressive symptoms, hand dominance, previous stroke, number of comorbid conditions and type of stroke on change in Box and Block Test as the outcome or 'y' variable was examined. Because the analysis carried out on the mobility outcomes revealed an interaction between treatment group and level of depressive symptoms,[117] we also examined this interaction in our data set.

Lastly, the effect size of the present study, based on the Box and Block Test was calculated by dividing the mean difference in the change score between the experimental and control groups by the standard deviation of the initial score of the control group.

Sample size

As this study had two hypotheses, one related to the effect of walking competency training where the primary outcome was the Six-Minute Walk Test and a second related to the effect of upper extremity training where the primary outcome was the Box and

Block Test, two sample size estimates were required. For the first hypothesis related to the waking intervention, the sample size calculation was based on the detection of a group difference of 28 m in average change in Six-Minute Walk Test performance (Type I error = 0.05, Type II error = 0.10, expected drop-out rate of 10%). This calculation emerges from results of a pilot trial of a similar intervention (group difference on Six Minute Walk Test = 37 m, SD = 30.4).[32] It was calculated that 60 persons were required.[109] For the second hypothesis, related to the subject of this article, the effect of the upper extremity intervention, 60 subjects would yield 90% power to detect a between-group clinically meaningful difference of 17.5 blocks using the Box and Block Test, assuming within-group standard deviation of 20.[118] To account for drop outs and the simultaneous testing of two hypotheses, we targeted 90 subjects for this study.

Results

Description of the Study Population

As previously noted,[109] 344 of the 1056 patients assessed for eligibility met the eligibility criteria. A total of 91 subjects agreed to participate and provided written, informed consent (85 in Montreal and 6 in Quebec City). Forty-seven subjects were randomized to the arm intervention group and 44 to the walking group. Both groups were similar in terms of their baseline characteristics. They are presented in Table 1. Out of the 47 participants in the arm intervention group, three people were missing postintervention evaluations due to illness. In the mobility group, two subjects withdrew due to pain or unwillingness to travel.

Subject Compliance

In the arm group, 72% (34) of the subjects attended 17 or 18 treatment sessions, four (9%) discontinued the treatment (Figure 1), four (9%) attended between 10 and 16 sessions and five (11%) were given a home programme. In the walking group, 86% (38) participated in 17 or 18 treatment sessions, three (7%) discontinued the intervention (Figure 1) and three (7%) attended between 12 and 14 sessions.

Figure 1 presents the flow of participants through the study. Postintervention data were missing for three people in the arm group due to illness (n=3) and for two people (n=2) in the walking group (one person unwilling to travel and one person experienced onset of groin pain preventing participation in therapy). One person was unable to complete the SF-36 at the post-intervention evaluation due to aphasia. One person was missing grip strength measurements at the postintervention evaluation. The method of last value carried forward was used to replace these missing data as it was thought that an illness or any other reasons why the data were missing would not have a direct impact on their arm function. In a clinical setting such as this one, when recovery is expected, this method can be regarded as conservative.[119] The TEMPA, the OARS-IADL as well the Barthel Index, were missing for six subjects. Values for these participants were estimated from baseline and postintervention scores on the Box and Block Test using simple linear regression. This same method was use for baseline grip strength measurements which were missing for one person.

Measures of arm activity limitation (capacity)

Table 2 presents the proportion of subjects in each group who, on the Box and Block Test, deteriorated, remained the same, improved between one and six blocks, or improved more than six blocks. There were no differences between the two groups on the proportions of persons in these categories (χ^2 =3df = 0.9, p=0.818).

Table 3 presents the scores on measures of activity limitation and impairment at baseline and postintervention for both treatment groups. Members of the arm group improved their score on the Box and Block Test by an average of one block more than members of the mobility group. This improvement is not significant and is not clinically relevant. Little or no change in scores was observed on the Nine-Hole Peg Test and the TEMPA in each group between the initial and postintervention evaluations.

Measures of arm impairment

Members of the arm intervention group improved their grip strength by an average of 0.5 kg more than the mobility group members. Again, this improvement was not significant and is not considered clinically important. Members of both study groups improved by an average of 3 points on the arm subscale of the The Stroke REhabilitation Assessment of Movement, resulting in a between-group difference of zero.

Indices of arm activity limitation (capacity)

Table 4 presents the proportion of members in each of the intervention groups who improved on each of the indices of arm activity limitation (capacity). Improvement was defined as having gained at least one point on an index. Again, there were no differences between the two intervention groups on the proportions of persons who improved for any of the questions.

Change in indices of performance

To explore factors contributing to change in arm function, we carried out a multiple linear regression with change in Box and Block Test as the outcome or 'y' variable and age, sex, depression, hand dominance, previous stroke, number of comorbid conditions and type of stroke as the predictors. The effect of these variables on outcome was examined one at a time with the indicator variable for group in the model.

In the multivariable analysis, none of the potential predictor variables, one at a time with group, was significantly associated with change in Box and Block Test so no further multivariate modeling was carried out. The interaction term with depression and group was also non-significant.

Lastly, the effect size for the present study, calculated using scores on the Box and Block Test is 0.06 (Figure 2).

Discussion

People assigned to receive a six-week programme of arm training did not improve their arm function to a greater extent than people assigned to receive walking training. The small differences observed on measures of arm impairment, activity limitation and performance in the arm training group between the baseline and the postintervention evaluations were not statistically significant or clinically meaningful. On the Box and Block Test, for example, an average gain of three blocks was observed in members of the arm group, but an improvement of at least seven blocks is necessary to translate to improve daily physical functioning. [120] The change on other measures of activity limitation, such as the Nine-Hole Peg Test and the Test d'Evaluation du Membre Supérieur des Personnes Agées (TEMPA), was also clinically unimportant for both treatment groups. Although data on the minimal clinically important change are unavailable for the Nine-Hole Peg Test or the TEMPA, a change of less than 1 point out of a possible 4 did not appear clinically relevant. Finally, changes of 1 and 0 points out of a possible 27 on the functional rating scale of the TEMPA were not meaningful to improve performance on activities of daily living. The changes observed on the measures of arm impairment as measured by grip strength were negligible. On the The STroke REhabilitation Assessment of Movement the change was 3 percentage points for both groups, which was not considered clinically meaningful. The three indices of arm performance did not reveal any statistically significant differences between the two treatment groups (Table 4). The tendency for people in the walking group to have improved their arm performance, especially in more integrated tasks that require use of both the arm and the lower extremity (eg. carrying groceries meal preparation and housework), may have masked a beneficial effect of task-oriented arm training. These results also support the task-specific effect of the walking and carrying task practiced in the walking intervention and the efficiency of training the arm and lower extremity simultaneously to improve specific functional activities.

In 2001, van der Lee [121] reviewed several randomized controlled trials aimed at evaluating the effects of rehabilitation; more specifically exercise therapy, for arm function. Although positive results were reported for six trials, the amount of therapy offered to the intervention and control groups sometimes differed and thus the observed results are inconclusive. In 2001, van der Lee et al. [122] reviewed the evidence from three randomized controlled trials that used this therapeutic approach and found that the evidence was inconclusive and that positive results may have been attributable to an increased amount of therapy on the more affected arm than the effects the constraint of the less affected arm. When the possibility of decreasing the amount of therapy given in a standard CIMT protocol from 6 h to 3 h a day was explored,[123] researchers found increased arm performance in both groups but the treatment effect was greater for the 6 h per day group. It is important to note that even by decreasing the number of hours from 6 to 3, the total time of treatment is still almost double the one in the present study.

Figure 2 presents effect sizes relative to the sample size of several randomized controlled trials aimed at evaluating different rehabilitation techniques to improve arm function post-stroke.[25,32,71,98,100-104,107,108] Effect sizes were calculated from results of postintervention evaluations just as in the present study. Some of these studies will be discussed below. In one of the studies using training as the intervention under investigation,[104] 20 h of additional training of the arm were offered to the experimental group. The latter improved more than the standard care group but the long-term benefits were significantly greater among persons with mild arm deficits. In another study[32] in which the treatments were offered three times a week for four weeks (similar to the present study), researchers found, on average, no significant difference between the experimental and control groups in grip strength or dexterity at the post-training or follow-up evaluation. A third study[108] examined the effect of therapeutic exercises in sub-acute stroke. The intervention consisted of 36 intervention sessions of 90 min each over a 12-to 14-week period. An improvement in arm performance was only observed in subjects mildly affected arm function upon entry into the study. The results of this study do not appear to agree with those of the author's earlier study in which it was found that the arm and lower extremity had similar recoveries.[13] Looking at Figure 2, it is apparent that studies with the smallest sample sizes are the ones that demonstrated largest effect size. This may be an indication that in these particular studies, patients selected to participate were the ones with at least minimal movement in their arm and hand which is known to be a good precursor for improvement in arm performance.

In present study, the intervention was limited to 90-min sessions, three times per week, for six weeks, not including the home exercise programme of 15 min a day. Also, we did not select patients based on their arm function at baseline and thus subjects with a wide range of arm dysfunction were included, over 16% (15) of patients were unable to move a single block using their more affected arm at the start of the study and none of these patients were able to move a single block at the postintervention evaluation. Furthermore, 30 subjects (68%) in the walking group had their dominant arm affected versus 25 subjects (53%) in the arm training group. Although a statistically significant difference was not detected between the two groups, this difference may have contributed to people in the walking training group using their affected arm more in everyday life and thus improving their performance despite not receiving therapy as part of the study. The noteworthy findings showing a tendency for the mobility group to have improved more on the indices of arm performance (OARS-IADL and SF-36) may also indicate that these subjects were using their arm in some of the tasks (holding on the railing for step-ups and for treadmill walking, carrying and walking) and this may have partially contributed to the gains in arm function observed.

Future research should include more intensive intervention that lasts for a longer period of time that also includes bilateral meaningful activities that integrate both the upper and lower extremities. Treatment should focus on the performance of activities of daily life and based on integrated tasks requiring the simultaneous use of both arm and lower extremity, including use meaningful objects as this has been shown to have a positive effect on the performance of tasks.[124] A major challenge still remaining is the treatment of those individuals whose arm is severely affected and cannot participate in task-oriented activities. There may be a need for the development of a different treatment strategy for individuals who have very little or no movement in their arm and the limitation of task oriented programs to individual who have some arm movement and dexterity at baseline.

Limitations

One of the limitations of this study is that we did not stratify on level of arm deficit. It was also difficult to challenge and motivate patients for whom no active movement of the affected arm was present. For these patients, vibration and assisted movements were clearly insufficient to improve function over the relatively short period of the study intervention. A specific intervention, tailored to their needs would have been required. Although the outcome measures used spanned a wide range of ability levels from impairment to activity performance, the sensitivity to change of the TEMPA, situated at the higher end of the ability scale, has not yet been demonstrated in a chronic stroke population. It may not have been useful for detecting small changes at the level of fine manual dexterity and higher level performance tasks. Questionnaires, such as the Motor Activity Log (MAL), on the amount of use of the affected arm in every day life may have been useful to detect changes in behaviours not always associated with a large improvement in motor ability.

Conclusion

The task oriented intervention did not improve arm function. In fact, although not statistically significant, greater improvement on the indices of arm performance in the mobility group seems to indicate that the performance of integrated functional tasks may be more beneficial. This study also indicates that for persons with no initial movement in their arm, very little gain is to be expected with therapies now in use.

Clinical Message

- An improvement in Upper Extremity function was not observed after a task-oriented intervention in the first year post-stroke
- An intervention geared to those who present with a severely affected arm is required
- Sensitive outcome measures are needed in order to detect small changes occurring at the level of arm performance.
-

Acknowledgments: Operating funds were provided by the Quebec *Réseau provincial de recherche en adaptation-réadaptation*, the Heart and Stroke Foundation of Canada, and the Canadian Stroke Network. Johanne Higgins was supported by a fellowship from the Canadian Stroke Network, McGill University Research Institute and the Fonds de Recherche en Santé du Québec. Nancy M. Salbach received a fellowship from the Canadian Institutes of Health Research.

Table 2.1 Demographic and Clinical Characteristics of Study Participants.

Subject Characteristic	Arm group (n = 47)	Walking group (n = 44)
Age Mean (SD)	73 (8)	71 (12)
Gender No. (%) male	30 (64)	26 (59)
Number of Comorbid conditions No. (%)		
0	3 (6)	2 (5)
1-2	19 (40)	13 (30)
3-4	18 (38)	17 (39)
>4	7 (15)	12 (27)
Type of stroke No. (%)		
Ischemic	36 (77)	40 (91)
Hemorrhagic	11 (23)	4 (9)
Number of stroke No. (%)		
1	41 (87)	39 (89)
2	5 (11)	5 (11)
4	1 (2)	0
Side of Hemiplegia No. (%)		
Right	22 (47)	17 (39)
Left	24 (51)	27 (61)
Bilateral	1 (2)	
Dominant arm affected No. (%)	25 (53)	30 (68)
Days post-stroke at first evaluation Mean (SD)	217 (73)	239 (83)

Abbreviations: SD, Standard Deviation; stroke, Cerebrovascular Accident

Figure 2.1 Flow of subjects through the trial

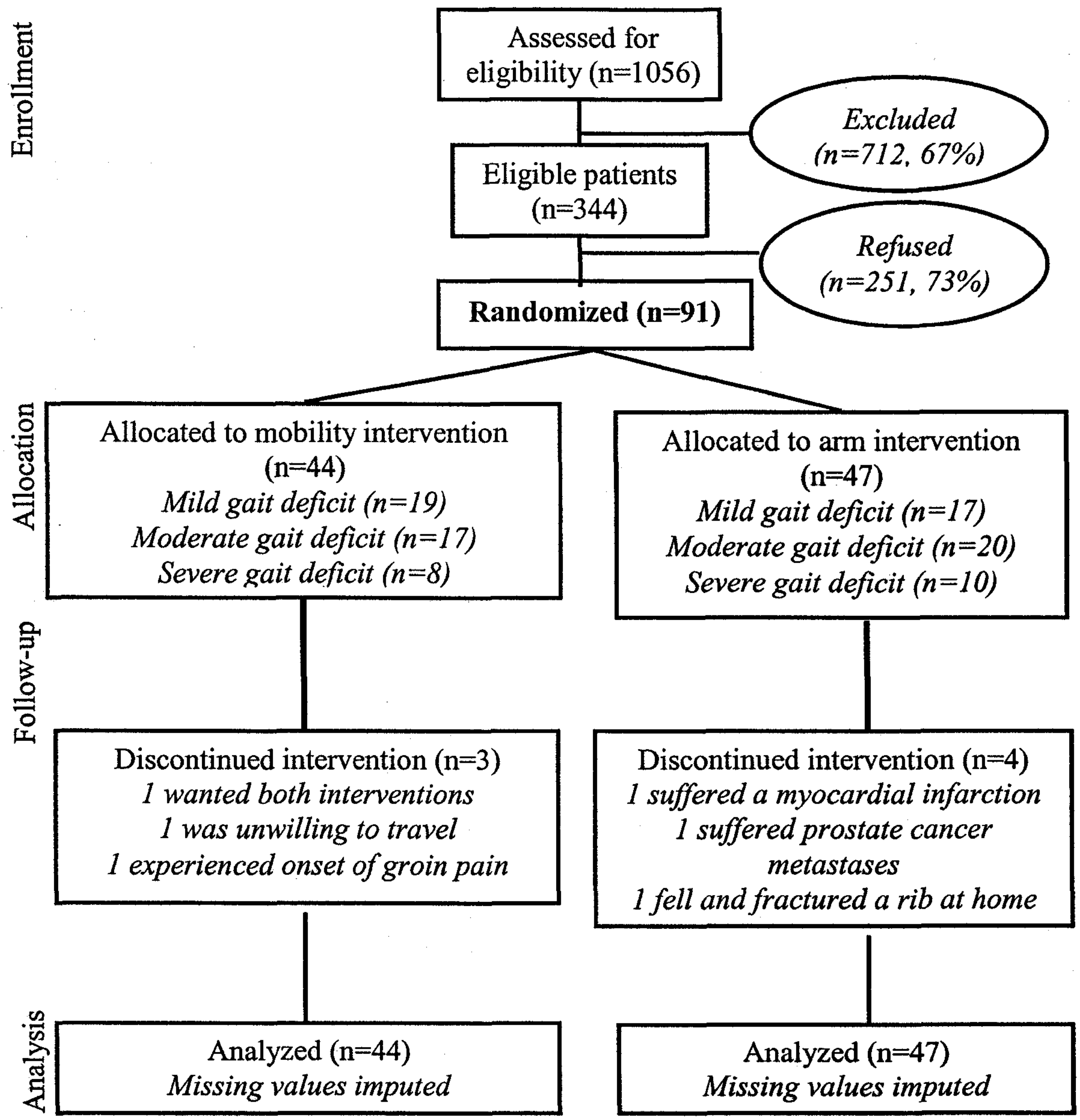


Table 2.2 Change in score on the Primary Outcome Measure: the Box and Block Test.

Change in Box and Block Mean Change (#blocks)	arm group (n=47) No (%)*	Walking group (n=44) No (%)*
<0	11 (24)	10 (23)
0	10 (21)	12 (27)
1-6	14 (30)	14 (32)
7-15	12 (26)	8 (18)

*$\chi^2$3 df =0.9; p=0.818

Abbreviations: df, degrees of freedom.

Table 2.3 Scores on Measures of Arm Impairment and Activity Limitation.

Measures		Arm training (n=47)			Walking training (n=44)			Group difference †(95% CI) *Wilcoxon Rank sum test
Activity Limitation		**Mean**	***SD***	***Median (Quartiles)‡***	**Mean**	***SD***	***Median (Quartiles)‡***	
Box and Block Test (# blocks)	*Pre*	26	16	31 (9, 38)	26	18	30 (6, 40)	
	Post	29	17	34 (16, 43)	28	19	32 (5, 41)	
	Change	3	5	3 (0, 7)	2	5	1 (0, 5)	1 (-1-3)
Nine-Hole Peg Test /3	*Pre*	1	1	1 (0, 1)	1	1	1 (0, 1)	
	Post	1	1	1 (0, 1)	1	1	1 (0, 1)	
	Change	0	0	0 (0, 0)	0	0	0 (0, 0)	p=0.6*
TEMPA /27 (Timed tasks)	*Pre*	9	3	9 (7, 11)	10	4	10 (7, 14)	
	Post	9	3	9 (7, 11)	10	4	10 (7, 12)	
	Change	0	2	-1 (-1, 3)	0	2	0 (-1, 0)	p=0.1*
TEMPA /27 (FR)	*Pre*	19	6	19	19	7	21	

				(13, 25)			(12, 26)	
	Post	20	7	22 (14, 26)	19	7	20 (12, 27)	
	Change	1	3	1 (-1, 3)	0	3	0 (-1, 1)	p=0.2*
Impairment								
Grip Strength (Kg)	*Pre*	16	10	16 (6, 24)	17	12	17 (8, 25)	
	Post	17	11	17 (8, 24)	18	12	18 (10, 26)	
	Change	1	5	1 (0, 4)	1	4	0	0.5 (1.3-2.4)
STREAM (arm subscale /100)	*Pre*	74	30	85 (60, 100)	71	34	88 (58, 100)	
	Post	76	30	90 (60, 100)	74	34	95 (50, 100)	
	Change	3	7	0 (0, 10)	3	9	0 (0, 5)	p=0.9*

Abbreviations: CI, Confidence Interval; SD, Standard Deviation; IQR, Interquartile Ranges; BI, Barthel Index; TEMPA, Test d'Evaluation du Membre Supérieur pour Personnes Agées, Test d'Evaluation du Membre Supérieur pour Personnes Agées; FR, Functional Rating; STREAM, STroke REhabilitation Assessment of Movement.

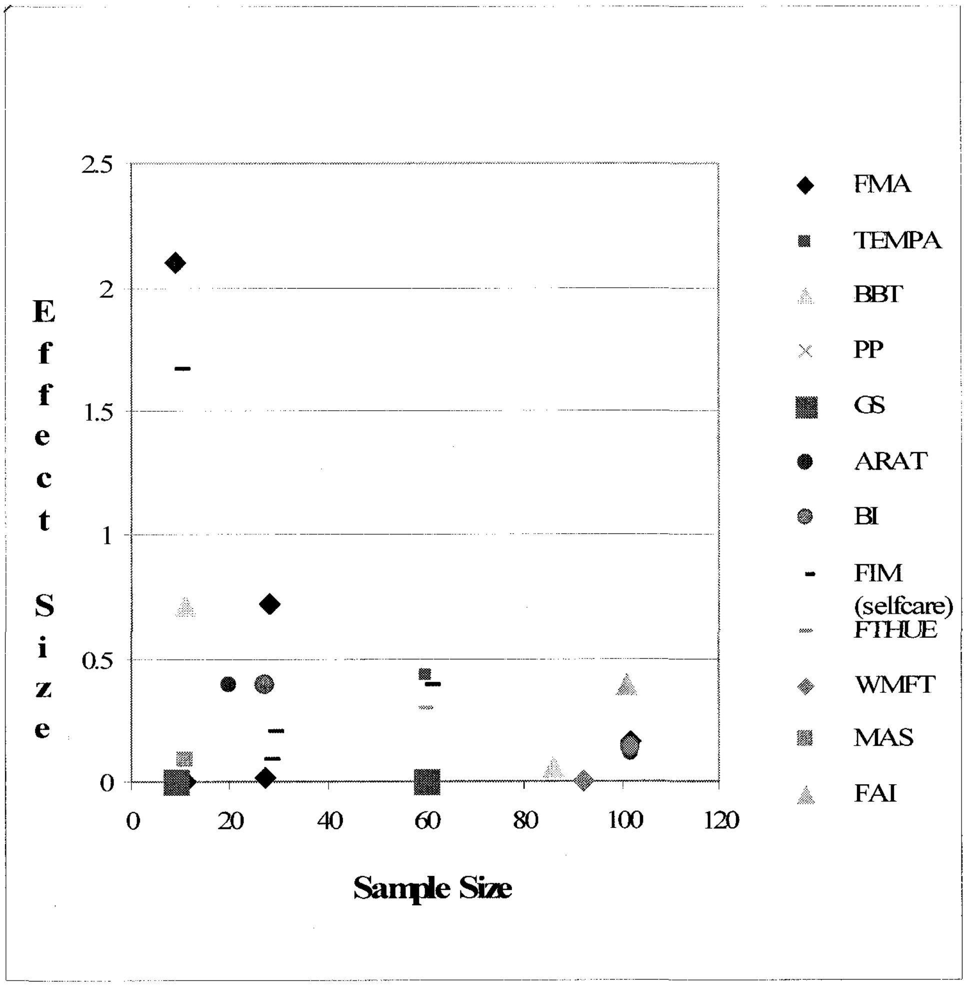
‡ 25th, 75th percentiles.

Table 2.4 Comparison of the 2 groups on Proportions of Persons who Improved on Indices of Arm Activity Limitation (capacity).

Indices of arm activity limitation		Arm group n = 47 No. (%)	Walking group n = 44 No. (%)	$\chi^2$3 df (p)
OARS IADL	**Meal preparation**	9 (19)	10 (23)	0.2 (0.675)
	Housework	6 (13)	10 (23)	1.6 (0.212)
SF-36	**Grocery carrying**	10 (21)	12 (27)	0.4 (0.504)
	Bathing/dressing	10 (21)	16 (36)	2.5 (0.111)
Barthel Index	**Feeding**	7 (15)	11 (25)	1.5 (0.226)
	Personal Hygiene	6 (13)	6 (14)	0.0 (0.902)
	Bathing	3 (6)	3 (7)	0.0 (0.933)
	Dressing/Undressing	10 (21)	7 (16)	0.4 (0.512)

Abbreviations: OARS-IADL, Older American Resources and Services Scale-Instrumental Activities of Daily Living; SF-36, Medical Outcomes Study 36-Item Short-Form Health Survey; df, degrees of freedom.

Figure 2 2 Effect sizes in relation to samples sizes of 12 Randomized Controlled Trials



Abbreviations: BBT, Box and Block Test;[103*] PP, Perdue Pegboard; [32] GS, Grip Strength;[32,104] TEMPA, Test d'Evaluation du Membre Supérieur pour Personnes Agées; [71] FMA, Fugl-Meyer Assessment; [104] ARAT, Action Research Arm Test; [98,100-103,108] BI, Barthel Index; [25,98] FIM, Functional Independence Measure; [100,104] FTHUE, Functional Test of the Hemiparetic Upper Extremity;[104] WMFT, Wolf Motor Function Test;[108] MAS, Motor Assessment Scale;[103] FAI, Frenchay Activities Index. [107] *Present Study

CHAPTER 3. The Science of Measurement

"If a yardstick measured differently because of the fact that it was a rug, a picture, or a piece of paper that was being measured, then to that extent the trust worthiness of the yardstick, as a measuring device, would be impaired." (Thurstone)

A large number of tests and indices are available to evaluate upper extremity function post-stroke. None of them, however, are considered to be a 'gold standard' among rehabilitation professionals or researchers. The importance of using adequate tests and indices that capture change and are psychometrically sound is undisputable in the search for better treatment strategies for the upper extremity of the increasing number of persons who are surviving a stroke and thus living longer with permanent disability and dependency.

In the field of rehabilitation, a lot of the variables we want to measure in individuals cannot be measured directly; these are usually assessed by observing behaviours that are related to the variables of interest using a standardized set of items scored on an ordinal scale.

Nunnaly[125] has defined measurement as: *"rules for assigning numbers to objects in such a way as to represent quantities of attributes"*. Sometimes these rules are self-explanatory and straightforward such as the use of a yardstick to measure length. One does not need to explain in what context or which yardstick was used by which person to measure the height of a table. One trusts and knows exactly what a centimeter means. It is a standard unit and does not vary according to the object being measured or the person taking the measurement. When measuring psychological attributes or personal traits however, this simplicity does not exist. To evaluate such attributes, tests have to be administered. Hence these attributes are called 'latent traits'. In order to 'measure' upper extremity function, tests or indices have to be administered and, through observation, an attempt is made at quantifying this attribute or trait. According to Wade[126] to measure is to quantify something by comparison with a standard unit. In neurological rehabilitation however, standard units do not yet exist.[126] In others words, there is no 'ruler' to measure upper extremity function.

3.1 Classical Test Theory (CTT)

Most of the tests and indices used in rehabilitation, as exemplified by those listed above and used in the randomized controlled trial presented in the first paper of this thesis, have an important common feature: they were developed using Classical Test Theory or CTT. This widespread theory has been used extensively over the past decades. Procedures were developed at the beginning of the 20th century by Spearman, among others.[127]

CTT defines a person's observed score as being equal to the true score minus the error.

OBSERVED SCORE = TRUE SCORE + ERROR

In this equation the true score represents an average of observed score if the person took parallel versions of a test many times and the error component is defined by the difference between the true and observed scores. Important features of this theory include the fact that the true score is 'test specific' i.e. that the score a particular examinee obtains depends on the particular test, and, thus, on the particular items that comprise that test. Other important characteristics of classical test theory concern the error term. The standard error of measurement applies to all scores obtained in a particular sample of examinees and is identical for all examinees.[128] Also, comparisons between individuals must be based on norms. Despite its predominance in the development of outcome measures in the field of rehabilitation, CTT has shortcomings:

1) The difficulty of the items is dependent on the ability of the group of persons to whom the items were administered. The characteristics/ability of the examinees or respondents and the characteristics of the test (the difficulty of the items comprising the test) cannot be separated. The reason for this is that the difficulty of an item is defined as the number of persons who succeeded on the item while the ability of a person is defined as the number of items that person successfully answered or performed. So, *"whether an item is hard or easy depends on the ability of the examinees being measured, and the ability of the examinees depends on whether the test items are hard or easy!"*.[129]

2) An index developed using this theory is composed of a finite set of items and the total test score is a simple summation of the score obtained on each of the items. A drawback of this approach is that because each item contributes the same amount to the total, it is not possible to compare persons between or within each other unless they were administered the same exact test or index. Furthermore, the interpretation of the test score has to be based on norms, so a comparison group is necessary.
3) As the total scores are a sum of the scores obtained on individual items that make up the index, the interpretation is flawed. The true distance between the different items and between the response options for each item is unknown and thus ratio characteristics that permit true comparisons between scores are lacking.

These shortcomings make the measurement of upper extremity function very challenging. If it is difficult to compare between persons or within the same person over time; it is also problematic to quantify change, to assess, for example, the difference in effectiveness between two treatments. Indeed the analysis of change scores along with the interpretation of the results are difficult using ordinal variables because we cannot assume equal changes have the same meaning.[130]

3.2 Psychometric properties in CTT

When reporting the psychometric properties of a test or evaluation, reliability and validity are most widely used.

3.2.1 Reliability

Under Classical Test Theory, the reliability of a test is used to estimate the amount of error, random or systematic, occurring in the measurement process. Under this theory, a longer test is more reliable than a shorter test.[131] Indeed, coefficient alpha, widely used to indicate internal consistency, is a function of the number of items and their degree of inter-correlation. When items within an index are highly correlated, the reliability of the index is high. It can be shown that the correlation between parallel test forms is equivalent to the reliability of a test.

$$Reliability = \frac{Subject\ variability}{Subject\ Variability + Measurement\ error}$$

The Standard Error of Measurement or SEM is a function of the standard deviation and the reliability is can be expressed as: SEM = $\sigma\sqrt{1-R}$.[131] The Standard Error of Measurement is constant across all examinees. There are many different types of reliability indices: test re-test, inter-rater, and responsiveness.

3.2.2 Validity

Classically, validity is defined as the *"extent to which an instrument measures what it is intended to measure"*.[132] As is the case for reliability, there are numerous types of measurement of validity. Sometimes validity of a particular index is established by correlating it with an existing index of known validity. Other types of validity require the administration of the index to known groups of patients. There are different types of strategies used, each particular to the type of validity sought. According to Messick,[132] however, there is only one type of validity that comprises all of the above subclasses. This author states that criterion and content validity are part of construct validity. The usefulness of distinguishing between the different types of validity comes into play when making inferences from scores. In educational and psychological measurement, validity is typically divided into six facets: content, substantive, structural, generalizability, external, and consequential aspects of construct validity[133] and performance evaluations should be evaluated using these same validity criteria.[134] Each of these facets has different sources of evidence to substantiate the extent of the validity. Different authors use different terminology when describing validity. The following table is derived from the work of Paul Stratford as presented in Chapter 4 of Physical Rehabilitation Outcome Measures.[135]

Table 3.1 Different Types of Validity.[135]

Definitions of Different Types of Validity
Face -appears to measure the construct
Content - evidence that the test or index is composed of a comprehensive sample of items that completely assess the domain of interest
Criterion Validity - extent to which the test or index provides results consistent with those obtained with a 'gold standard' test *Concurrent* - at approximately the same point in time *Predictive* – predicts a subsequent criterion event – something known to result from the attribute or construct being measured
Construct - the extent to which the test or index provide results that are consistent with theoretically driven relationships *Cross-sectional* – at same point in time *Convergent* – the extent to which the results of a test or index agree with the results of another test or index that is believed to be assessing the same attribute or construct *Known-groups* – extent to which the test or index differs across groups known to represent different levels of the attribute or construct of interest *Discriminant* - the extent to which a test or index correlates with attributes or constructs that are different from the ones intended
Longitudinal validity - ability of a test or index to detect change over time *Sensitivity to change* - the ability of a test or index to measure change in the attribute or construct regardless of whether it is relevant or meaningful, a necessary but insufficient condition for responsiveness.[27] *Responsiveness* - the ability of a test or index to measure a meaningful or clinically important change in the attribute or construct.[27]

3.3 Scores vs. Measures

Most of the observations gathered from the tests or indices described above result in ordinal data. When the scores from different items comprised in the instrument are added together, the result is a 'raw score'. They are counts of observed events, but NOT measures. Counts are regularly mistaken for measures although they are ordinal scales that do not possess an additive structure.

The example of the SF-36 scale, a well known index of health-related quality of life commonly used in rehabilitation, is used to illustrate this point.

In general, would you say your health is: (circle one)

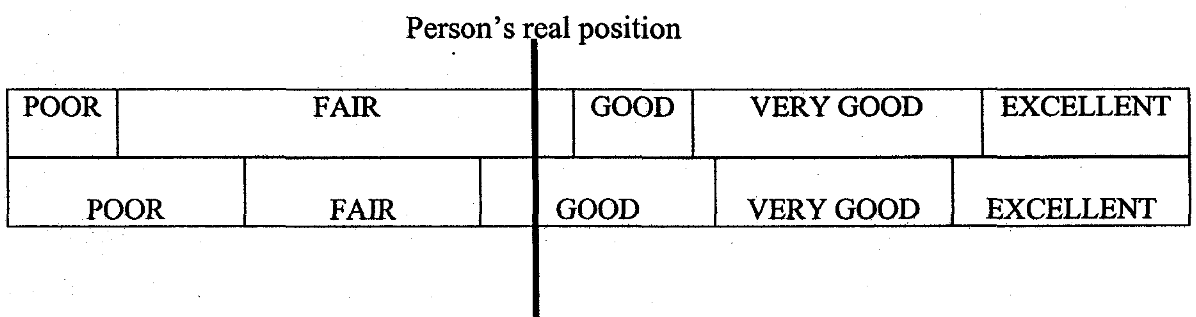
Excellent 1
Very good 2
Good 3
Fair 4
Poor 5

In this particular example, a numeral and NOT a number is assigned to each of the statements. Unfortunately these numerals are customarily treated as if they were numbers. That is, as if the statements were equally distanced from each other just as the distance between the numbers one and two is equal to the distance between the numbers two and three and so forth. Below is a graphical representation of these presumed equal distances.

POOR	FAIR	GOOD	VERY GOOD	EXCELLENT

In reality, the distance between excellent and very good may not be the same as the distance between very good and good. Shown below is a hypothetical representation of how distances between each of the descriptors may be perceived by the respondents.

Figure 3.1. Ordinal Scale with Unknown Distances.



This is an ordinal scale that does not possess an additive structure, yet, the responses to each of the questions are routinely added together to get a 'score'. The rehabilitation literature abounds with scales such as this one where numerals are treated as if they were numbers and ordinal scales are treated as if they were interval scales.

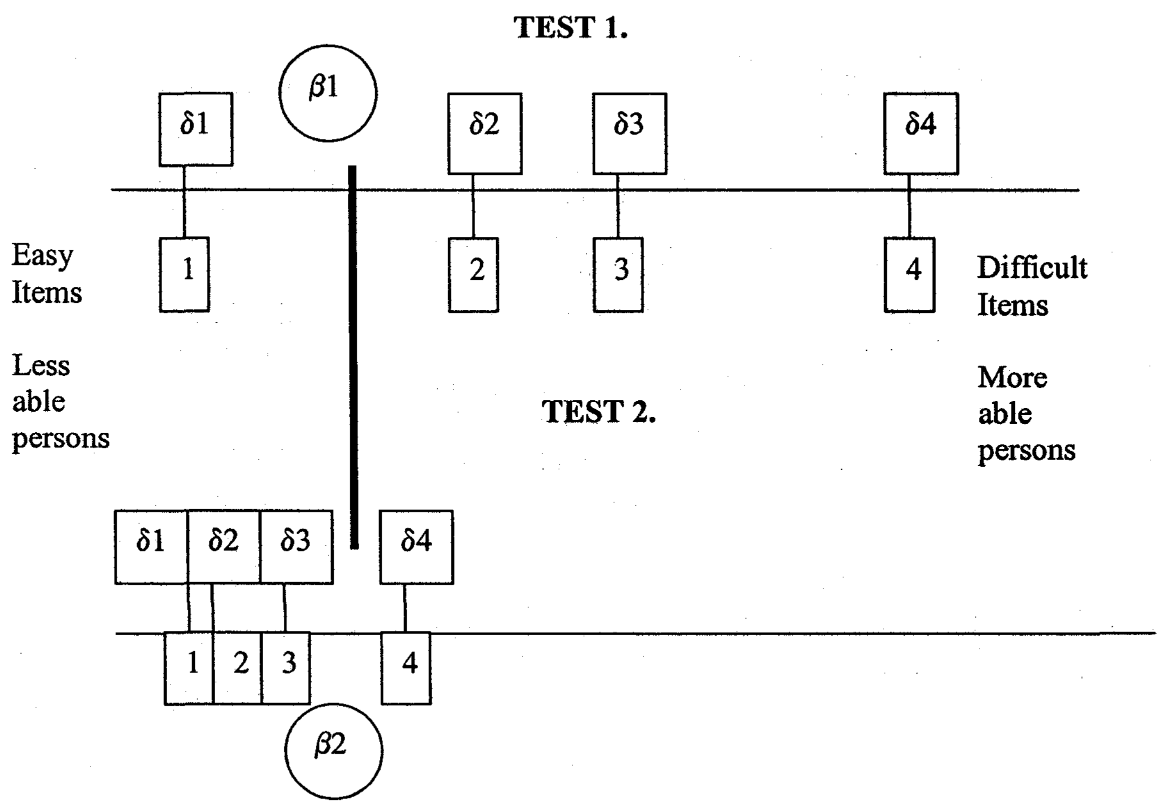
This practice stems from Steven's representational theory where numbers are used to represent relations between objects.[136] Followers of this theory define measurement as the *"assignment of numerals to objects or events according to a rule"*. However, the ordinal structure of the data is usually ignored and wrongly treated as if it were interval and total scores are subsequently used in statistical analyses.[137] Means and standard deviations should not be used where ordinal data are concerned as is presented in this replication of a table from the 1946 article from Stevens.[136] (Table 3.2.).

Table 3.2 Permissible Statistics (Adapted from Stevens (1946)).[136]

Scale	Basic Empirical Operations	Mathematical Group Structure	Permissible Statistics (invariantive)
Nominal	Determination of Equality	Permutation group x = f(x) f(x) means any one-to-one substitution	Number of cases Mode Contingency Table
Ordinal	Determination of greater or less	Isotonic group x = f(x) f(x) means any monotonic increasing function	Median Percentiles
Interval	Determination of equality of intervals or differences	General linear group X = ax + b	Mean Standard deviation Rank-order correlation Product-moment correlation
Ratio	Determination of equality of ratios	Similarity group X = ax	Coefficient of variation

The consequences of treating ordinal data as if they were interval are numerous. Because the true distances between the items and also between the responses options of the items are unknown, wrong conclusions can be drawn about differences between people as well as about change. Furthermore, total scores depend on the level and spread of item difficulties and this can be represented graphically (Figure 3.2). The ability of a person is represented by the symbol β and the difficulty of an item by the symbol δ. Person 1 with ability $\beta 1$ scores a 1 on this particular test because her ability level is superior to the difficulty level of item 1 ($\delta 1$) but inferior to the difficulty level of item 2. Person 2, on the other hand, scores a 3 because her ability level is situated between items of difficulty levels 3 and 4. Despite these two persons having the exact same ability level, they get different scores because they took different tests.

Figure 3.2 Level of Spread of Item Difficulties (figure adapted from Best Test Design.[138]

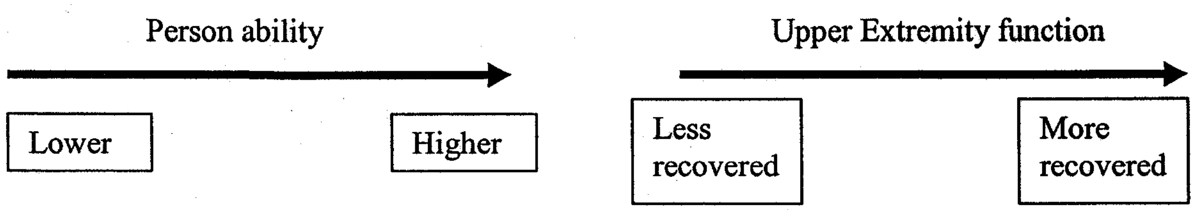


As demonstrated above, two persons of the same ability can obtain different scores because of the spread of the item difficulties that are hypothetically different but unknown in Classical Test Theory. According to Wright and Masters,[139] there are four requirements that must be met for true measurement to occur:

1) *The reduction of experience to a one dimensional abstraction*

In simple terms, this indicates that we cannot measure two different attributes and subsequently add them together. Just as when reporting the measurements of a table, we do not add height, width and length together. Similarly, we cannot add the scores on different items in an index before unidimensionality has been ascertained. This unidimensionality is represented graphically by a horizontal line.

Figure 3.3 Unidimensionality represented by an arrow.



2) *More or less comparisons among persons and items.*

- Just like valid comparisons can be made between the lengths of different objects, longer vs. shorter, adequate comparisons should be made between items, harder vs. easier, and between persons, able vs. less able or more recovered vs. less recovered.

3) *Linear magnitude allowing for the positioning objects or persons along a line.*

- Meaning that the unidimensional construct measured allows for the positioning of persons and items along one continuum from less to more.

4) *A unit determined by a process which can be repeated without modification over the range of the variable.*

- A centimeter remains the same length whether measuring an infant or a full grown person and the units of measurement of upper extremity function should remain constant throughout stages of recovery.

3.4 Moving from Indices to Measures: The Rasch Measurement Model

Before the discussion about Rasch theory, a brief definition of the most common terms used within this framework are presented for reasons of clarity. These definitions are adapted from Bond and Fox.[137]

Item: a question (including its response choices) in a test or index.

Item difficulty: the estimate of an item's underlying difficulty. It describes how hard or how easy this item is to answer, to perform or to endorse. This statistic describes the location of the item along the construct continuum. In health outcomes measurement, the term difficulty may apply when measuring physical functioning. An item with high difficulty may ask if a person can perform vigorous activities like running and an item with low difficulty may ask if a person can walk one block.

Person: any individual answering or being tested on an item.

Person ability: an estimate of how much of a particular trait a person possesses. The higher the ability estimate, the more of the trait the person possesses. A person with high ability, in the context of health outcome measurement would be able to perform more vigorous activities than a person with less ability.

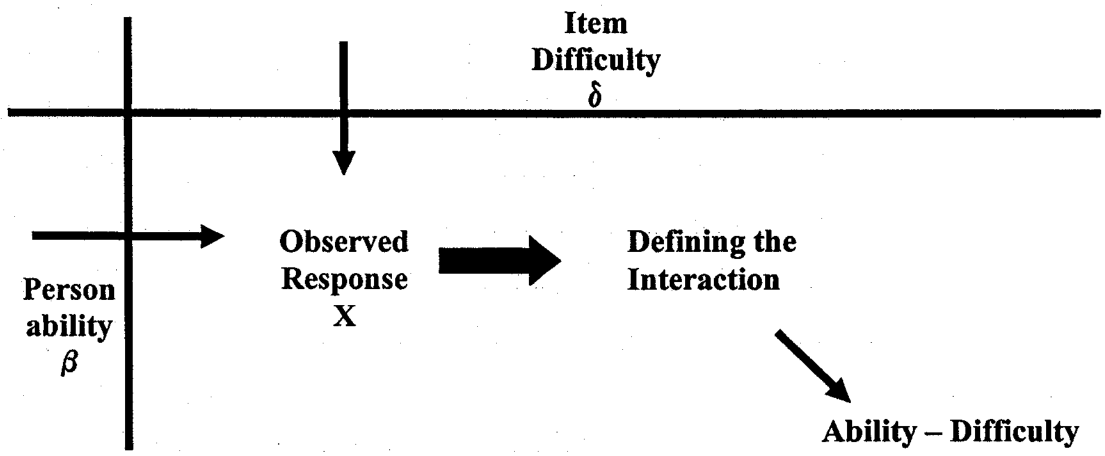
Logit: the unit of measurement that results when the Rasch model is used to transform raw scores obtained from ordinal data to log odds ratios on a common linear scale.

Fisher [1] argued that rehabilitation professionals need to progress towards the use of a common language for functional assessment. In order to meet this requirement, a method of measuring outcomes that is independent of the particular items that compose the existing test or indices is needed; a way to overcome the limitations of CTT is called for. The road to constructing such a measure is complex and the basic principles underlying this theory of measurement must first be understood.

George Rasch, a Danish mathematician, first introduced his models between 1952 and 1960.[140] Rasch developed his first model as he was analyzing data from several tests given to children who had trouble reading. He wanted to 'measure' each child independently of the test used or of the circumstances around each testing situation. To 'measure' the child meant measuring how much of a particular property the child possessed, in this case, reading ability. Because such properties are not directly measurable by the use of instruments, such as measuring temperature with a thermometer, they are called 'latent traits'.

Rasch wanted to find a way to define the interaction taking place when a person confronts an item on a test. An easy way to represent this interaction is to subtract the difficulty of an item from the ability of the person.[138]

Figure 3.4 Relationship between item and person (figure adapted from Best Test Design).[138]



When a person possesses more ability than the difficulty of the item administered, he or she should be able to succeed on that item. This is similar to the deterministic Guttman scale defined by Guttman as follows:

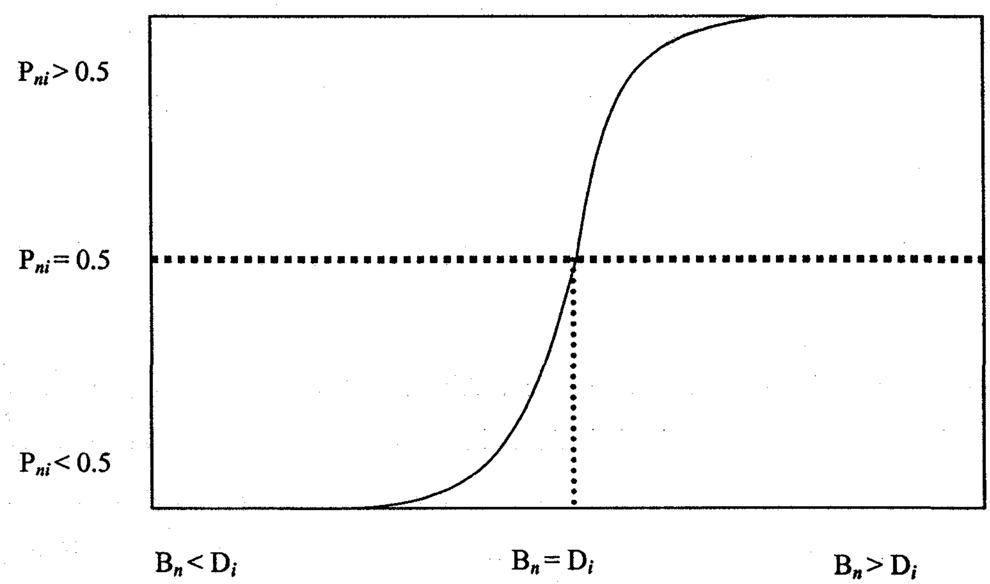
If a person endorses a more extreme statement, he should endorse all less extreme statements if the statements are to be considered a scale.... We shall call a set of items of common content a scale if a person with a higher rank than another person is just as high or higher on every item than the other person.[141]

However, this is not always the case and persons and items rarely behave in such a deterministic fashion. A probabilistic model is thus more appropriate. Rasch defined it as:

If a person's ability is greater than the difficulty of the item, the probability of succeeding should be greater than one and conversely, when a person's ability is lower than the *difficulty of the item they should have a probability of succeeding smaller than one half. When the ability is the same as the difficulty, the probability should be exactly one half.*[142]

The Information Characteristic Curve (ICC) is the basis of the Rasch measurement model. It describes the relationship between a person's ability and an item's difficulty. On the Y axis is the probability of a correct answer and on the X axis are the persons' abilities. As a person's ability increases from left to right, the probability of a correct answer increases. When the ability of the person is equal to the difficulty of the item, the probability of the person answering the item correctly is 50%. The slope of the ICC, for dichotomously scored items is 0.25.[143]

Figure 3.5 The Item Characteristic Curve (adapted from Best Test Design.[138]



Because the difference between *difficulty* and *ability* varies from – infinity to + infinity but the probability of a correct response must remain between zero and one, the difference is applied as the exponent of base *e*.

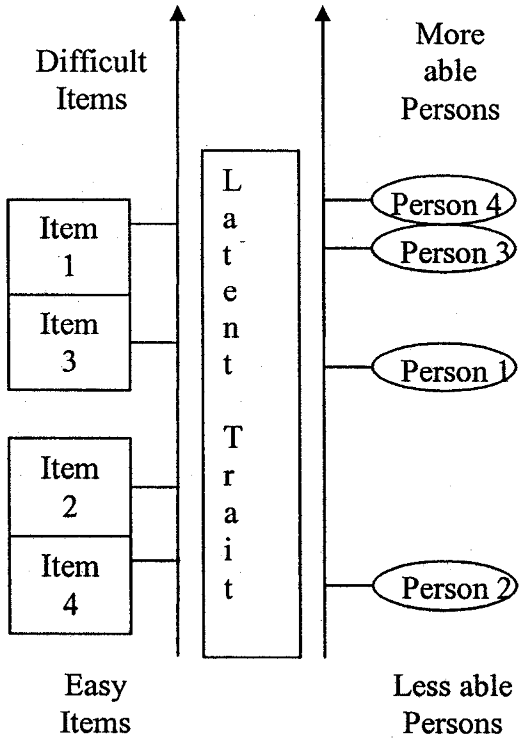
$e^{(\beta-\delta)} = \exp(\beta - \delta)$ this will vary between zero and infinity thus further transformation

$$\frac{\exp(\beta - \delta)}{1 + \exp(\beta - \delta)}$$

will allow the difference to remain between zero and one. This formulation describes the shape of the curve in Figure 3.5 and defines the probability of a successful answer to an item. It is the Rasch model and is the key to creating linear measures. In fact, the Rasch model is the only way to convert ordinal measures into interval-like, meaningful measures. The assumptions underlying this model are unidimensionality, or measuring a unique latent trait, and local independence. The latter assumes that responses to items are not correlated once the effect of the latent trait has been removed. These two conditions are highly similar as a violation of local independence occurs with the presence of multidimensionality.

When measuring an attribute, the person is positioned along the continuum of this attribute or latent trait by presenting the person with a set of items. The items are also positioned along the continuum of the latent trait being measured. A graphical representation is given below.

Figure 3.6 Distribution of Items and Persons on the same ruler.



In this figure, person 4 has the most ability and it can be stated how much more able this person is in comparison to the other persons. In the same way, it is known which items are the most difficult and by how much. But further than that, it is known that person 4 has a high probability of answering all questions correctly because his/her ability is higher than the difficulty of the items. On the other hand, person 1 has a high probability of correctly answering items 2 and 4 but a low probability of answering items that are higher on the scale than their ability level (items 1 and 3). The person ability and item difficulty estimates on this scale, having been log-transformed, are now called logits. The Rasch model also overcomes the inter-dependence between items and persons present under CTT and separates person ability from item difficulty. The abilities of two persons can be compared independently from the items each was administered. To illustrate, consider persons A and B with respective abilities β_A and β_B. These two persons are administered a set of items and the number of correct and incorrect responses are recorded.

Under the Rasch model their abilities can be estimated as[144]

$$\mathrm{Ln}\,(N_{10}/N_{01})$$

where N10 is the number of times A was right and B was wrong and N01 was the number of times A was wrong and B was right. Using this model, any set of items can be chosen without influencing the relative abilities of persons A and B. The same can be applied to the items. The use of iterative pairwise comparisons allows the relative locations of persons to be estimated. Abilities are no longer test-dependent and difficulties are no longer sample-dependent. In other words, two persons can be administered two totally different sets of items and still be compared objectively. It is the same as using two different rulers to measures two persons' height, the comparison is made possible by 'standard units of measurement'.

3.4.1 Extensions of the Rasch Measurement Model

The basic Rasch model, represented by the formula given above, was developed to analyze dichotomous data, scored 0 and 1, Yes/No or Right/Wrong. There are, however, numerous Rasch models. Binomial Trials, Poisson Counts and Ranks models, to name a few, are used to analyze data presented in different types of scoring formats. The Rating Scale Model developed by Andrich[145] lends itself to situations where the outcome is given more gradations than 0 or 1, that is, polytomous data. An example would be a questionnaire asking participants to rate how they agree with certain statements: 'strongly disagree/disagree/agree/strongly agree'. The implication is ordered categorization, such as when a score of 0 is awarded for a poor performance, a 1 is given for partial accomplishment and 2 for excellent. In such situations, the ICCs are called Category Characteristic Curves (CCCs) and they represent probability of a response for each of the categories. The Partial Credit Model was developed by Wright and Masters.[139] It is similar to the Rating scale model but is useful for a number of diverse situations when all items within a test or index are not scored in the same way. For example, this model can be applied when credits are allotted for items that are partially correct or when items require a sequence of steps to be completed. It is also useful when creating a bank of items from a pool where items are scored differently, have a different number of scoring categories and where the categories have different meanings. It can be represented by the formula:

$$\ln\left(\frac{P_{nix}}{P_{nix-1}}\right) \equiv B_n - D_i - F_{ix}$$

The particularity of this model is that each item has its own threshold parameters. *Thresholds "specify the points where the probability of a response of either 0 or 1, and 1 or 2 respectively, are equally likely".*[143] Once the appropriate model is chosen in relation to the data at hand, the parameters of the Rasch model, item difficulty, and person ability, must be inferred from the data, that is, their relative positions on the latent trait continuum must be determined. Because of the processes involved in the estimation methods, and the amount of data at hand, computers are now an indispensable tool.

3.4.2 Estimation Procedures for Person Ability and Item Difficulty

Estimation refers to the positioning of the position of items and persons along the linear continuum of the latent trait being measured. There are numerous estimation methods used by different statistical packages. They can be divided into non-iterative and iterative methods.[146] Iterative estimation methods are based on initial approximate starting values for the estimates. A comparison is made between observed estimates and what is expected from the model. This process is repeated until the discrepancies between observed and estimated are considered small and unimportant. When this stage is reached, the estimation process is considered to have converged.[146]

The Rasch Unidimensional Measurement Models (RUMM)[147] program uses pair-wise estimation (PAIR). This iterative approach uses the relative frequencies of observations in the data to estimate the parameters.[146] Only if the data fit the model can we be confident that the estimates obtained for item difficulties and persons abilities lay on the same linear continuum (unidimensionality). This property allows for the calculation of means and variances. Rasch models also provide estimates of modeled error variance for each of the calibrated difficulties and abilities. These estimates give an indication of how precise the estimation is and allow for the calculation of confidence intervals.[148] A noteworthy advantage of the PAIR iterative procedure is that it is useful for analyzing incomplete data and this, in turn, is particularly useful in situations where overlapping test forms are being co-calibrated to create an item bank.

3.4.3 How well do the Data fit the Rasch Model – The Overall Analysis of Fit

Only in the instance that the data fit the model do the characteristics of that particular model hold true.[149] In Rasch analysis, fit statistics are fundamental. If parts of the data do not fit the model, a decision to modify or reject part of the data needs to be taken. According to Stone,[150] there are two necessary conditions for an appropriate fit. First, more able persons should have higher probabilities of succeeding on an item than less able persons. Also, easier items should be successfully performed more often than are harder items. This theory arises from the work of Guttman who put forward the deterministic model described earlier. However, as the Rasch model is probabilistic it requires a certain degree of randomness.
Wright and Panchapakesan[151]in 1969 were the first to introduce a global fit statistic. It was based on the difference between the number of correct responses by a group of examinees with similar raw scores and the number of correct responses predicted based on the model. These statistics are called between-group fit statistics and are based on the Pearsonian chi-square. They are the most widely used indices of fit[149] and thus the ones that were used to carry out this project.

<u>Pearsonian Family of Fit Indices – Residual Analysis</u>
Residuals are always obtained from the subtraction of the expected from the observed scores. Because the values of the residuals have different variances, they can be standardized by dividing them by their standard deviation.[149] The magnitude, either positive or negative, of the standardized residual is an indication of how unexpected a response is. When the standardized residual is large (eg. ±3) the probability of a correct

response is either very high or very low in the unexpected direction. Squaring the standardized residual is a solution to removing the sign. This value can then be treated as a chi-square with one degree of freedom. These chi-squares can then be interpreted for the entire data (overall fit), or for interpreting the fit of the persons or for the fit of the items. The details of this particular approach will be discussed in more depth in the discussion on item and person fit. Another method of analyzing fit is the likelihood ratio chi-square approach was developed and elaborated by Andersen,[152] Gustafason[153] and van de Wollenberg (1982).[154] This method will not be discussed in this paper. Smith and Hedges [155] have demonstrated that these two statistics, the Pearsonian chi-square and the likelihood ratio chi-square are highly correlated and that the former can be used with confidence.

Other methods

Parametric fit indices such as those presented above are now being criticized for inadequately detecting departures of the data from the model, because, as their name implies, they are derived directly from the data. [156] The standardized Z residual uses the subtraction of nonlinear ordinal scores. In reality, Z is an ordinal score. A residual-free statistic has been proposed and it has the form: [156]

$$G = \frac{\sum^{AllItemPairs} X_{nh}(1 - X_{ne})}{r(L - r))}$$

r = item total score

L = number of test items

Xnh= response to most difficult item in the pair

Xne = response to easiest item in the pair

The numerator in the equation is the number of Guttman response errors observed across item pairs. This formula, however, does not take into account the size of the Guttman error.[156]

A Q statistic that weighs each Guttman error has been proposed.[157] Levine [158] and Drasgow [159] and Klauer [160] also devised a way to detect departures from the model called the optimal response. This particular statistic is used for the assessment of person fit and involves testing a null model of a normal Rasch-like response behavior against a non-normal one.

Non-parametric approaches can also be used to verify the shape of the ICC's and consideration will be given to collapsing items if necessary.

3.5 The analysis of Item Fit

The fit of each individual item to the model can be evaluated and interpreted separately. Certain computer programs present three types of item fit indices: (1) total fit, (2) between fit and (3) within fit. As discussed previously, they are based on squared residuals. The difference between the three indices is that they are summarized differently.[149]

All three of these fit statistics can be calculated as weighted (INFIT) and unweighted (OUTFIT). The analysis of fit provides statistics that are indicative of how well the data fit the Rasch measurement Model. The INFIT and the OUTFIT statistics are the most widely used misfit indices. They have the general form of a chi-square statistic divided by their degrees of freedom. The INFIT is the weighted fit statistic and is less affected by abnormal responses of persons with ability far from the difficulty of the item. The OUTFIT, the unweighted fit statistic, is outlier-sensitive. Their values are expected to be around one. Values of less than 1 indicate that the data varied less than what is predicted by the model whereas values superior to 1 indicate excess noise in the data.[161] According to Wright and Linacre, [162] the INFIT and OUTFIT statistics should range between 0.5 and 1.7 for clinical observations.

Table 3.3 Interpretation of Fit Statistics. Adapted from Bond and Fox.[137]

Mean Squares	t	Response Pattern	Variation	Misfit Type
> 1.3	>2.0	Too random	Too much	Underfit
<0.75	<2.0	Too determined	Too little	Overfit

As would be expected, the smallest standard errors are obtained when the data fit the model. Although the overall chi-square statistic was initially developed for dichotomous items, it can also be used for polytomously scored items as well.

In RUMM, both a χ^2 and F statistics are provided. If they are statistically non-significant, it indicates fit to the model.

3.6 The Analysis of Person Fit

As is the case with the item fit statistics, there are three types of person fit statistics: (1) total fit, (2) between fit and (3) within fit and two versions of each can be calculated as the weighted and unweighted.[148] Person fit statistics, as the name implies, give information of how examinees answered questions. Departures of the observed data from what was expected are indications that persons did not answer or perform in the expected manner. This can be due for example to cheating, guessing or misunderstanding of the question(s). The reasons for these departures need to be examined and a decision is required regarding the deletion of part(s) of the sample from the analyses.

In RUMM, fit statistics for persons are presented as a Z statistic that approximates a standard normal deviate. Just as with the t statistic described earlier, a large + value indicates a large deviation from the Guttman pattern and a large negative value indicates overfit or too Guttman-like. Ideally, it should be around 0.

3.7 The analysis of Category Misfit or Ordered thresholds.

In the case where there are more than two answers or scoring options for the items, the fit of each category must be assessed. Rasch analysis is well suited to investigate the usefulness of the categories used in a rating scale. All the indices, including those originally scored on an ordinal scale have to be scrupulously examined for the quality of the categories in their rating scale. According to Guilford,[163] categorization of the response options should be well-defined as well as mutually exhaustive. The graph presented below, the polytomous version of an ICC, is an example of appropriate categorization. Each of the scoring options (0, 1, 2) are ordered adequately, 0 being at the left where persons of less ability are situated, 1 is located right next to the 0 and the 2 comes at the right of the graph where persons with the most ability are located and are thus more likely to endorse that response option. Furthermore, each category has its own 'hill'. This means that an ability level exists for which that particular score is most likely to be used. Looking at the second graph, category 1 does not possess its own hill and this indicates that this category has not been used to score that particular item and may be useless. As is the case when items or persons do not fit the model, the data often require manipulation. Optimal categorization needs to be reached in order for the item to fit. This can be done by collapsing categories. It is frequently necessary to combine categories together, when one of them is not used or is under-utilized.

Figure 3.7 Example of adequate categorization of rating scale.

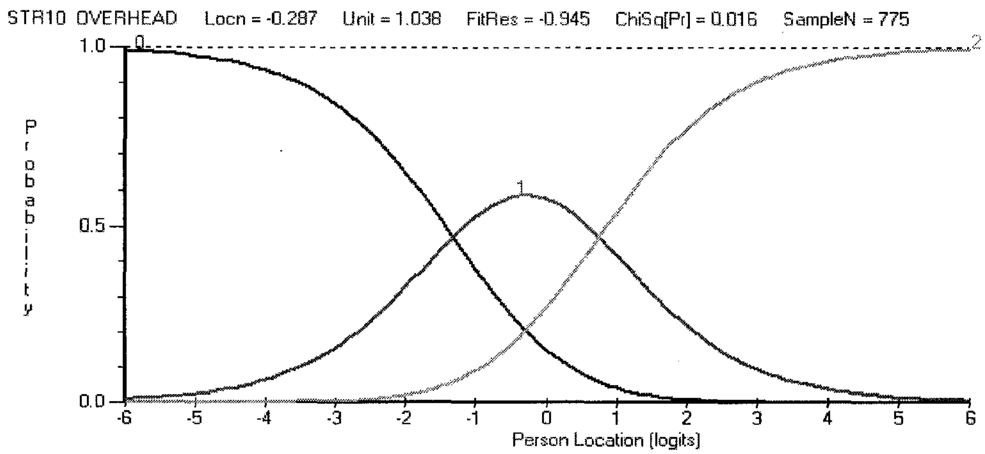
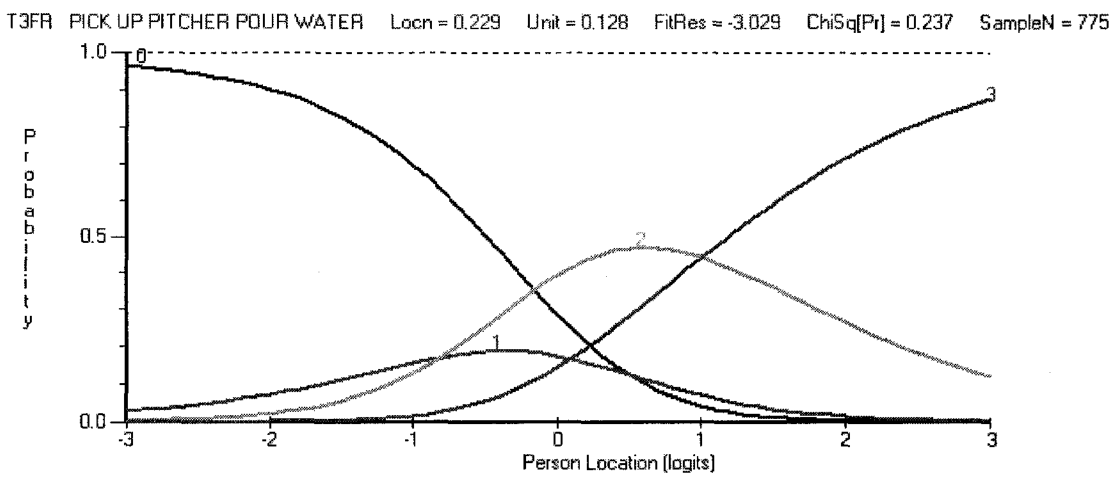


Figure 3.8 Example of inadequate categorization of rating scale.



Analysis of category fit can should also be done numerically. Linacre[164] proposes several guidelines in order to verify the usefulness and the performance of the categories.

(1) There must be at least ten observations representing each of the categories. In other words, if less then ten persons chose a particular category, this may be an indication that this particular choice is under-utilized and thus not useful.

(2) The distribution of the observations must be regular. A uniform distribution of observations *across categories is ideal. Other acceptable distributions are the unimodal*

where the center category is more frequently used, or bimodal distribution where extreme categories at each end are preferred.

(3) The average measure statistics (average item location in logits) must advance monotonically with category.

(4) The OUTFIT mean square must be less than 2.0.

(5) Step calibrations must advance or increase as this will indicate that a person possessing more of the trait will choose or score in categories that are higher up in the hierarchy of the scale.

(6) Step difficulties between rating categories advance by at least 1.4 logits.

(7) Step difficulties between rating categories advance by less than 5.0 logits.

3.8 The Ascertainment of Unidimensionality

One of the important assumptions of Rasch models is that the construct being measured is unidimensional. Item fit statistics, described above can be used to detect multi-dimensionality. Several groups along the years have stated that if the data fit the model, it is evidence of unidimensionality. Others have suggested that the fit indices provided with a Rasch analysis are not sufficient to confirm unidimensionality. Smith[165] has compared two methods: the Rasch fit statistic and Principal Component Analysis (PCA). He concluded that the number of response options is important in determining which method is most appropriate to use. PCA may be more appropriate when there is no clear definition of the construct being measured and when one has no prior knowledge whether the items are measuring the same construct. In simple terms, PCA provides an estimate of how many constructs or factors are represented by the set of items being analyzed. There are several other different ways of verifying unidimensionality. Three main approaches are discussed in the literature:

1) Prior testing such as Factor Analysis and Principal Component Analysis
2) Fit to the Rasch Model

3) Post-hoc testing: Principal component Analysis of the Residuals after fit to the Rasch model and Smith's independent t-test approach.

According to Tennant,[166] using an exploratory factor analysis *a priori*, with parallel analysis to indicate significant eigenvalues, gives indication of the dimensionality of the items before fitting the data to the Rasch model. Examining the Principal Component Analysis of the residuals after the fit to the Rasch model is then recommended. As well, comparisons of person estimates derived from different subsets of items, using Smith's independent t-test approach should be used.

3.9 Targeting

The ability to place items and persons along the same linear continuum is inherent to the Rasch measurement model. It allows determining whether the particular items administered to a group of persons were targeted to them. The most practical way to assess targeting is visually, through the use of item-person maps. If along the central line, which represents the attribute being measured, the items are clustered towards the top of the line and the persons are at the bottom, this particular set of items would be considered too difficult for the group of persons assessed. The impact of this is to decrease the ability to estimate with precision the true position of persons along the construct. Conversely, if the items are located lower than the persons, these items would be considered too easy, mis-targeting this particular population and again, decreasing the precision of the ability estimates of the persons.

If there are no items located in the vicinity of the persons or if there are important gaps or distances between the items, the location of the ability of these persons cannot be estimated with precision. Ideally, items should be spread evenly on the continuum and range from -3 to +3 logits.

3.10 Detecting Differential Item Functioning

An important step when constructing a new measure is to investigate the presence of Differential Item Functioning (DIF). DIF occurs when there is a *"loss of invariance of item estimates across testing occasions."*.[167] For example, when estimates of item difficulty vary in their position on the linear continuum according to different populations being assessed (e.g. males vs. females) this item is said to demonstrate DIF. Item difficulty estimates have to be compared across the samples of interest. The between fit statistic is the fit index that will detect the presence of DIF. Item bias or DIF violates the property of invariance inherent to the Rasch model and items displaying this characteristic must be looked at carefully. DIF can also be detected using the ICC's. If the location of the curves differ while the slopes are identical, it indicates the presence of uniform DIF. When the slopes are not parallel and cross each other, it is indicative of non-uniform DIF.

Another way of detecting DIF is to use ordinal regression. This method can be used to analyze ranked outcomes in which the ordinality of the data can be maintained.[168]

A relatively new method of detecting DIF is through hierarchical generalized linear models (HGLM) for polytomous items.[169] According to the authors, however, this model imposes a lot of restrictions (constant item discriminations and fixed threshold values across items) and the impact of the distribution of the latent trait on the estimation has not yet been fully evaluated.

Figure 3.9 Example non-parallel curves demonstrating DIF.

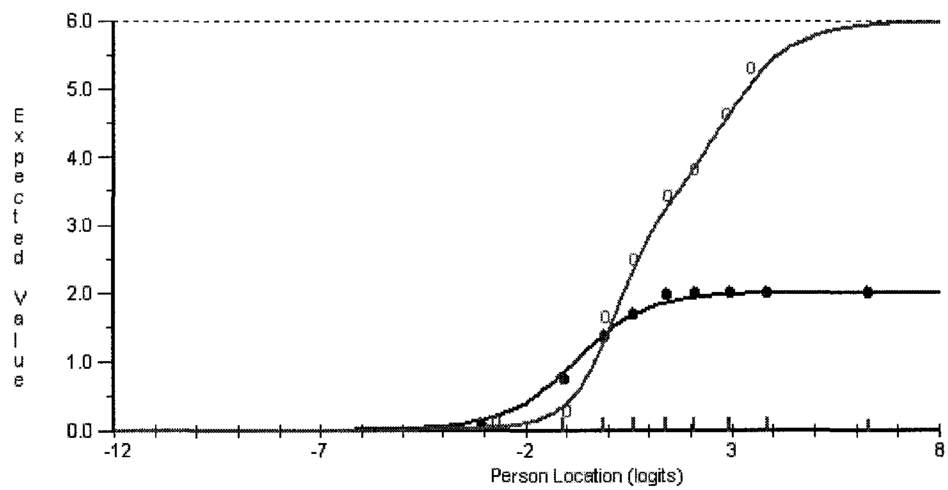
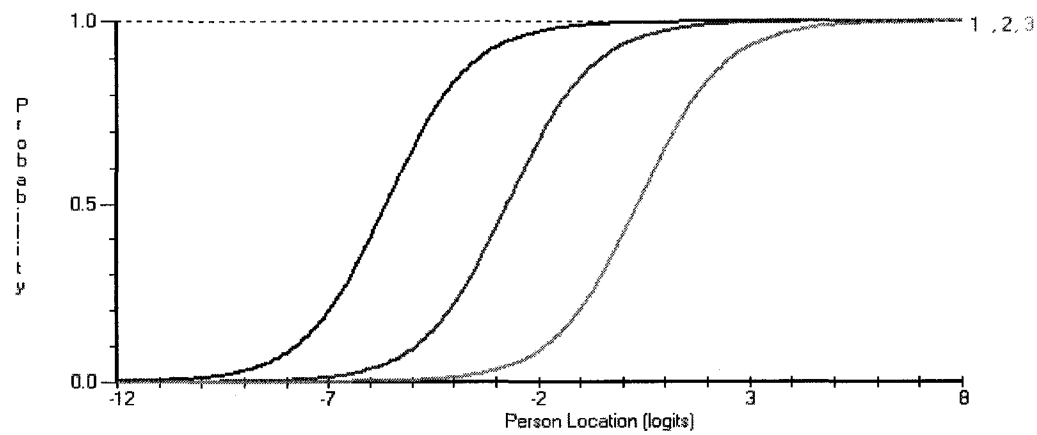


Figure 3. 10 Example of parallel curves demonstrating the absence of DIF.



3.11 Item Banking

An item bank is a large collection of questions that are organized, calibrated, and matched to a given construct or task.[170] Items in a bank are available for the design of an adaptive test. Items in a bank are used to form the item pool for any particular adaptive test. According to Flaugher,[171] the quality of the item pool, and thus the bank, is important in order to benefit from all the advantages of adaptive testing. It must contain a sufficient number of well-written items that span the range of proficiency. The new item bank for the creation of an upper extremity measure must contain items from the easiest movements that can be performed by the most severely affected patients to the hardest ones that can only be accomplished by those with near normal upper extremity function. In the following sections, the steps required to create the item bank designed to measure upper extremity function will be described.

3.12. Linking

Linking is a process that places items from different indices, each index measuring the same latent trait, onto the same linear continuum. The steps involved in this procedure when using RUMM statistical software and when complete data are available include: (1) calculating sufficient statistics for item parameters, (2) estimating item parameters using pairwise estimation algorithm, (3) estimating person location parameters using the values for item parameters estimated in the second step.[172]

In the process, the items from the pool are placed along the linear continuum of upper extremity function. In order to fit these items onto a line, they need to be connected in some way. There are numerous designs than can be used to co-calibrate indices. Data previously collected as part of various research projects where some tests and indices overlap between studies can be used to create an item bank

3.13 Psychometric properties in Rasch measurement

Wolfe and Smith[173] present different ways of ascertaining validity within the Rasch framework. Seven different types of validity are discussed and how the evidence can be generated when developing measures:

Content Validity is concerned with *content relevance*, based on expert opinions, *representativeness of the items*, examining the empirical hierarchy and spread of the item calibrations along the variable and *technical quality*, assessed via item fit statistics.

Substantive Validity refers to *"theoretical rationales for the observed consistencies in test responses along with empirical evidence that the theoretical processes are actually engaged by respondents in the assessment tasks."* This type of validity may be addressed by verifying the definition of the variable intended by the researchers (confirmation of the intended item hierarchy) and examination of person fit statistics.

Structural validity is ascertained when the chosen measurement models requirements, such as unidimensionality, are satisfied.

Generalizability is concerned with the degree to which inferences based person measures or item calibrations are invariant, across different tasks, time, groups, or contexts.

External validity is the correspondence between different measures of the same construct, and discriminant evidence, the lack of correspondence from measures of distinct constructs. It is the equivalent of the traditional contruct validity.

Consequential validity is concerned with the implications of score interpretation and the consequence of the use of the measure itself.

Interpretability is the degree to which qualitative meaning can be assigned to quantitative measure and corresponds to the extent to which the meaning of a score can be communicated and interpreted.

Reliability

In the Rasch models, variance of the measurement error is computed from the measurement error that accompanies the patient ability and item difficulty estimates.[167,174] Standard Errors indicate the precision of each estimate of item difficulty and person ability.[148] It is the 'space' within which the true difficulty and true ability should fall. Repeated testing is thus not necessary to analyze reliability. When using RUMM, there are two reliability indices provided. The first is the Person Separation Index represented the degree to which the relative variation among the persons is not random.[143]and the second is Cronbach's Alpha. These two statistics are very similar to the ones seen in CTT. According to Smith,[148] the person separation index represented by G and calculated as follows should be used:

G = $\sqrt{R / (1 - R)}$ where R is the Rasch person reliability named persons separation index in RUMM. To avoid confusion, the person separation index in RUMM will be called person reliability and G will be called person separation index in this thesis. The more dispersed are the person measures on the construct being measure, the higher will be the value of G. The STRATA, computed by:

STRATA = (4G +1)/3 and indicates the number of distinct ability level separated by three errors of measurement. Rasch item reliability is determined in the same way.

3.14 Computer Adaptive Testing

Once the item difficulties have been estimated using a Rasch model and placed along a single ruler, an item bank is formed from which any subset of items can be drawn to make up a test. This ability to order all the items on the same scale has been at the root of a relatively new form of testing called Computer Adaptive Testing or CAT. In simple terms, a CAT is a form of test where the examinee is presented with an item on a computer, enters their answer and the computer, through a programmed algorithm, will select the next question to be presented to the examinee depending on their answer to the previous question. This process will go on until the person's ability estimate has been calculated with a satisfactory standard error of measurement set by the test administrator.

3.15 The Rasch Measurement Model vs Item Response Theory (IRT)

The Rasch model is termed by some as the most parsimonious of the IRT models. L.L. Thurstone is one of the pioneers who developed the foundation for Item Response Theory in a parallel line with Rasch. Although these two theories are often presented side by side, as being part of the same family, they are different. Indeed, the philosophy at the basis of their development is different in several aspects. The Rasch model is a definition of measurement, a model to obtain stable either linear measures or on a monotonic scale, from ordinal data.[175] On the other hand, IRT models were developed to explain which model is best to describe the data at hand. [176] In a recent article, Massof[177] concludes that Rasch models are valid measurement models and IRT models are not. The battle between the proponents of these two diverging theories is not resolved. According to Hobart,[176] if we want to construct a true linear measure such as those used in physical science, as opposed to explaining data, Rasch item analysis should be used. Because the main objective of this work is to develop a linear measure of upper extremity function, as opposed to explaining the data at hand, the Rasch model has been chosen.

CHAPTER 4 Rationale and Study Objectives *(Manuscripts 2, 3 and 4)*

Outcome measurement in the field of rehabilitation has entered a new era. Its importance in the discovery and development of effective interventions has been recognized. For years, Classical Test Theory has been the main pillar upon which tests and indices have been developed. In recent years, however, new methods have penetrated the field. This method has been used extensively in the educations and psychology. At the present time, there are separate indices of upper extremity function, each one capturing only one narrow aspect of this construct, there is no link between the tests, and it is difficult to extract and communicate meaningful information. The rehabilitation field is predominated by tests and indices that are scored differently, each having their own scaling and whose total scores do not have any inherent meaning. Another noteworthy disadvantage is the burden to clients and research participants as the assessments can be lengthy and demanding.

The overall aim of this the second part of this project is to create a method to parsimoniously and hierarchically quantify upper extremity function post-stroke along an identifiable construct using Rasch analysis.

The specific objectives were:

1) To summarize the use of Rasch analysis and Item Response Theory in rehabilitation;

2) To identify how many unique constructs comprise upper extremity function from existing pools of upper extremity function indices;

3) To use the Rasch measurement model to identify items that are able to discriminate between levels of recovery along the construct(s);

4) To hierarchically align items to create a bank of items and allow the assignment of a recovery level for individuals;

5) To develop a paper version of an adaptive measure of upper extremity function for clinical use of the new measure.

CHAPTER 5 Manuscript 2

Preface to Manuscript 2

The use of Rasch Analysis and Item Response Theory in Rehabilitation: A Review of the Literature.

The first manuscript of this thesis presented a randomized controlled trial that showed that an upper extremity task-oriented intervention was not effective in improving arm function among persons in the chronic phase post-stroke. Difficulties were experienced in administering a series of tests and indices to capture upper extremity function to persons with stroke. There were also challenges to analyze and summarize results from several tests and indices scored differently and for which total scores are difficult to interpret. This lead to the desire to find a new method of measuring upper extremity function that would overcome these obstacles and fill an obvious void in the measurement of outcomes in rehabilitation, more specifically, in the measurement of upper extremity function after stroke. After summarizing the effects from twelve randomized controlled trials aimed at evaluating the effectiveness of a rehabilitation intervention on upper extremity function after stroke, it was clear that there were no agreement between researchers as to which upper extremity outcome assessment to use. There are no 'gold standards'. Most researchers use several different tests and indices are then subsequently faced with the difficulties in analyzing and interpreting the results.

A solution to this problem is imperative if we are to find effective treatment for the upper extremity. When the author became aware, through readings and conferences, of new 'modern' psychometric models that could overcome the problems faced when analyzing results emerging from tests and indices that were developed in a classical way, it was important to find out if there were existing 'true measures' of upper extremity function developed for persons having suffered a stroke.

Rasch analysis and item response theory, used extensively in the fields of psychology and education, have received an increasing amount of attention from researchers who understand the need to propel the field of rehabilitation into evidence-based practice.

In the second manuscript, an extensive literature review is presented. The objective is to present studies that have used Rasch analysis or Item Response Theory in the field of rehabilitation either to: (1) develop a new measure, (2) improve or determine the psychometric properties of an existing test or index or (3) analyze results by fitting the data to Rasch or Item Response Theory model or (4) collect their data using a measure that has been developed through the use of Rasch or Item Response Theory.

This extensive review of the literature will provide an accurate picture of the use of modern psychometric models in the field of rehabilitation. It will inform the candidate about potentially existing measures that have been developed to assess upper extremity function after stroke.

Title Page

The use of Rasch Analysis and Item Response Theory in Rehabilitation: A Review of the Literature.

Authors: Johanne Higgins, MSc, O.T., Nancy E. Mayo, BSc(PT), Ph.D.

Posts of the Authors at the Time of the Study:

Higgins J: PhD candidate, School of Physical and Occupational Therapy, Faculty of Medicine, McGill University.

Mayo NE: Associate Professor, School of Physical and Occupational Therapy, Faculty of Medicine, McGill University, Associate member, Department of Epidemiology and Biostatistics and Occupational Health, Faculty of Medicine, McGill University

Work conducted in: Montreal, Quebec, Canada, and Quebec City, Quebec, Canada

Address correspondence to: Nancy Mayo, Royal Victoria Hospital, Division of Clinical Epidemiology, 687 Pine Avenue West, Ross 4.29, Montreal, Quebec, H3A 1A1 CANADA, Telephone: (514) 934-1934 ext. 36922, Fax: (514) 843-1493, e-mail: nancy.mayo@mcgill.ca

Abstract

Objective: To summarize the use of Rasch analysis and Item Response Theory in rehabilitation.

Methods: MEDLINE, CINAHL, Cochrane Library, EMBASE, ERIC and psycINFO were searched for articles using Rasch analysis and Item Response Theory in the field of rehabilitation. The different goals for using these methods were classified into four categories: (1) develop a new measure, (2) improve or determine the psychometric properties of an existing test or index or (3) analyze results by fitting the data to a Rasch or Item Response Theory model or (4) use a measure that had been developed through the use of Rasch or Item Response Theory (IRT).

Results: A total of 357 articles were retained. The year of publication of the articles ranged from 1991 to the first 4 months of 2007. The majority (65%) of the articles aimed at determining or improving the psychometric properties of an existing test or index by the use of Rasch analysis or Item Response Theory. Only a small number of authors (7%) used previously Rasch- or IRT-developed measures to conduct their study. An important number of measures were developed for evaluating the impact of neurological conditions such as stroke and brain injury.

Conclusion: This review demonstrated an increased use of Rasch analysis and Item Response Theory in the field of rehabilitation, especially for the development of new measures and the determination of their psychometric characteristics. The use of previously Rasch-developed measures is still limited and an increased effort to make these measures known to clinicians and researchers is required.

Introduction

The quest for psychometrically sound outcome assessments in the field of rehabilitation is ongoing. This search is justified by the need to use reliable and valid measures to quantify the impact of disability and to estimate the effect of interventions on change in function. Whether measures are used to evaluate individual patients or for research purposes, rehabilitation professionals need several different tests and indices to ensure comprehensive assessment of the phenomenon under study. An enormous number of generic and condition-specific indices and tests have been developed for use with rehabilitation populations but using numerous evaluations has disadvantages. The inability to convert scores from one instrument to another is a major challenge of outcome measurement. It is also difficult to interpret the results of clinical trials because it is difficult to interpret clinically the change scores on all scales.[178] Furthermore, most existing measures are quantified by summing the individual item scores to yield an ordinal measurement scale which is then most often treated as having interval properties for statistical analyses. This can result in incorrect conclusions about differences between groups or change over time within individuals.

The emergence of evidence-based medicine has lead researchers to shift from outcomes that are relevant to the immediate consequences of the condition such as tests of impairments to tests that quantify the severity of the condition by assessing, for examples, participation of the person in daily activities or their quality of life.[178]

Constructs such as upper extremity function, quality of life, depression or satisfaction, that are intrinsic to each individual, cannot be measured directly using an instrument. They are assessed by observing related behaviors, that are unitless.[179]

Rasch proposed a statistical model that allows for the transformation of cumulative raw scores, into linear measures, meeting the assumptions made in measurements in physical sciences: linearity, additivity, equal distances and sample independence.

Rasch and Item Response Theory (IRT) models have been increasingly used in the measurement of rehabilitation outcomes within the last decade.

Objective

The objective of this review of the literature is to describe the extent to which, Rasch analysis and Item Response Theory are used in the field of rehabilitation.

Methods

Literature search

A literature search was conducted in MEDLINE, CINAHL, Cochrane Library, EMBASE, ERIC and psycINFO. The search strategy was built on Rasch analysis or Item Response Theory (type of analysis) and Rehabilitation (type of measure developed). Rasch analysis, Item Response Theory and Rehabilitation were used as MeSH and keywords for the electronic databases. Only articles written in English or French were included. Articles dealing with conditions that are of particular interest to rehabilitation professionals: occupational therapists, physical therapists and speech and language pathologists were included in this review. Articles dealing only with methodological, statistical or mathematical issues and that did not include a particular outcome assessment were excluded.

Classification of the Articles

Articles were first classified according to the main objective or aim the authors wanted to reach. (1) Several articles used latent theory models (Rasch model or Item Response Theory model) toward developing a brand new measure. (2) Others aimed at determining or improving the psychometric properties of an existing test or index. (3) Some authors analyzed the results of their study by first transforming scores on an ordinal test or index by using latent trait theory. (4) Lastly, some articles describe studies is which researchers are simply using a measure that had previously been developed using latent trait theory to conduct their study. The articles were also classified according to the population or type of condition for which the measures were developed or further refined.

Results

A total of 357 articles were found. The year of publication of the articles ranged from 1991 to the first 4 months of 2007. There was a steady increase in the number of publications since the beginning of the 1990's. Figure 5.1 shows the number of articles by year of publication. Up to 1999, between 2 and 13 articles using Rasch or Item Response Theory were published per year. In 2006, a total of 55 articles were found. Eighty-eight percent of the articles found make use of Rasch models in their studies while the rest claim to use an Item Response Theory Model.

Classification of the Articles

Aim

Most of the articles (62%) dealt with the determination or the improvement of the psychometric properties of an existing test or index. (P). A total of 77 (22%) articles described the development of new measures either from brand new items or from items that were taken from existing tests and indices (N). Thirty-five articles (10%) depicted the use of a Rasch or Item Response Theory model to transform scores on tests and indices prior to analyzing the data (A). Finally, in 27 (7%) of the articles, the authors use a previously Rasch-developed measure to assess subjects in their study (U).

Populations

The articles were also classified according to the different population groups targeted by the measures they are describing. Only studies aimed at either developing a new measure or determining and improving the psychometric properties of an existing test or index through Rasch or Item Response Theory analysis are presented (Tables 5.1 to 5.9). For each population category, the year of publication, the name of the measure, the aim of the study (P or N) and a summary of the use of the measure and findings are presented.

1-Stroke

A total of 50 articles were dedicated to the creation or further refinement of measures for stroke either exclusively or along with other conditions. Thirty-four of the articles dealt with stroke exclusively while sixteen articles included other conditions including stroke. Table 5.1 presents thirty-four articles presenting measures that were either developed or further examined and improved through Rasch analysis (labeled as P or N) with stroke populations exclusively. Eight of the articles present the development of brand new measures. Among the new measures, one evaluates the impact of a stroke; other measures assess mobility, functional and motor recovery, activities of daily living and awareness of disability. The remaining twenty-five articles were aimed at determining or improving the psychometric properties of existing tests or indices. The sixteen articles presenting 'generic' measures designed for stroke survivors along with other conditions are presented in Table 5.2.

2-Traumatic or Acquired Brain Injury

Thirty-six studies creating or investigating measures for brain injury were found. Eighteen of these studies were exclusive to a brain injury population (Table 5.3). Four articles pertained to the development of new measures and they assessed (1) needs and outcomes of children and youth with acquired brain injury, (2) functional change in patients, (3) performance and (4) disability. The fourteen remaining articles describe the psychometric evaluation or improvement of existing measures.

3-Cerebral Palsy

Table 5.4 presents articles (8) investigating measures for cerebral palsy. Only one article pertained to the development of a new measure: the ABILHAND-Kids was specifically developed to measure manual ability in children with cerebral palsy as well as providing guidelines for goal setting in treatment planning. Seven articles were classified as determining or improving the psychometric properties of an existing test or index.

4- Multiple Sclerosis

Two articles (Table 5.5) describe measures that were developed for multiple sclerosis. One of these was recently published in 2006. The MSSS-88 is a reliable and valid, patient-based, linear measure of the impact of spasticity in multiple sclerosis. The other one, the EQUISCALE was published in 1997 and is designed to assess balance in persons with this condition.

5-Parkinson's Disease

Articles related to measures designed for persons who have Parkinson's Disease are described in Table 5.6. The Impact on Participation and Autonomy (IPA) questionnaire, the Parkinson's disease Questionnaire (PDQ-39) and the Nottingham Health Profile were examined for the quality of their psychometric properties. The IPA-I shows promise as a tool for measuring participation in people with Parkinson's disease for it has acceptable psychometric properties for measuring perceived problems in participation. The PDQ-39 and the Nottingham Health Profile, measures of health status require developmental work before they can be considered suitable for this particular population.

6-Low Vision

A total of sixteen articles are related to patients with low vision and they are presented in Table 5.7. The medical conditions associated with this impairment are mainly cataracts and glaucoma. Six articles discuss the development of new measures. The new measures were designed to assess activities of daily living, vision disability, quality of life, functioning and a questionnaire for measuring vision difficulties in persons who receive low-vision rehabilitation.

7-Back Pain

Two new measures were developed for back pain. The Back Pain Functional Scale-Physical Functioning Items [180] is a lumbar spine specific Computer Adaptive Test to assess lumbar spine functional status. The Quebec Back Pain Disability Scale[181] comprises 20 items and has been shown to discriminate between different levels of disability in persons with back pain. The other articles assessed and examined the psychometric properties of different tests and indices. Among those were: the Roland Disability Questionnaire, the National Health and Nutrition examination Surveys Activities of Daily Living Instrument, the SF-36, the Oswestry Low Back Pain Questionnaire, the Worker Role Interview (WRI) and the Occupational Rehabilitation Data Base (ORDB) function capacity instrument. All articles relating to back pain and back impairments are presented in Table 5.8.

8- Arthritis and other related inflammatory diseases

Table 5.9 presents fourteen articles related to rheumatoid arthritis, osteoarthritis and other inflammatory diseases. Ten of those studies describe the development of a new measure. Among the measures presented, some assess the quality of life, work instability and ADLs and IADLs in persons with rheumatoid arthritis. One measure is designed to evaluate pain in persons with knee osteoarthritis, another aims at evaluating the severity of osteoarthritis of the hip and the ABILHAND assesses manual ability.

9-Other Conditions

Other articles deal with tests, indices and measures used with several other conditions. Among these are: fibromyalgia, spina bifida, spinal cord injury, depression, amyotrophic lateral sclerosis, Alzheimer's disease, ankylosing spondylitis, cancer, Behçet's disease, Guillain-Barré syndrome and more. As well some 'generic' measures were developed or psychometrically assessed in mixed neurological and musculoskeletal conditions populations.

The Measures

A total of 77 articles describe the development of a new measures using Rasch analysis or Item Response Theory and 82% of them were published starting in the year 2000 with a marked increase in 2006-2007. Among the measures published earlier in the 1990's, is the Assessment of Motor and Process Skills (AMPS) which is also one of the most frequently used in subsequent studies, after the Functional Independence Measure transformed scores (Table 5.10).

There were 26 articles (7%) in which the study made use of an already Rasch-developed measure. Among the measures used are: previously Rasch-analyzed Functional Independence Measure subscales (9 articles), the Assessment of Motor and Process Skills (4 articles), The Test of Playfulness (2 articles). Other measures, that were each used in only 1 article include: The Pediatric Evaluation of Disability Inventory (PEDI) Function Skills and Self-care domains, the ABILHAND-Kids, the Child and Adolescent Scale of Participation (CASP) the Craig Handicap Assessment and Reporting Technique (CHART), the Supervision Rating Scale (SRS), the Activity Measure for Postacute care (AM-PAC) the Test of Infant Motor Performance (TIMP), the Mayo-Portland Adaptability Inventory (MPAI), the Motor Readiness Questionnaire for stroke (MRQS), the Medical Rehabilitation Follow Along (MRFA) and two Avlund mobility scales, mobility-tiredness and mobility-help. A description of the studies is presented in Table 5.11, including the populations in which the measures were used.

Conclusion

Rasch analysis and item response theory, also called 'modern psychometric methods', are increasingly being applied in the field of rehabilitation. In 1991, two publications demonstrating the use of Rasch analysis or IRT models in the field of rehabilitation appeared; in 2006, a total of fifty-five were found. As professionals become more aware of the advantages of using these techniques to measure outcomes of rehabilitation, ordinal indices that were previously developed using classical test theory have been subjected to latent trait analyses to convert measurement scales to have interval-like properties. New measures have also been developed using these modern psychometric methods particularly in neurologic populations such as stroke, brain injury, and cerebral palsy. An important number of measures are also being developed and refined for persons with rheumatoid arthritis, vision impairments, and back pain.

On the other hand, very few studies are using measures that have already been Rasch-developed. This may be due to the fact that very few Rasch-developed measures have been published, psychometrically assessed, and ready to be used for research purposes. The use of these models is relatively new in the field of rehabilitation and it may take some time before the new measures are used routinely utilized as part of research projects.

These modern psychometric mothods have a great potential to transform the way outcome measurements are being developed and administered in the field of rehabilitation. Using linear measures instead of the traditional raw scores obtained from the current tests and indices permits the calculation of true change within individuals and between groups, thus allowing for the evaluation of the effectiveness of the treatments that are being delivered. This is of the utmost importance especially at a time when therapeutic interventions are being evaluated and scrutinized in terms of their effectiveness and the spending of health care dollars needs to be justified. Moreover, these modern psychometric techniques can resolve many of the current problems that traditional measurement tests and indices cannot. Rasch analysis and Item Response Theory make it possible to quantify latent traits such as patient's feelings about their quality of life, the difficulties they face when accomplishing certain tasks, and the impact their disabilities have on their level of participation in life activities. This is exemplified by the analyses and creation of such scales as the Burden of Stroke Scale (BOSS),[182] the Stroke Impact Scale (SIS),[183]and Self-Assessment Instrument for Measuring Independent [184] mobility which assesses perceived visual ability for independent mobility. These measures are designed to take into account clients' feelings about their disabilities and the difficulties they are encountering.

Some of the measures presented are psychometrically sound and ready to be used while others still require changes and further testing of their psychometric properties. In a stroke population (Table 5.1), the Rivermead Mobility Index,[185] the Catherine Bergego (unilateral Neglect)[186] the Functional Independence Measure motor subscale,[187] the Stoke Impact Scale-16,[90,188] the EG-Motor Index,[189] the Trunk Impairment Scale,[190] the Repeatable Battery for Assessment of Neuropsychological Symptoms (cognitive ability)[191], and the Stroke Impairment Assessment Set [192,193]are all ready to be used. The P-Drive, to assess driving ability,[194] the ABILHAND, a measure of manual ability in everyday activities [195]as well as a comprehensive assessment of ADL function obtained by combining the items of the Barthel Index and Frenchay Activities Index [196]can also be used in a stroke population.

For brain injury (Table 5.3), measures such as the Galveston Orientation and Amnesia Test [197] to assess post-traumatic amnesia, Mayo-Portland Adaptability Inventory (MPAI-4) and its sub-scales are ready to be used to assess impairment, activity, and participation.[198] The Cognistat [199] can only be used to classify multiple levels of cognitive status in both acute and postacute traumatic brain injury (TBI) settings. The 8-item Participation Index (M2PI) of the Mayo-Portland Participation Index can be used with special attention to rater biases.[200] The Middlesex Elderly Assessment of Mental State[201,202] can be used in a Turkish population as a cognitive impairment screening tool.

In a cerebral palsy population (Table 5.4), the ABILHANDS-Kids, is ready for use to assess manual ability.[203] Also ready to be used are the Gross Motor Function Measure-66 (for children) [204-206] and the WeeFIM (motor function in children).[207]

In patients with Multiple Sclerosis (MS) (Table 5.5), the Multiple Sclerosis Spasticity Scale (impact of spasticity)[208] needs to be further examined and tested for responsiveness. The EQUI-SCALE [209] can be used to assess balance as in a study by Cattaneo in 2002. [210]

In a population with Parkinson's disease, none of the measures presented, the Impact on Participation and Autonomy (IPA) questionnaire,[211] the Parkinson's disease Questionnaire (PDQ-39) and the Nottingham Health Profile are ready for use. Although promising, they all require further testing to be considered suitable.

For low-vision patients, the Assessment of Function Related to Vision,[212] appears valid but advantages and disadvantages remain to be evaluated and modification to item(s) may be required. The Melbourne Low-Vision ADL Index [213] and the Impact of Vision Impairment Scale [214,215] are all ready to be used with patients presenting with vision impairments although the sensitivity of the IVI subscales should be assessed (Table 5.7). As well, an activity index comprising 41 activity goals and its telephone version [216,217]can be administered to evaluate vision disability and limitations in functional vision. The American Medical Association's (AMA) vision disability scale [177]is based only on preliminary data. The 6-tiem VisQOL[218] to assess vision and Quality of Life, the 48-Item Veterans Affairs Low-Vision Visual Functioning Questionnaire [219-221] and the Self-Assessment Instrument for Measuring Independent mobility [184]are ready to be administered. As for the Visual Function-10, authors do not mention whether the shortened version is sufficiently psychometrically sound.[222]

In the back pain population (Table 5.8), the National Health and Nutrition examination Surveys Activities of Daily Living Instrument [223] needs to be evaluated for test-retest reliability and responsiveness. The Low Back SF-36 PF18's measurement properties need to be tested on an independent sample[224] and the Quebec Back Pain Disability Scale [181] should be tested under different conditions and populations. The Roland-Morris Disability Questionnaire [225] can be used in Turkey but has not been tested across other cultures. The Oswestry Low Back Pain Questionnaire should include higher level items and be tested on other samples.[226,227]

In a population with arthritis or other inflammatory diseases, several measures are ready for use (Table 5.9). They are, two multiattribute preference weight functions,[228] the

Cedars-Sinai Health-Related Quality of Life in Rheumatoid Arthritis (although authors suggest that a prospective validation in clinical trial settings is warranted), the ABILHAND,[229] the Short Arthritis Assessment Scale SAS[230]and the Foot Impact Scale.[231] The Turkish version of the Stanford Health Assessment Questionnaire [232] is ready for use in international studies between the United Kingdom and Turkey only. The Joint-Specific Multidimensional Assessment of Pain [233] can be used but only for osteoarthritis of the knee. The Health Assessment Questionnaire II[234] can be used to measure functional status in rheumatology and the Regional Pain Scale to identify patient with fibromyalgia (FM) or develop a new definition of FM.[235] Finally, an item bank was also created to measure Quality of Life in 5 domains relevant to arthritis.[236]

The field of rehabilitation will benefit greatly from using true measures to assess clients. An effort needs to be deployed in order for the new measures to reach the clinics as well as the researchers. Rehabilitation professionals need to be informed of these new trends. A debate is ongoing as the best ways of translating knowledge but until they are found, researchers will need to present their findings on Rasch analysis and Item Response Theory in conferences and in journals to insure the rehabilitation community is exposed to these new measurement trends. It is to be expected to take a few years before the rehabilitation community embraces measures developed using modern psychometrics.

Figure 5.1 Number of articles by year of publication

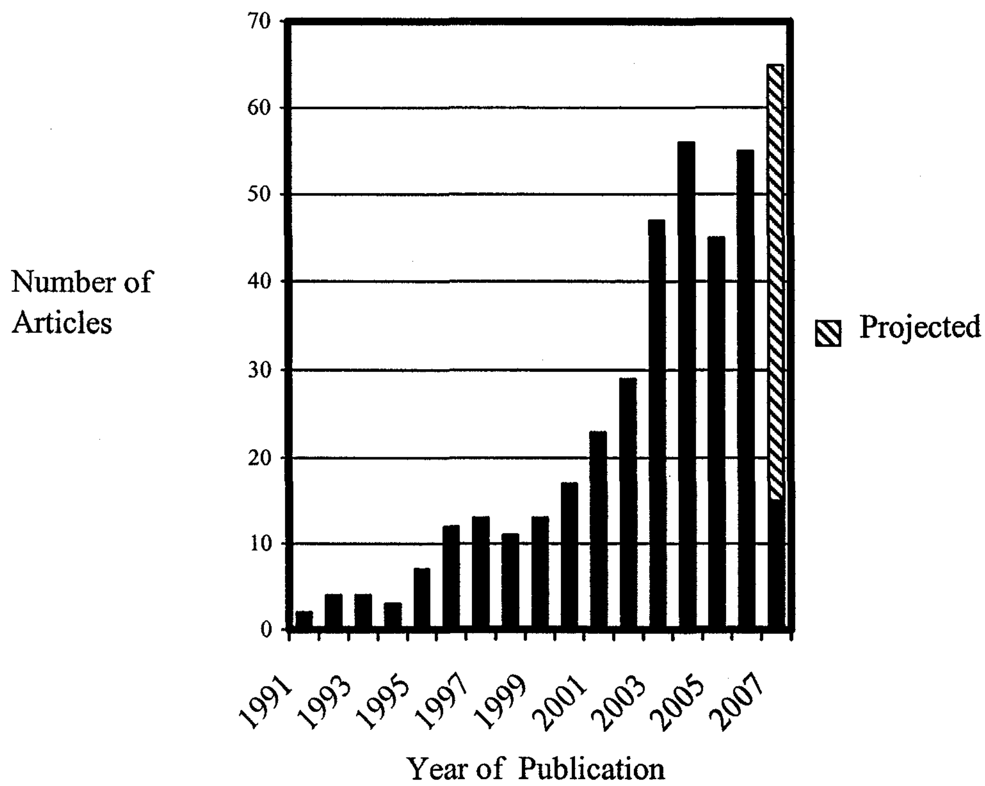


Table 5.1 Measures Developed for Stroke, or Used in a Stroke Population.

Authors	Year	Measure(s)/Use	Study Conclusions*	Aim
Aamodt G., et al.[237]	2006	Motor Assessment Scale (MAS)/ To measure motor functioning for stroke patients.	The scalability of the different items is not optimal. For some items, the top and/or bottom levels are overrepresented and for other items the middle levels tend to cluster. There seems to be a single construct except for one of the items. The instrument also showed the same meaning for the different subgroups.	P
Antonucci G., et al.[185]	2002	Rivermead Mobility Index (RMI)/ To measure mobility status.	The RMI is a unidimensional scale with a hierarchy of easy-to-hard test questions. Item difficulty level was stable when processed on different groups of patients assessed on different occasions.	P
Azouvi P., et al.[186]	2003	Catherine Bergego Scale (CBS) for Unilateral Neglect/ To assess spatial neglect in everyday life.	Behavioral assessment proved to be more sensitive than conventional paper and pencil tasks. Both conventional statistics and Rasch analysis suggest that the CBS is reliable and valid, and that the 10 items define a homogeneous construct.	P
Bjorkdahl A., et al.[238]	2004	European Brain Injury Questionnaire/ To capture the social, emotional, and cognitive impacts of a stroke.	With some alterations the European Brain Injury Questionnaire seems to be a useful instrument in clinical practice and research, and a help to capture the social, emotional, and cognitive impacts of a stroke.	P

Table 5.1 Measures Developed for Stroke, or Used in a Stroke Population. (cont.).

Authors	Year	Measure(s)/Use	Study Conclusions*	Aim
Brock KA., et al.[187]	2002	Functional Independence Measure (FIM) motor subscale/ To determine quality of outcome for poststroke rehabilitation.	Rasch estimates of the FIM motor subscale provide a discriminative measure for evaluating outcomes and change in ability achieved in stroke rehabilitation.	P
Doyle PJ., et al.[239]	2005	-Burden of Stroke Scale (BOSS)-Stroke Impact Scale (SIS)-World Health Organization's Disability Assessment Schedule (WHO-DAS)/ To measure communicative functioning.	(a) A mature understanding of communicative functioning as a measurement construct will require further research, (b) patients with stroke related communication disorders will be better served by the development of instruments measuring a wide range of communicative functioning ability, and (c) the theoretical and methodological tools provided by the Rasch family of the measurement models may be productively applied to these efforts.	P

Table 5.1 Measures Developed for Stroke, or Used in a Stroke Population. (cont.).

Authors	Year	Measure(s)/Use	Study Conclusions*	Aim
Duncan PW., et al.[188]	2003	The Stroke Impact Scale (SIS)-16/ To assess physical function approximately 1 to 3 months post-stroke.	Compared to the Barthel Index (BI), the SIS-16 is an excellent collection of items suitable for assessing a wide range of physical function limitations of patients with stroke at 1 to 3 months post-stroke. Because of a less pronounced ceiling effect, the SIS-16 can differentiate lower levels of disability as compared to the BI.	N
Duncan PW., et al.[90]	2003	The Stroke Impact Scale (SIS)-16/ To assess a wide range of physical function limitations 1-3 months post-stroke.	Rasch analysis further established the validity of the SIS. The domains are unidimensional, the items have an excellent range of difficulty, and the domain scores differentiated patients into multiple strata. The activities of daily living/instrumental activities of daily living, mobility, strength, composite physical, and participation domains have the most robust psychometric characteristics. The composite physical domain is most able to discriminate difficulty in function in individuals after stroke, while the communication, memory, and emotion domain items only capture limitations in function in the more impaired groups of patients.	P
Engberg A., et al.[189]	1995	EG Motor Index/ To assess mobility.	The final rating scale has three special characteristics: 1) it reflects the regularity in the recovery of mobility after stroke; 2) the sum of item scores comprises the information contained in the 10-item subscores; 3) the score sum is independent of age, side of hemiparesis, and gender of the patient. Latent trait analysis (Rasch) was found to be an ideal model for statistical investigation of these properties.	N

Table 5.1 Measures Developed for Stroke, or Used in a Stroke Population. (cont.).

Authors	Year	Measure(s)/Use	Study Conclusions*	Aim
Finch L., et al.[240]	2007	The Functional Recovery Measure/ To quantify function as represented by the ICF.	A 12-item measure captured the concept of functioning that could be used as a prototype to quantify recovery post-stroke. These items could form the basis for a measure of functioning.	N
Fujiwara T., et al.[190]	2004	Trunk Impairment Scale (TIS)/ To assess trunk impairment after stroke.	Our newly developed TIS is reliable, valid and responsive for use in stroke outcome research.	P
Heinemann AW., et al.[241]	1997	National Institute of Health (NIH) Stroke Scale/ To characterize stroke impairments.	The items in a revised NIH Stroke Scale worked well together to define the severity of impairment resulting from stroke that is observed during medical rehabilitation. Directions regarding limb ataxia should be modified to indicate untestability due to hemiplegia.	P
Hsueh IP., et al.[196]	2004	-Barthel Index (BI)-Frenchay Activities Index (FAI)/ To assess comprehensive Activities of Daily Living (ADL) function in stroke patients.	A clinically useful assessment of the comprehensive ADL function of patients at or later than 1 year after stroke can be obtained by combining the items of the BI and FAI (excluding 2 FAI items) and simplifying the responses into dichotomous categories. It is also demonstrated that the items of the new scale measure comprehensive ADL function as a single unidimensional construct when assessed at 1 year after stroke.	N

Table 5.1 Measures Developed for Stroke, or Used in a Stroke Population. (cont.).

Authors	Year	Measure(s)/Use	Study Conclusions*	Aim
Hsueh IP., et al.[242]	2006	STroke REhabilitation Assessment of Movement (STREAM)/ To assess motor function.	The S-STREAM demonstrates high Rasch reliability, unidimensionality, and concurrent validity with the STREAM in patients with stroke. Furthermore, the S-STREAM is efficient to administer, as it consists of only half the number of items in the original STREAM. Additional studies to examine other psychometric properties (eg, predictive validity and responsiveness) of the S-STREAM or its psychometric properties in various recovery stages after stroke are needed to further establish its utility in both clinical and research settings.	N
Hula W., et al.[243]	2006	Revised Token Test Performance (RTT)/ To assess auditory processing impairments in adults with aphasia.	The findings generally support the RTT's validity, but a minority of items appears to respond to a different construct. Also, within-subtest differences in item difficulty suggest the need for further examination of variability in impaired language performance. Finally, the results suggest an equivocal advantage for Rasch scores in detecting change over time.	P
Jette AM., et al.[244]	2003	-FIM-OASIS (Outcome and Assessment Information Set for Home Health care)-MDS (Minimum dataset for skilled nursing and subacute rehabilitation.-PF-10/ To assess functional outcome.	This article presents an empirical comparison of four functional outcome instruments used in Post Acute Care (PAC) with respect to their content, breadth of coverage, and measurement precision. Results illustrate limitations in the range of content, breadth of coverage, and measurement precision in each outcome instrument. None appears well-equipped to meet the challenge of monitoring quality and functional outcomes across settings where PAC is provided. Limitations in existing assessment methodology has stimulated the development of more comprehensive outcome assessment systems specifically for monitoring the quality of services provided to PAC patients.	P

Table 5.1 Measures Developed for Stroke, or Used in a Stroke Population. (cont.).

Authors	Year	Measure(s)/Use	Study Conclusions*	Aim
Koh CL., et al.[245]	2006	Action Research Arm Test (ARAT)/ To assess upper extremity function.	Our findings indicated that the 19-item ARAT constituted a unidimensional construct measuring upper-extremity function in stroke patients. However, the results did not support the premise that the raw sum scores of the ARAT can be transformed into interval Rasch scores. Thus, the raw sum scores of the ARAT can provide information only about order of patients on their upper extremity functional abilities, but not represent each patient's exact functioning.	P
Lai J., et al.[246]	2003	-Stroke Impact Scale (SIS)-SF-36/ To assess physical functioning, and social well-being.	Both the physical and participation subscales of the SIS cover a wider range of item difficulty than their counterparts from the SF-36. Compared with the SF-36 PF and SF domains, the SIS-16 and SIS Participation are better able to capture physical functioning and social well-being in patients with strokes.	P
Larson EB., et al.[191]	2003	Repeatable Battery for Assessment of Neuropsychological Symptoms (RBANS)/ To assess cognitive ability.	The present findings support the use of cognitive evaluations of patients with acute stroke to assist with prediction of outcome to be used in treatment planning.	P
Linn RT., et al.[247]	1999	Functional Assessment Measure (FAM) Functional Independence (FIM)/ To assess functional status.	Results indicate that use of the entire FAM as an adjunct to the FIM reduces test efficiency while providing only minimal additional protection against ceiling effects.	P
Liu M., et al.[193]	2002	Stroke Impairment Assessment Set (SIAS)/ To assess stroke impairment.	The SIAS is a useful measure of stroke impairment with well-established psychometric properties.	P

Table 5.1 Measures Developed for Stroke, or Used in a Stroke Population. (cont.).

Authors	Year	Measure(s)/Use	Study Conclusions*	Aim
Lundgren-Nilsson A., et al.[248]	2005	Functional Independence Measure (FIM)/ To assess functional status in different countries.	Clinical collected data from FIM for patients with stroke cannot be pooled in its raw form, or compared across countries. Comparisons can be made after adjusting for country-specific Differential Item Functioning, through the adjustments for Differential Item Functioning and rating scales may not generalize to other samples.	P
Millis SR., et al.[249]	2007	National Institutes of Health Stroke Scale (NIHSS)/ To detect neurologic impairment.	Our findings provide initial support for the use of individual, targeted scales for measurement of impairment after ischemic stroke. Low person separation reliability may be a consequence of the sample, which included only people with large ischemic cortical strokes.	P
Patomella AH., et al.[194]	2006	Performance Analysis of driving Ability (P-Drive)/ To assess the quality of the participant's driving performance in a simulator.	The findings from this study indicated that P-Drive is an assessment tool with properties of internal scale validity, person response validity, and which also contains aspects of reliability in relation to precision of the estimates and separation. P-Drive seems to be a valid and stable assessment tool for assessing the driving ability in a simulator of people with stroke.	P
Penta M., et al.[195]	2001	ABILHAND questionnaire/ To measure manual ability in everyday activities.	The ABILHAND questionnaire results in a valid person-centered measure of manual ability in everyday activities. The stability of the item-difficulty hierarchy across different patient classes further supports the clinical application of the scale to measure manual ability in everyday activities, the ABILHAND is valid. The item-difficulty hierarchy is stable which supports the clinical application of the scale.	P

Table 5.1 Measures Developed for Stroke, or Used in a Stroke Population. (cont.).

Authors	Year	Measure(s)/Use	Study Conclusions*	Aim
Sabari JS., et al.[250]	2005	Motor Assessment Scale for Stroke (MAS)/ To assess upper extremity function.	The findings lead to suggestions for changes in the behavioral criteria hierarchy for upper-limb items on the MAS and highlight the importance of using statistical analyses to test the validity of proposed hierarchies of behavioral criteria in functional assessments to assess upper extremity function, changes are suggested to the behavioral criteria hierarchy for upper-limb items on the MAS.	P
Segal ME., et al.[251]	1997	The Physical Ability Stroke Scale (PASS)/ To measure physical ability for long-term outcomes.	In conclusion, Rasch analysis was useful in identifying a brief, 10-item physical ability scale appropriate for follow-up of this sample of stroke patients. The next step is to validate the PASS on a larger sample of patients, perhaps including those with diagnoses other than stroke, to determine its general applicability for rehabilitation follow-up.	N
Tennant A., et al.[252]	2004	-Mini-Mental State Examination (MMSE) -Functional Independence Measure Motor Scale (FIM)/ To use the FIM and the MMSE (cognition) in international studies.	Where data are to be pooled for international studies, analysis of Differential Item Functioning (DIF) by culture is essential. Where DIF is observed, adjustments can be made to allow for cultural differences in outcome measurement.	P

Table 5.1 Measures Developed for Stroke, or Used in a Stroke Population. (cont.).

Authors	Year	Measure(s)/Use	Study Conclusions*	Aim
Tennant A., et al.[253]	1996	Barthel Index/ To assess activities of daily living.	Calibration of the 10 items in the Index shows considerable differences in the degree of difficulty (weight), and these differences are not compensated for by the current scoring. Thus adding together the items produces a scale whose intervals vary considerably, particularly between intervals at the lower or upper ends of the scale, and those at the centre. This can give rise to considerable differences between the change score based on the Rasch transformation (taking into account item difficulty) and the change score based on raw scores. These findings confirm the ordinal nature of the Barthel Index. Further questions are raised about the unidimensionality of the Index, and the context in which it should be used.	P
Tham K., et al[254]	1999	Assessment of Awareness of Disability (AAD)/ To assess awareness of disability.	The results indicated that AAD measures a single construct and that it discriminates among clients with different levels of awareness of disability. While the results of this pilot study are promising, the validity and reliability of the AAD need to be further examined on a larger and more varied group of clients.	N
Tsuji T., et al.[192]	2000	Stroke Impairment Assessment Set (SIAS)/ To assess impairment from stroke.	The unidimensionality of the SIAS was confirmed, and the SIAS total scores proved useful for stroke outcome prediction.	P

Table 5.1 Measures Developed for Stroke, or Used in a Stroke Population. (cont.).

Authors	Year	Measure(s)/Use	Study Conclusions*	Aim
Tsuji T., et al.[255]	1995	Functional Independence Measure (FIM)/ To use the FIM in Japan.	The item difficulty patterns in Japan differ slightly from those in the United States because of cultural differences. As countries show different patterns of difficulty, we must be careful when making international comparisons of FIM data converted by Rasch analysis.	P
Van de Winckel A.[256]	2006	Motor evaluation Scale for Upper Extremity in Stroke Patients (MESUPES)/ To assess arm function.	The MESUPES-arm and MESUPES-hand meet the statistical properties of reliability, validity and unidimensionality. Both tests provide a useful clinical and research tool to qualitatively evaluate arm and hand function during recovery after stroke.	N
Van Hartingsveld F., et al.[257]	2006	Barthel Index/ To assess activities of daily living.	We improved the psychometric properties and clinical interpretation of the Barthel index.	P

N = Development of a new measure; P = Determine the psychometric properties of an existing test or index or improve an existing test or index.

*Conclusions are a direct citation from the authors in each article.

Table 5.2 Generic Measures Developed or Psychometrically Assessed in Stroke Along with Other Conditions.

Authors	Year	Populations	Study Conclusions*	Aim
Bode RK., et al.[30]	2003	-Persons being treated for or having a history of cancer -HIV -Stroke -Multiple Sclerosis	The illustrations show that (1). core items, which functioned similarly across 4 diagnostic groups, can be identified and used to construct instruments measuring physical function that are tailored to each of these groups, and (2). items from 3 separate datasets can be linked to create a dataset that can serve as an initial pain item bank.	P
Dallmeijer AJ., et al. [258]	2007	-Stroke-Multiple Sclerosis -Amyotrophic lateral sclerosis	SF-36 physical functioning scale (PF10)./All items of the PF10, except one for the amyotrophic lateral sclerosis group, form a unidimensional scale, supporting the use of a sum score as a measure of physical functioning within these diagnostic groups. When comparing the data of patients after stroke, with that of patients with multiple sclerosis and/or amyotrophic lateral sclerosis patients, adjustments for differential item functioning are required.	P
Dallmeijer AJ., et al.[259]	2005	-Stroke-Multiple Sclerosis -Traumatic Brain Injury	Functional Independence Measure/ Differential Item Functioning is present in several items of both the motor and cognitive domain of the FIM™. Adjustments for differential item functioning may be required when FIM™ data will be compared between groups or will be used in a pooled data analysis.	P
Fisher WP.Jr[1]	1995	-Brain injury-neuromuscular -Musculoskeletal-Spinal cord -Stroke	-Functional Independence Measure-Patient Evaluation and Conference System/Either instrument's ratings are easily and quickly converted into the other's using the common unit of measurement, the rehabit (rehabilitation measuring unit). This article argues that the stability of the PECS and FIM item difficulty estimates over thousands of subjects, dozens of hospitals, hundreds of raters, and years of assessment is convincing evidence in support of the widespread use of their cocalibrated, common scale values as a functionometric ruler.	N

Table 5.2 Generic Measures Developed or Psychometrically Assessed in Stroke Along with Other Conditions. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
Gandek B., et al.[260]	2006	-Neurologic-Musculoskeletal-Medically complex	Participation Measure for Post-Acute Care (PM-PAC)/ PM-PAC is a promising new measure of patient-reported participation as defined by the International Classification of Functioning, Disability, and Health. Information about PM-PAC items from this study will be useful in developing a computerized, adaptive measure of participation.	N
Hart DL., et al.[261]	2002	-Spinal Impairments -Knee Impairments -Upper and Lower extremity impairments -Cervical and lumber spine impairments -Stroke -Orthopedic conditions -Pain syndromes-Neurologic conditions	An index of physical functional health status (FHS)/ Results support the reliability and validity of FHS-36 measures in the present sample. Analyses show the potential for a dynamic, computer-controlled, adaptive survey for FHS assessment applicable for group analysis and clinical decision making for individual patients.	N
Heinemann AW., et al.[262]	1993	Various impairment groups including -Stroke -Brain dysfunction -Spinal cord dysfunction -Amputation -Arthritis etc...	Functional Independence Measure (FIM) (Scaled measures/ Adequate clinical precision of the FIM was demonstrated, though suggestions for improvement emerged. The frequency of misfit between patients and the performance scales varied across impairment groups, but was acceptable. The results of this project will enable clinicians and researchers to plan cost-effective treatment by providing a valid measure of disability.	P

Table 5.2 Generic Measures Developed or Psychometrically Assessed in Stroke Along with Other Conditions. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
Holman R., et al.[263]	2005	-Residents of Supported Housing -Chronic Pain-Stroke inpatients -Parkinson's disease	Academic Medical Center Linear Disability Score (ALDS)/ The Academic Medical Center linear disability score item bank has promising measurement characteristics for the mixed patient population described in this paper. Further studies will be needed to examine the measurement properties of the item bank in other populations.	P
Kirby RL., et al [264]	2002	Wheelchair users (amputation, stroke, musculoskeletal disorders, spinal cord injury, neuromuscular disorders).	Wheelchair Skills Test (WST)./ The WST is practical, safe, well tolerated, exhibits good to excellent reliability, excellent content validity, fair construct and concurrent validity, and moderate usefulness. This pilot study makes an important contribution toward meeting the need for a well-validated outcome measure of manual wheelchair ability.	P
Kucukdeveci AA., et al.[265]	2001	-Stroke-Spinal Cord Injury	Adaptation of Functional Independence Measure in Turkey (FIM)/ Adaptation of the FIM has been successful and it can be used in Turkey as long as the limitations are recognized.	P
Kucukdeveci AA., et al.[266]	2000	-Stroke -Spinal cord injury	Modified Barthel Index /In conclusion, adaptation of the modified Barthel Index has been successful and it can be used in Turkey as long as its limitations are recognized.	P

Table 5.2 Generic Measures Developed or Psychometrically Assessed in Stroke Along with Other Conditions. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
Lundgren-Nilsson A., et al.[267]	2006	-Stroke-Spinal cord injury-traumatic brain injury	Functional Independence Measure (FIM)in Scandinavia/ The seven-category response function is a problem for the FIM instrument, and a reduction of responses might increase the validity of the instrument. Likewise, the removal of items that do not fit the underlying trait would improve the validity of the scale in these groups. Cross-diagnostic Differential Item Functioning (DIF) is also a problem but for clinical use sum scores on group data in a generic instrument such as the FIM can be compared with appropriate adjustments. Thus, when planning interventions (group or individual), developing rehabilitation programs or comparing patient achievements in individual items, cross-diagnostic DIF must be taken into account.	P
Nilsson AL., et al.[268]	2005	Neurological disorders: -Stroke -Spinal Cord Injury -Traumatic Brain Injury	Rasch-transformed Functional Independence Measure/ Decreasing the categories from seven to four may be one way of dealing with problems of disordered thresholds. Further studies are also needed in order to try the suggested scale in clinical settings and to compare it with the original FIM scale.	P

Table 5.2 Generic Measures Developed or Psychometrically Assessed in Stroke Along with Other Conditions. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
Ostir GV., et al.[269]	2006	Inpatient rehabilitation: -Orthopedic-Stroke -Neurologic-Brain injury -Cardiac and pain-Arthritis and others	Instrument of home and community participation (PAR-PRO)/ The 20-item PAR-PRO instrument of home and community participation displayed good psychometric characteristics. The instrument shows promise as a broad measure of home and community involvement for persons with disabilities. Further work is needed to support its application for people without disability.	N
Silverstein B., et al.[270]	1992	-Neuromuscular -Brain Injury -Spinal Cord Injury -Stroke -Musculoskeletal	Patient Evaluation and Conference System (PECS)/ The analyses were performed to determine the extent to which four item subsets identified in a previously reported factor analysis of the PECS comprise interval measures of functional independence status. Results indicate that the PECS scales meet these requirements to varying degrees. The analyses also identify areas in which measurement quality can be improved.	P
Smith RM., et al.[271]	2004	Patients admitted to : -Brain Injury -Neuromuscular -Musculoskeletal-Spinal Cord Injury -Stroke programs	14 Functional Independence Measure Motor Skills items- Patient Evaluation and Conference System (PECS) Motor Skills LifeScale/ The results indicate that it is possible to construct a common equal interval translation between the PECS and FIM for the two scales. Measures on the common metric can be based to either scale and are independent of the number of items completed. This use of these anchored scales will allow institutions using either the PECS and FIM to make direct comparisons of clinical outcomes with other institutions, independent of the particular outcome tool used to evaluate patients.	P

N = Development of a new measure; P = Determine the psychometric properties of an existing test or index or improve an existing test or index.

*Conclusions are a direct citation from the authors in each article.

Table 5.3 Measures Developed for Brain Injury, or Used in a Brain Injury Population.

Authors	Year	Measure(s)/Use	Study Conclusions*	Aim
Bedell GM., et al.[272]	2004	Child and Family Follow-up Survey (CASP)/ To monitor needs and outcomes of children and youth with acquired brain injury.	The results presented are preliminary yet promising. The follow-up survey is being further developed and tested. Future research also will investigate clinical utility, differences in scores among children and youth with different diagnostic conditions, and feasibility of creating separate versions of the participation measure for different age groups.	N
Bedell GM., et al.[273]	2002	Prototype clinical performance measure/ To assess change in performance in pediatric inpatient rehabilitation.	Findings provide evidence that the activity scales have the potential to be responsive and point to issues that will need to be addressed in future measurement development.	N
Bode RK., et al.[197]	2000	Galveston Orientation and Amnesia Test (GOAT)/ To assess post-traumatic amnesia.	Equal-interval measures of Post-Traumatic Amnesia (PTA) were developed that exhibited good reliability and validity. A self-scoring key was developed to more efficiently assess PTA.	P
Cantagallo A., et al [274]	2006	MINDFIM/ To measure disability in high functioning outpatient.	MINDFIM is a valid and useful measure of disability in Traumatic Brain Injury (TBI) outpatients	N

Table 5.3 Measures Developed for Brain Injury, or Used in a Brain Injury Population. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
Doninger NA., et al.[199]	2006	Neurobehavioral Cognitive Status Examination (Cognistat)/ To classify multiple levels of cognitive status in both acute and postacute Traumatic Brain Injury settings.	As a screening instrument, the Cognistat reliably classifies multiple levels of cognitive status in both acute and postacute Traumatic Brain Injury settings; however, this measure is unsuitable for generating a profile of neurocognitive strengths and weaknesses.	P
Elhan AH., et al.[275]	2005	Mini-Mental State Examination (MMSE)/ To screen cognitive status.	The Turkish version of the Mini-Mental State Examination can be used as a cognitive screening tool in acquired brain injury. Cross-cultural validity between Italy and Turkey is supported, given appropriate adjustment for differential item functioning. However, shortfalls in reliability at the individual level, as well as the presence of differential item functioning suggest that a better instrument should be developed to screen for cognitive deficits following acquired brain injury.	P

Table 5.3 Measures Developed for Brain Injury, or Used in a Brain Injury Population. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
Hart T., et al.[276]	2006	Moss Attention Rating Scale (MARS)/ To assess and study different dimensions of disordered attention in acute patients, and to monitor change over time and treatment response.	The factor-scored MARS has potential utility as a quantitative observational method with which to assess and study different dimensions of disordered attention in acute Traumatic Brain Injury, and to monitor change over time and treatment response within these dimensions.	P
Johnston., et al.[277]	2006	Glasgow Coma Scale (GCS) Functional Independence Measure (FIM) Glasgow Outcome Scale (GOS) Community Integration Questionnaire (CIQ) Disability Rating Scale (DRS) Neurobehavioral Functioning Inventory (NFI) Supervision Rating Scale (SRS)/ To assess global outcomes.	Actual dimensionality was distinct from the named scales employed. A unidimensional measure model fit the data much better than expected. This outcome dimension might be called 'general community functioning'. In the future, it should be possible to develop more valid and parsimonious measures of community outcomes following Traumatic Brain Injury.	P

Table 5.3 Measures Developed for Brain Injury, or Used in a Brain Injury Population. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
Kothari DH., et al.[278]	2003	Pediatric Evaluation of Disability Inventory (PEDI)/ To measure functional change in a pediatric population.	The Acquired Brain Injury (ABI)-specific scales added relatively little improvement in sensitivity compared with the generic PEDI scales of the Mobility and Self-care domains. Thus, for group analyses, the authors recommend use of the generic PEDI subscales for children with ABI. Future work with ABI-specific subscales may improve the physical therapist's ability to describe an individual's pattern of functional recovery.	N
Kutlay S., et al.[201]	2007	Middlesex Elderly Assessment of Mental State (MEAMS)/ To screen for cognitive impairments.	The reliability and validity of the Turkish version of MEAMS as a cognitive impairment screening tool in acquired brain injury has been demonstrated.	P
Malec JF., et al.[279]	2000	Mayo-Portland Adaptability Inventory (MPAI)/To assess impairment, activity, and participation.	Impairment, activity, and participation define a single dimension of brain injury sequelae. The MPAI shows promise as a measure of this construct.	P

Table 5.3 Measures Developed for Brain Injury, or Used in a Brain Injury Population. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
Malec JF., et al.[200]	2004	8-item Participation Index (M2PI) of the Mayo-Portland Participation Index (MPAI)/ To assess impairment, activity, and participation.	The M2PI, particularly in composite indices and with attention to rater biases, provides an outcome measure with satisfactory psychometric qualities and the potential to represent the varying perspectives of people with acquired brain injury, significant others, and rehabilitation staff.	P
Malec JF., et al.[198]	2004	Mayo-Portland Adaptability Inventory (MPAI-4) and sub-scales/ To assess impairment, activity, and participation.	The MPAI-4 possesses satisfactory internal consistency regardless of rating source. A composite measure based on ratings made independently by people with ABI, significant others and staff may serve as a 'gold standard' for research purposes. In the clinical setting, assessment of varying perspectives and biases may not only best represent outcome as evaluated by all parties involved but be essential to developing effective rehabilitation plans.	P
Malec JF., et al.[280]	2003	Mayo-Portland Adaptability Inventory (MPAI)/ To evaluate clinical cases and rehabilitation programs.	Outcome after Acquired Brain Ijury is represented by the unitary dimension described by the MPAI. MPAI subscales further define regions of this dimension that may be useful for evaluation of clinical cases and program evaluation.	P
Tennant A., et al.[202]	2006	Middelsex Elderly Assessment of Mental State (MEAMS}/ To assess cognitive state for the adult Turkish population.	The MEAMS offers a valid assessment of cognitive state for the adult Turkish population, and the revised cut points accommodate for age and education. Further studies are required to ascertain the validity in different diagnostic groups.	P

Table 5.3 Measures Developed for Brain Injury, or Used in a Brain Injury Population. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
Tesio L., et al.[281]	1998	Functional Assessment Measure (FAM)/ To assess functional status.	The FAM was on average too easy. The most difficult item (a new one, Employability) did not attain the average ability of the subjects. Also, it was only slightly more difficult than the most difficult FIMSM item (Memory). The FAM does not seem to improve the FIMSM as a far as Traumatic Brain Injury outpatients are to be assessed.	P
Whyte J., et al.[282]	2003	Moss Attention Rating Scale (MARS)/ To assess attention-related behaviors.	Our results provide preliminary support for the viability of developing an observational attention rating scale for use in inpatient TBI rehabilitation. Further research will need to explore the existence of subdimensions and provide further validation with reference to other neuropsychologic measures of attention and knowledge of lesion severity and localization.	P
Williams GP., et al.[283]	2005	High-level Mobility Assessment Tool (HiMAT)/ To assess higher-level mobility requirements of people with TBI for return to pre-accident social, leisure and sporting activities.	The HiMAT, which assesses higher-level mobility requirements of people with Traumatic Brain Injury for return to pre-accident social, leisure and sporting activities, is a uni-dimensional and discriminative scale for quantifying therapy outcomes. Further investigation is required to establish the reliability, concurrent validity and responsiveness of the HiMAT.	P

N = Development of a new measure; P = Determine the psychometric properties of an existing test or index or improve an existing test or index.

*Conclusions are a direct citation from the authors in each article.

Table 5.4 Measures Developed for Cerebral Palsy, or Used in a Population with Cerebral Palsy.

Authors	Year	Measure(s)/Use	Study Conclusions*	Aim
Arnould C., et al.[203]	2006	ABILHAND-Kids/ To measure manual ability in children with Cerebral Palsy and provide guidelines for goal setting in treatment planning.	The ABILHAND-Kids is a functional scale specifically developed to measure manual ability in children with Cerebral Palsy (CP) providing guidelines for goal setting in treatment planning. Its range and measurement precision are appropriate for clinical practice.	N
Avery LM., et al.[204]	2003	Gross Motor Function Measure (GMFM-66)/ To measure gross motor function for children with cerebral palsy.	The GMFM-66 is an interval-level measure of gross motor function for children with Cerebral Palsy (CP); it should improve the scoring, interpretation, and overall clinical and research utility over the original GMFM.	P
Chen CC., et al.[207]	2005	WeeFIM instrument/ To assess motor function in children with disabilities.	Similar to the adult Functional Independence Measure instrument, the WeeFIM instrument has two distinct dimensions. The motor items form a unidimensional construct with acceptable measurement properties. Developmental differences in motor task mastery among children with disabilities are assumed but rarely tested. As evidenced by the age-specific item hierarchies found in this study, developmental differences among children with disabilities mimic that of children without disabilities. Finally, the results suggest that clinicians attend to the differences in the WeeFIM item definitions and are able to use it accurately and effectively.	P

Table 5.4 Measures Developed for Cerebral Palsy, or Used in a Population with Cerebral Palsy. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
Grimby G., et al.[284]	1996	Physical items from Functional Independence Measure(FIM)-Instrumental Activity Measure (IAM)/To assess disability in community-living persons.	The combination of physical items from FIM and instrumental activity measure (IAM) are useful for disability assessment in community-living persons and should be further studied in other impairment groups.	P
Kottorp A., et al.[285]	2003	The Assessment of Motor and Process Skills (AMPS)/ To assess activities of daily living for people with developmental disabilities.	Further research is needed to evaluate the use of the AMPS in clinical assessment and intervention planning for this group of clients.	P
Russell DJ., et al.[286]	2000	66-item version of the Gross Motor Function Measure (GMFM)/ To assess motor function.	This study demonstrated that the GMFM-66 has good psychometric properties. By providing a hierarchical structure and interval scaling, the GMFM-66 can provide a better understanding of motor development for children with CP than the 88 item GMFM and can improve the scoring and interpretation of data obtained with the GMFM.	P

Table 5.4 Measures Developed for Cerebral Palsy, or Used in a Population with Cerebral Palsy. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
Shi W., et al[206]	2006	Gross Motor Function Measure (GMFM-66)/ To assess the gross motor functions of children <3 yrs old with cerebral palsy.	Results indicated that the GMFM-66 had good reliability and validity in assessing the gross motor functions of children <3 yrs old with Cerebral Palsy. The GMFM-73 derived in the present study did not function significantly better for young children than GMFM-66 to assess the gross motor functions of children <3 yrs old with cerebral palsy, GMFM-66 had good reliability and validity. The GMFM-73 derived in the present study did not function significantly better for young children than GMFM-66.	P

N = Development of a new measure; P = Determine the psychometric properties of an existing test or index or improve an existing test or index.

*Conclusions are a direct citation from the authors in each article.

Table 5.5 Measures Developed for Multiple Sclerosis, or Used in a Multiple Sclerosis Population.

Authors	Year	Measure(s)/Use	Study Conclusions*	Aim
Hobart JC., et al.[208]	2006	Multiple Sclerosis Spasticity Scale (MSSS-88)/ To measure the impact of spasticity.	The 88-item Multiple Sclerosis Spasticity Scale (MSSS-88) is a reliable and valid, patient-based, interval-level measure of the impact of spasticity in multiple sclerosis. It has the potential to advance outcomes measurement in clinical trials and clinical practice, and provides a new perspective in the clinical evaluation of spasticity.	N
Tesio L., et al.[209]	1997	EQUI-SCALE/ To assess balance.	The remaining 8 items made up a scale (called EQUI-SCALE) complying with the requirements of unidimensionality and reliability. The item scores remained stable in a sub-sample of 24 patients tested before and after ten 1-hour exercise sessions, thus supporting the homogeneity of the items.	N

N = Development of a new measure; P = Determine the psychometric properties of an existing test or index or improve an existing test or index.

*Conclusions are a direct citation from the authors in each article.

Table 5.6 Measures Developed for Parkinson's Disease, or Used in a Population with Parkinson's Disease (PD).

Authors	Year	Measure(s)/Use	Study Conclusions*	Aim
Franchignoni F., et al.[211]	2007	Impact on Participation and Autonomy (IPA) questionnaire/ To measure participation.	IPA-I shows promise as a tool for measuring participation in people with Parkinsons's Disease. IPA-II has acceptable psychometric characteristics for measuring perceived problems in participation. Additional steps to improve their metric properties and further studies in people with different kinds of health conditions need to be carried out.	P
Hagell P., et al.[287]	2003	-Parkinson's disease (PD) Questionnaire (PDQ-39) -Nottingham Health Profile/ To measure health status.	Although promising, both questionnaires warrant further developmental work and stronger support of measurement validity before they could be considered fully suitable for use in Parkinson's Disease, in particular in earlier stages of the disease.	P

N = Development of a new measure; P = Determine the psychometric properties of an existing test or index or improve an existing test or index.

*Conclusions are a direct citation from the authors in each article.

Table 5.7 Measures Developed for Vision Impairments, or used in a Population with Vision Impairments.

Authors	Year	Measure(s)/Use	Study Conclusions*	Aim
Altangerel U., et al.[212]	2006	Assessment of Function Related to Vision (AFREV)/ To assess visual functioning.	To AFREV performance-based measure, a new test of a spectrum of activities correlates well with some standard measures of visual function and certain aspects of self-report assessments. AFREV appears to be a valid measure of performance ability that may provide information not obtainable from standard measures of visual function or subjective surveys.	P
Babcock-Parziale J., et al.[288]	2005	-VA-132 -Functional Assessment of Self-Reliance on Tasks (FAST)/To assess low-vision and blind rehabilitation outcome.	Our conclusions show that the two instruments are currently not compatible for calibration; however, this can be improved with proper attention to scaling inadequacies, test administration times, and content coverage.	P
Haymes SA., et al.[213]	2001	Melbourne Low-Vision ADL Index (MLVSAI)/To assess activities of daily living.	The MLVAI is a highly valid and reliable standardized test of ADL performance for the general low-vision population. It may be used to assess patients with low vision and has the potential to be used as a measure of low-vision rehabilitation outcomes.	N
Lamoureux EL., et al.[214]	2006	Impact of Vision Impairment Scale (IVI)/To assess outcomes of low-vision rehabilitation.	The results provide support for the measurement properties of the Rasch-scaled 28-item version of the IVI and of its potential for assessing outcomes of low-vision rehabilitation. A raw score-to-Rasch person measure conversion is supplied.	P

Table 5.7 Measures Developed for Vision Impairments, or used in a Population with Vision Impairments. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
Lamoureux EL., et al.[215]	2007	Impact of Vision Impairment Scale (IVI)/ To assess the effects of different types of low-vision rehabilitation programs.	An examination of the IVI dimension confirmed a three-subscale structure that displays interval measurement characteristics likely to provide a valid and reliable assessment of restriction of participation. The findings provide an opportunity for a more detailed measurement of the effects of different types of low-vision rehabilitation programs.	P
Massof RW., et al.[216]	2005	41 activity goals (Activity Inventory (AI)/ To assess visual disability.	The 2 variables that define visual disability-value of independence and visual ability-are valid constructs that can be estimated accurately and reliably from patient ratings of the importance and difficulty of activity goals.	P
Massof RW., et al.[177]	2002	American Medical Association's (AMA) vision disability scale/ To assess visual disability.	A tutorial review of psychometrics-classical test theory, item response theory, and Rasch analysis-shows how vision disability measurement scales can be estimated from Likert-type visual function rating scales. We conclude that preliminary data relating measures of vision disability to measures of visual acuity and visual fields support the new AMA vision disability scale.	N
Massof RW., et al.[217]	2005	Telephone-administered Activity Inventory/ To assess visual disability.	Our study results confirm the hierarchical structure of the Activity Breakdown Structure model and show how the individualized Activity Inventory can produce measures of limitations in functional vision.	P

Table 5.7 Measures Developed for Vision Impairments, or used in a Population with Vision Impairments. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
Massof RW., et al.[289]	1998	A method (algorithm) of measuring vision disabilities/ To measure vision disabilities.	Rehabilitative demand, an algorithm for defining vision disability, incorporates interval measures of visual ability and the value of independent living estimated from patient-based assessments, a social scale of the value of activities relative to independent living, and a consensus opinion of low vision experts on the prioritization of the need for rehabilitation.	N
Misajon R., et al.[218]	2005	6-item VisQoL/ To assess vision and quality of life.	The short 6-item VisQoL has excellent psychometric properties as a simple summative instrument. It can be used in its present state as a condition-specific outcome measure for the evaluation of healthcare interventions for the visually impaired. The descriptive model is also suitable for generating utility values for the economic evaluation of vision-related programs and services.	N
Pesudovs K., et al.[290]	2003	Activities of Daily Vision Scale for Cataract Surgery Outcomes (ADVS)/ To assess activities of daily living in cataract surgery outcome.	Despite careful traditional validation, the ADVS data contained inadequacies exposed by Rasch analysis. Through Rasch scaling, particularly with response scale reduction, the ADVS can be improved, but additional questions seem to be needed to suit the more able, including patients undergoing second eye cataract surgery. There remains a need to develop Rasch-scaled measures of visual disability for use in ophthalmic outcomes research.	P
Stelmack J., et al.[219]	2004	48-ITEM VETERANS Affairs Low-Vision Visual Functioning Questionnaire (VA LV VFQ-48)/ To assess visual functioning.	Preliminary analysis indicates that the questionnaire items are applicable to persons of differing abilities. The Rasch person-item map demonstrates that the field-test version of the VA LV VFQ-48 has good range and is well centered with respect to the person measure distribution. Construct validity and reliability are also demonstrated.	N

Table 5.7 Measures Developed for Vision Impairments, or used in a Population with Vision Impairments. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
Stelmack J., et al.[220]	2004	48-ITEM VETERANS Affairs Low-Vision Visual Functioning Questionnaire (VA LV VFQ-48)/ To assess visual functioning.	The VA LV VFQ-48 is valid and reliable and has the range and precision necessary to measure visual ability of low-vision patients with moderate to severe vision loss across diverse clinical settings.	P
Szlyk JP., et al.[221]	2004	Veterans Affairs Low Vision Visual Functioning Questionnaire/ To measure vision difficulties in persons who receive low-vision rehabilitation.	The results demonstrate that the questionnaire is an effective instrument for measuring vision difficulties in persons who receive low-vision rehabilitation.	N
Turano KA., et al.[184]	2002	Self-Assessment Instrument for Measuring Independent mobility/ To measure perceived visual ability for independent mobility in patients with glaucoma.	The instrument developed for patients with retinitis pigmentosa (RP), to determine difficulty across a range of mobility situations, is a valid measure of perceived ability for independent mobility in patients with glaucoma.	P

Table 5.7 Measures Developed for Vision Impairments, or used in a Population with Vision Impairments. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
Turano KA., et al.[184]	2002	Self-Assessment Instrument for Measuring Independent mobility/ To measure perceived visual ability for independent mobility in patients with glaucoma.	The instrument developed for patients with retinitis pigmentosa (RP), to determine difficulty across a range of mobility situations, is a valid measure of perceived ability for independent mobility in patients with glaucoma.	P
Velozo CA., et al.[222]	2000	Visual Function-14 (VF-14)/ To assess visual function through self-report.	Relative to the original VF-14, the resulting VF-10 showed less redundancy of items while person separation (2.20) and Cronbach's alpha (.89) remained relatively intact. The study demonstrates that Rasch analysis, while effective in elucidating the metrics of an original instrument, can also be useful in designing modifications of instruments that are both efficient and psychometrically sound.	P

N = Development of a new measure; P = Determine the psychometric properties of an existing test or index or improve an existing test or index.
*Conclusions are a direct citation from the authors in each article.

Table 5.8 Measures Developed for Back Pain or Back Impairments, or Used in a Population with Back Pain or Back Impairments.

Authors	Year	Measure(s)/Use	Study Conclusions*	Aim
Cook CE., et al.[223]	2006	National Health and Nutrition examination Surveys Activities of Daily Living Instrument (NHANES ADL)/ To assess activities of daily living in persons with low back pain.	Individual use of selected items of the NHANES ADL instrument may further improve the capacity of the health car provider in measuring and recording dysfunction associated with Low Back Pain (LBP).	P
Davidson M., et al.[224]	2004	Low-Back SF-36 PF18/ To assess function in persons with low back pain.	The Low-Back SF-36 PF18 comprises the 10-item SF-36 PF scale and four items each from the Oswestry and Quebec back pain questionnaires. The possible total score ranges from 0 to 100, with a higher score indicating better function. The new scale appears to offer advantages over the used of the original scale for the assessment of functioning in patients with low back pain.	P
Garatt AM.[291]	2003	Roland Disability Questionnaire/ To measure health outcome for back pain.	The Roland Disability Questionnaire largely satisfies the Rasch model for unidimensionality. However, the instrument could be improved through the removal of poorly fitting items and the addition of items at the upper and lower points of the scale hierarchy. The distribution of Roland Disability Questionnaire scores should be carefully considered before statistical testing is undertaken. Rasch transformed scores can be used to deal with deficiencies in the scale hierarchy.	P

Table 5.8 Measures Developed for Back Pain or Back Impairments, or Used in a Population with Back Pain or Back Impairments. (cont.).

Authors	Year	Measure(s)/Use	Study Conclusions*	Aim
Granger CV., et al.[292]	2003	BMC, Placid, and Painfree measures. -One-item LVAS (LIFEware System visual Analog Scale)/ To determine the rate per day of improvement expected for outpatients with low back.	Findings may be useful for clinical application because the actual rate of improvement may be compared with the expected rate. There was only a weak relationship shown between affective factors and pain improvement.	P
Hart DL., et al.[180]	2006	Back Pain Functional Scale-Physical Functioning Items (PF)/ To measure function in patients with lumbar spine impairments.	A body part specific simulated CAT developed from a lumbar spine functional status(LFS) item bank was efficient and produced precise measures of LFS without eroding discriminant validity. Further studies of the practical use of this LFS CAT to track functional change in patients receiving rehabilitation for lumbar impairments are needed.	N
Kopec JA., et al.[181]	1996	The Quebec Back Pain Disability Scale/ To assess functional disability in patients with back pain.	Items that were highly effective in discriminating between different levels of disability were selected for the final, reduced scale. The scale has 20 items, representing six empirically derived categories of activities affected by back pain. Measurement properties of this instrument have been previously discussed.	N

Table 5.8 Measures Developed for Back Pain or Back Impairments, or Used in a Population with Back Pain or Back Impairments. (cont.).

Authors	Year	Measure(s)/Use	Study Conclusions*	Aim
Kopec JA., et al.[181]	1996	The Quebec Back Pain Disability Scale/ To assess functional disability in patients with back pain.	Items that were highly effective in discriminating between different levels of disability were selected for the final, reduced scale. The scale has 20 items, representing six empirically derived categories of activities affected by back pain. Measurement properties of this instrument have been previously discussed.	N
Kopec JA., et al.[181]	1996	The Quebec Back Pain Disability Scale/ To assess functional disability in patients with back pain.	Items that were highly effective in discriminating between different levels of disability were selected for the final, reduced scale. The scale has 20 items, representing six empirically derived categories of activities affected by back pain. Measurement properties of this instrument have been previously discussed.	N
Kucukdeveci AA., et al.[225]	2001	Roland-Morris Disability Questionnaire (RMDQ)/ To assess disability in low back pain.	The RMDQ is a robust unidimensional ordinal measure, largely free of differential item functioning, which works well in the Turkish population. Nonparametric effect sizes of ordinal scales are found to overestimate or underestimate the true effect size depending on the nature of the scale and the distribution of patients at baseline.	P
McArthur DL., et al.[293]	1991	Extensive battery of behavioral and psychological tests/To demonstrate use of Rasch analysis used in the assessment of rehabilitation.	Although ridit analysis provides potentially interesting group-aggregated information, Rasch analysis provides a full exposition of the overall quality of measurement, the quality of measurement for each variable, and the relation of respondents to measures. With this model, the underlying continuum upon which ordered categorical assessments are based can be illuminated.	P

Table 5.8 Measures Developed for Back Pain or Back Impairments, or Used in a Population with Back Pain or Back Impairments. (cont.).

Authors	Year	Measure(s)/Use	Study Conclusions*	Aim
McArthur DL., et al.[293]	1991	Extensive battery of behavioral and psychological tests/To demonstrate use of Rasch analysis used in the assessment of rehabilitation.	Although ridit analysis provides potentially interesting group-aggregated information, Rasch analysis provides a full exposition of the overall quality of measurement, the quality of measurement for each variable, and the relation of respondents to measures. With this model, the underlying continuum upon which ordered categorical assessments are based can be illuminated.	P
Page SJ., et al.[226]	2002	Oswestry Low Back Pain Questionnaire/ To measure low back pain disability.	By using the abbreviated scale, suggested item order, and predicted responses, abbreviated versions of the instrument can be applied to measure LBP disability more efficiently.	P
Velozo CA., et al.[294]	1999	Worker Role Interview (WRI)/ To assess psychosocial capacity for return to work in injured workers.	This initial series of studies present a theoretically based instrument, which shows promising psychometric qualities. While the predictive study indicated that the WRI was not useful in predicting return to work, this finding may have been a function of the small n-size in the study or that the WRI may mediate its effect through other variables.	P
Velozo CA., et al.[295]	2006	Occupational Rehabilitation Data Base (ORDB) function capacity instrument/ To evaluate functional capacity.	These findings suggest that Handling items should not be included as a dictionary of occupational titles (DOT) measure for clients with back pain. While the above psychometric study supports using client or therapist ratings as independent instruments, the lack of concordance between these ratings requires further investigation.	P

Table 5.8 Measures Developed for Back Pain or Back Impairments, or Used in a Population with Back Pain or Back Impairments. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
White LJ.,et al.[227]	2002	Oswestry Low Back Pain Questionnaire/ To assess self-reported disability levels quantitatively.	Rasch analysis produced disability categories in the Oswestry that are linear and, therefore, useful for quantitatively assessing self-reported disability levels.	P

N = Development of a new measure; P = Determine the psychometric properties of an existing test or index or improve an existing test or index.

*Conclusions are a direct citation from the authors in each article.

Table 5.9 Measures Developed for Arthritis and Other Inflammatory Diseases, or Used in a Population with Arthritis and Other Inflammatory diseases.

Authors	Year	Measure/Use	Conclusions*	Aim
Chiou C., et al.[228]	2006	Two multiattribute preference weight functions (MAPWF)/ To assess health states and associated preference weights of patients with rheumatoid arthritis.	Our results reveal that the new measures are reliable and valid in assessing health states and associated preference weights of patients with rheumatoid arthritis.	N
Chiou C.[296]	2004	Cedars-Sinai Health-Related Quality of Life in Rheumatoid Arthritis (CSHQ-RA)/ To evaluate quality of life in rheumatoid arthritis.	The abridged CSHQ-RA short form is a valid and reliable instrument that can be used to examine the impact of RA on patients' health-related quality of life. Prospective validation in clinical trial settings is warranted.	N
Dawson J., et al.[297]	2005	Lequesne Index of severity for osteoarthritis of the hip (LISOH)/ To assess severity of osteoarthritis of the hip in an elderly population.	The current study identifies major limitations with the LISOH – particularly if used as a single composite measure.	P

Table 5.9 Measures Developed for Arthritis and Other Inflammatory Diseases, or Used in a Population with Arthritis and Other Inflammatory diseases. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
Gilworth G., et al.[298]	2003	The Work Instability Scale for Rheumatoid Arthritis (RA-WIS)/ To evaluate work disability in rheumatoid arthritis.	The WIS can be scored in 3 bands indicating low, medium, and high risk of work disability.	N
Helliwell P., et al.[231]	2005	Foot Impact Scale for rheumatoid arthritis/ To assess the impact of rheumatoid arthritis and to measure the effect of interventions.	A foot impact scale to assess the impact of RA and to measure the effect of interventions has been developed. The 2 scales comprising the instrument demonstrate good psychometric properties.	N
Kopec JA., et al.[236]	2006	Item Response Theory-based Item Bank/ To measure quality of life in 5 domains relevant to arthritis.	We have developed IRT-based item banks to measure HRQL in 5 domains relevant to arthritis. The items in the final item banks provide adequate psychometric information for a wide range of functional levels in each domain.	N

Table 5.9 Measures Developed for Arthritis and Other Inflammatory Diseases, or Used in a Population with Arthritis and Other Inflammatory diseases. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
Kucukdeveci AA., et al.[232]	2004	Turkish version of the Stanford Health Assessment Questionnaire/ To use the Stanford health Assessment Questionnaire in international studies between UK and Turkey.	Cross-cultural validity was found to be sufficient for use in international studies between the UK and Turkey. Future adaptation of instruments should include analysis of DIF at the field testing stage in the adaptation process.	P
O'Malley KJ., et al.[233]	2003	Joint-Specific Multidimensional Assessment of Pain (J-MAP)/ To measure joint-specific pain of patients with knee osteoarthritis.	The J-MAP is a reliable, valid, and responsive measure for assessing joint-specific pain at a single time point, or changes over time for one or a group of patients with knee osteoarthritis. With this initial evidence of its psychometric rigor, further testing of the measurement properties of the J-MAP in other joints and in other populations should be undertaken.	N
Penta M., et al.[229]	1998	ABILHAND questionnaire/ To measure manual ability.	Even in a small sample of patients, using the Rasch methodology enabled the investigators to produce a useful scale of manual (dis)ability and to define manual ability as a unique construct, at least in patients with rheumatoid arthritis.	N
Ryser L, et al.[299]	1999	Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)/ To assess patients with osteoarthritis.	Applying Rasch analysis to the 3 WOMAC domains confirms their unidimensionality. However, the pain and function items seem to represent the same construct. Thus, it may be possible to simplify these domains to avoid redundancy. The impact of reducing items must nevertheless be examined.	P

Table 5.9 Measures Developed for Arthritis and Other Inflammatory Diseases, or Used in a Population with Arthritis and Other Inflammatory diseases. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
Sheehan TJ., et al.[300]	2000	Activities of Daily Living (ADL) and Instrumental Activities of Daily Living, items from National Health and Nutrition Examination Survey and Follow-up Study (NHEFS)/ To assess ADLs and IADLs in patients with rheumatoid arthritis.	The fit between the items and the samples, indicating the compatibility between the test and subjects, is seen much more clearly with Rasch with more than half of the general population measuring the extremes. Since research on disability depends on measures with known properties, the superiority of Rasch over Likert is evident.	P
Wolfe F., et al.[230]	2004	Short Arthritis Assessment Scale (SAS)/ To evaluate arthritis severity in research and clinical practice.	The SAS is a 4-item arthritis severity questionnaire that can be easily administered in primary care for patients with OA, but is suitable for use across all arthritis illnesses. Scoring is simple, requiring only the addition of four 10-point scales, and interpretation is straightforward. The SAS may have a role in rapid assessment of the arthritis patient in primary care practice.	N
Wolfe F., et al.[234]	2004	Health Assessment Questionnaire II (HAQ-II)/ To measure functional status in rheumatology.	The HAQ-II is a reliable and valid 10-item questionnaire that performs at least as well as the HAQ and is simpler to administer and score. Conversion from HAQ to HAQ-II and from HAQ-II to HAQ for research purposes is simple and reliable. The HAQ-II can be used in all places where the HAQ is now used, and it may prove to be easier to use in the clinic.	N

Table 5.9 Measures Developed for Arthritis and Other Inflammatory Diseases, or Used in a Population with Arthritis and Other Inflammatory diseases. (cont.).

Authors	Year	Populations	Study Conclusions*	Aim
Wolfe F., et al.[235]	2003	Regional Pain Scale (RPS)/ To identify patient with fibromyalgia (FM) or develop a new definition of FM.	The RPS is a valid scale of pain extent. It can be useful to identify patients with FM or can be used to develop a new definition of FM, even among patients with concomitant illnesses such as RA and OA. In addition, it is a measure of pain extent that is disease independent, and works as well in RA and OA as in FM to identify patients with increased severity and resource utilization..	N

N = Development of a new measure; P = Determine the psychometric properties of an existing test or index or improve an existing test or index.
*Conclusions are a direct citation from the authors in each article.

Table 5.10 Studies Developing New Measure(s) Using Rasch or Item Response Theory.

Measure	Population
ABILHAND questionnaire[229]	Rheumatoid arthritis
ABILHAND-Kids: A Measure of manual ability (questionnaire)[203]	Cerebral Palsy (children)
Activity Measure for Postacute care (AM-PAC)[301]	Post Acute Care-Neurologic-Musculoskeletal-Medically Complex
Activity Measure for Postacute care (AM-PAC)[302]	Volunteers receiving skilled rehabilitation services-Neurologic-Musculoskeletal-Medically Complex
AMA vision disability scale[177]	Low vision
Assessment of Awareness of Disability (AAD)[254]	Stroke
Assessment of Motor and Process Skills (AMPS) – Online Performance[303]	N/A
Assessment of Motor and Process Skills(AMPS)[304]	N/A
Assisting Hand Assessment (AHA)[305]	Children with unilateral upper limb impairments
Audiovisual computer-based testing (AVCVT) platform to assess health status and preference-based HRQOL[306]	Vulnerable Populations (chronic illness, low literacy)
Back Pain Functional Scale-Physical Functioning Items (PF)[180]	Lumbar Spine impairments
Barthel Index (BI)-Frenchay Activities Index (FAI)[196]	Stroke
BD-QoL: a quality of life measure specific to Behçet's disease[307]	Behçet's disease
Caregiver Self-Efficacy Scale for Transfers (CSEST)	Caregivers of children requiring Transfer Assistance
Cedars-Sinai Health-Related Quality of Life in Rheumatoid Arthritis (CSHQ-RA)[296]	Rheumatoid Arthritis
Child and Family Follow-up Survey (CASP)[272]	Acquired brain injury (children)
Community Ambulatory Physical Performance Scale (CAPPS)[308]	Community-dwelling with difficulties in mobility
Community-dwelling persons with disability[309]	Functional Independence Measure (FIM[TM])

Table 5.10 Studies Developing New Measure(s) Using Rasch or Item Response Theory. (cont.).

Measure	Population
Compuperized Adaptive Version of Shoulder Functional Status (SFS) items[310]	Shoulder impairments
Computerized adaptive tests (CATs)from the Lower Extremity Functional Scale (LEFS) designed to assess lower extremity functional status[311]	Patients with hip, knee, or foot/ankle impairments
DASH (Disabilities of the Arm, Shoulder and Hand) Outcome Measure.[312]	Patients with upper-limb musculoskeletal disorders
Dizziness Handicap Inventory Short Form (DHI-sf)[313]	Patient attending neuron-otologic outpatient center due to dizziness and/or imbalance complaints
Dynamic gait index (DGI)[314]	Balance disorders
EG Motor Index[189]	Stroke
EQUISCALE[209]	Multiple Sclerosis
Falls Efficacy Scale[315]	Community living older adults (over 65)
Fatigue Item Bank from the FACIT-Fatigue Scale[316]	-Anemic cancer patients-General population
Flexilevel Scale of Shoulder Function (FLEX-SF)[317]	Shoulder-related dysfunction
Foot Impact Scale for Rheumatoid Arthritis[231]	Rheumatoid Arthritis
Frequency of Forgetting scale of the Memory Functioning Questionnaire (MFQ).[318]	Members from Family HealthPlan (FHP), a Southern California-based health maintenance organization (HMO)
Functional Independence Measure-Patient Evaluation and Conference System[1]	Brain injury-neuromuscular-Musculoskeletal-Spinal cord-Stroke
Functional Recovery Measure	Stroke
Health Assessment Questionnaire II (HAQ-II)[234]	Rheumatic diseases
Instrument of home and community participation (PAR-PRO)[269]	Inpatient rehabilitation: -Orthopedic-Stroke-Neurologic-Brain injury-Cardiac and pain-Arthritis and others
International Knee documentation Committee Subjective Knee Form (IKDC) 5832	Subjects with a variety of knee problems.

Table 5.10 Studies Developing New Measure(s) Using Rasch or Item Response Theory. (cont.).

Measure	Population
IRT-based Item Bank[236]	Arthritis
Items from Patient Reported Outcomes Measurement Information System (PROMIS)[319]	Musculoskeletal Disease
Joint-Specific Multidimensional Assessment of Pain (J-MAP)[233]	Knee Osteoarthritis
Late Life Function and Disability Instrument[320]	Adults over 60 with a range of functional limitations
Late Life Function and Disability Instrument[321]	Community-dwelling adults over 60
Manual Ability Measure (MAM-16)[322]	Variety of hand impairments
MAS Word Learning Test Trials-WMS-R Digit Span-Animal category fluency-Letter fluency[323]	Dementia
Measure of IADL[324]	70-year old persons who do not depend on help
Melbourne Low-Vision ADL Index[213]	Visions impairments from various causes
Method of measuring vision disabilities[289]	Vision disabilities
MINDFIM[274]	Traumatic Brain Inury
Motor evaluation Scale for Upper Extremity in Stroke Patients (MESUPES)[256]	Stroke
MPS Physical Performance Measure (MPS-PPM)[325]	Mucopolysaccharidosis type I (MPS I)
Multiple Sclerosis Spasticity Scale (MSSS-88)[208]	Multiple Sclerosis
New 25-item composite measure (Mini-Mental State Examination-Blessed Information Memory concentration Test-Blessed-Roth dementia rating Scale)[326]	
Nottingham Health Profile (NHP) (Spanish version)[327]	General population-Different clinical pathologies
Older adults Desire for Physical Competence (DPC)[328]	Older adults (aged 60-95)
Participation Measure for Post-Acute Care (PM-PAC)[260]	Neurologic-Musculoskeletal-Medically complex
Pediatric Evaluation of Disability Inventory (PEDI)[278]	Children with Acquired Brain Injury
Physical Ability Stroke Scale (PASS)[251]	Stroke

Table 5.10 Studies Developing New Measure(s) Using Rasch or Item Response Theory. (cont.).

Measure	Population
Physical Functioning Items based on:-Activity domains of the ICF - Input from measurement and content experts-focus groups of persons with disabilities-Literature review[329]	Neurologic-Musculoskeletal-Medically complex
Play Assessment for Group Setting (PAGS)[330]	Children aged from 2 to 8 years
Prototype clinical performance measure[273]	Paediatric brain injury inpatient rehabilitation
Quebec Back Pain Disability Scale[181]	Back Pain
Rasch-derived Patient Activation Measure (PAM)[331]	Adults aged 45 and older
Regional Pain Scale (RPS)[235]	Rheumatic Disease
Self-administered survey of functional status[332]	Individuals over 65 from outpatient ambulatory clinic (arthritis, hypertension, hearing problems, back problems, diabetes)
SF-36-Condition-specific HRQOL questionnaires-Oswestry Low Back Pain Disability Questionnaire-Neck Disability Index-Lysholm Knee Questionnaire[261]	Spinal Impairments-Knee Impairments-Upper and Lower extremity impairments-Cervical and lumber spine impairments-Stroke-Orthopedic conditions-Pain syndromes-Neurologic conditions
SF-36V Physical Functioning Scale[333]	Veterans with Spinal Cord Injury
Short Arthritis Assessment Scale (SAS)[230]	Osteoarthritis-Rheumatoid arthritis-Fibromyalgia
Special Interest Group in Amputee Medicine (SIGAM) mobility grades[334]	Lower Limb Amputees
Stroke Impact Scale (SIS)[188]	Stroke
STroke REhabilitation Assessment of Movement (STREAM)[242]	Stroke
Systemic Lupus Erythematosus-specific quality of life instrument (SLEQOL)[335]	Systemic Lupus Erythematosus
Tufts Assessment of Motor Performance (TAMP)[336]	Children and adults with impairments in motor performance due to neurological or musculoskeletal diagnosis
Two multiattribute preference weight functions (MAPWF)[228]	Rheumatoid Arthritis

Table 5.10 Studies Developing New Measure(s) Using Rasch or Item Response Theory. (cont.).

Measure	Population
University of Illinois at Chicago Fear of Falling Measure[337]	Community Dwelling Elderly
Veterans Affairs Low Vision Visual Functioning Questionnaire[221]	Persons with low vision
VETERANS Affairs Low-Vision Visual Functioning Questionnaire (VA LV VFQ-48)[219]	Low-vision patients
VisQoL (6-item)[218]	Visually impaired
Women's Quality of Life Questionnaire (WOMQOL)[338]	Quality of Life in Women of reproductive age with no known pathology
Work Instability Scale for Rheumatoid Arthritis (RA WIS)[298]	Rheumatoid Arthritis

Table 5.11 Studies Using Previously Rasch-Developed Measures.

Measure	Population
Children's version of the ABILHAND (ABILHAND-Kids)[339]	Children with deficiency of the upper limb
Assessment of Motor and Process Skills (AMPS)[340-343]	Stroke
	Adolescents with cerebral palsy
	Adults with mental retardation
	Nonwell persons
Avlund mobility scales, mobility-tiredness and mobility-help.[344]	75-year old men and women with follow-up at age 80
Child and Adolescent Scale of Participation (CASP) [345]	Children and youth with traumatic and other acquired brain injuries.
The Craig Handicap Assessment and Reporting Technique (CHART) [346]	Spinal cord injury
Functional Independence Measure (FIM)[347-355]	Burn Injury
	Spinal cord injury
	Traumatic Brain Injury
	Family Members
	Spinal Cord Injury
	Stroke-Other impairment groups
	Guillain-Barré syndrome
	Stroke
	Spinal Cord Injury
Mayo-Portland Adaptability Inventory(MPAI)[356]	Acquired brain injury
Medical Rehabilitation Follow Along (MRFA)[357]	Musculoskeletal impairments.
Motor Readiness Questionnaire for Stroke (MRQS)[358]	Stroke
Participation Measure for Post-Acute Care (AM-PAC)[359]	Major chronic disabling conditions:-neurologic disorder-lower-extremity orthopedic trauma-medically complex condittions

Table 5.11 Studies Using Previously Rasch-Developed Measures. (cont.).

Measure	Population
Pediatric Evaluation of Disability Inventory (PEDI) [360]	N/A
Supervision Rating Scale (SRS) [361]	Traumatic Brain Injury
Test of Infant Motor Performance(TIMP) [362]	Infants tested at 7, 30, 60, and 90 days after term age.
Test of Playfulness (ToP) [363,364]	Children with motor delay Adolescent males both typically developing (TD) and with severe emotional disturbance (SED)

CHAPTER 6 Manuscript 3

Preface to Manuscript 3

Development and Initial Psychometric evaluation of an Item Bank created to Measure Upper Extremity Function in persons with Stroke

The ultimate goal of all rehabilitation interventions is to enable persons to resume the performance of their daily activities and fulfilment of their life roles after being affected by an injury or a disease. Persons with stroke often have an impaired arm following the event, preventing them from independently accomplishing even simple tasks such as bathing and dressing. The severity of the impairment is unique to each person, ranging from a complete paralysis of the arm to a loss of coordination, proprioception or even sensation in the affected arm. Numerous rehabilitation strategies have been used throughout the years, based on different models of recovery and different theories. To date, none of them has constantly been proven to be superior to the others. The comparison of the different studies evaluating the impact of the interventions is made difficult by the absence of consensus on the appropriate outcome assessments used to evaluate the extent to which the interventions were successful. The implementation of effective treatment strategies lies on our ability to measure change adequately. At the present time there is a lack of rigor in measurement strategies used in rehabilitation. Most evaluation tests and indices were developed under the framework of Classical Test Theory where numerals are assigned to describe how well a specific task is accomplished through the observation of the subject or though the administration of questionnaires. Those numerals, as opposed to actual numbers, are ordinal in nature and are subsequently used in statistical procedures where they are wrongly treated as if they were interval-level data. Wrong inferences can be drawn about the effectiveness of a particular treatment when such practices are employed. Also, it is not unusual, in clinical practice and in research, to assess persons with stroke with three or four different tests of capacity for skilled activity and several questionnaires designed to capture actual use of the arm in daily activity (performance). The measurement protocol, under Classical Test Theory, currently requires all items of all tests be administered even though there are evident redundancies. Summarizing recovery is a difficult process and communication across disciplines and rehabilitation settings is arduous. There is no common language for describing recovery and no common currency. There is a need to create objective, scientific measures in the field of rehabilitation. New psychometric methods, Rasch and Item Response Theory models have progressively penetrated health sciences over the past few years as tools to construct such measures. Although Rasch analysis and Item Response Theory are often considered as similar, with the Rasch model being the most 'parsimonious' of the Item Response Theory models, the philosophy behind their development is quite different. Because the goal of the following manuscript is to create a method to objectively measure upper extremity function, the author believes the Rasch analysis is an appropriate method to use. Indeed, Rasch analysis meets the requirements for true measurement: unidimensionality, linearity, sample independence, scale independence, as well as invariance.

Rehabilitation professionals need to move toward a common language of functional assessment.[1] Rasch Analysis has been used extensively in the fields of education and psychology[1] and now, as shown in the previous manuscript, has been studied extensively in rehabilitation but not used. There is a pressing need to harmonize the assessment of the upper extremity.

The main objective of this third manuscript is to create a bank of items measuring upper extremity function. This new bank will be unique as it will comprise both performance items as well as questions on how the patients feel they are capable or how difficult or limited they are in accomplishing an activity or a task. The responses to tests and questionnaires from over 4000 persons with stroke from eight studies were analyzed with the Rasch model. The data were collected longitudinally, a core set of tests and indices crosses studies and some studies have unique tests and indices. Rasch modeling can handle this data structure very well. The tests and questions retained for the analysis were those pertaining to upper extremity function. The World Health Organization, through its International Classification of Functioning, Disability, and Health (ICF) has provided a framework and classification system to guide the development of rehabilitation outcomes instruments and has classified health and health related domains. These domains include body, individual, societal and environmental perspectives. The domains are: (1) *body structure and function* and (2) *activity and participation*. Functioning can be classified by the term *capacity*, which is what a person is able to do in a standard environment or test situation, and *performance*, which is what a person actually does in their familiar environment in the community or at home. One of the aims of the ICF is to provide a common language to improve communication across users: health care professionals, researchers, policy makers, as well as clients.[365] A group of researchers gathered evidence from the literature in order to identify the most common problems among stroke patients using the ICF checklist and formed a 'core set' of items for stroke. Under the Activities and Participation component, four items (toileting, eating, washing oneself and dressing) requiring the use of the upper extremity were included in the preliminary core set.[266] Thus, using indices of impairment to assess motor deficits, indices of performance and indicators of use of the affected upper extremity and finally using indices of health related quality of life from a total of eight studies were used produce a bank of items to measure upper extremity function. These items are thought to cover the entire spectrum of upper extremity function, and define, by the same token, the concept of upper extremity function. Because Rasch analysis can be used to transform ordinal level Likert scale responses into 'interval-like' level item difficulty estimates, it makes it an ideal to model to 'measure' self-reported responses of how a patient feels they are limited or how difficult an activity is to accomplish.

Measuring change using a 'true' measure is a first step towards finding effective treatment interventions for the affected arm of persons with stroke.

To meet the objectives of this third manuscript, the candidate was responsible for assembling the data from each of the eight studies and preparing the data for Rasch analysis as well as carrying out all statistical analyses.

Title Page

Development and Initial Psychometric Evaluation of an Item Bank created to Measure Upper Extremity Function in Persons with Stroke.

Authors: Johanne Higgins, MSc, O.T., Lois E. Finch, MSc, P.T., Jacek Kopec, MD, Ph.D, Nancy E. Mayo, BSc (PT), Ph.D.

Posts of the Authors at the Time of the Study:

Higgins J: PhD candidate, School of Physical and Occupational Therapy, Faculty of Medicine, McGill University.

Finch LE: PhD candidate, School of Physical and Occupational Therapy, Faculty of Medicine, McGill University.

Kopek J: Assistant Professor Health Care and Epidemiology Faculty of Medicine, Mather Building University of British Columbia.

Mayo NE: Associate Professor, School of Physical and Occupational Therapy, Faculty of Medicine, McGill University, Associate member, Department of Epidemiology and Biostatistics and Occupational Health, Faculty of Medicine, McGill University

Work conducted in: Montreal, Quebec, Canada, and Quebec City, Quebec, Canada

Address correspondence to: Nancy Mayo, Royal Victoria Hospital, Division of Clinical Epidemiology, 687 Pine Avenue West, Ross 4.29, Montreal, Quebec, H3A 1A1 CANADA, Telephone: (514) 934-1934 ext. 36922, Fax: (514) 843-1493, e-mail: nancy.mayo@mcgill.ca

Abstract

Title: Development and Initial Psychometric Evaluation of an Item Bank created to Measure Upper Extremity Function in Persons with Stroke.

Objective: To create and illustrate the development of a method to parsimoniously and hierarchically assess upper extremity function in persons with stroke.

Design: Data were analyzed using Rasch analysis.

Setting: Re-analysis of data from 8 research studies involving persons with stroke.

Patients: Over 4000 patients with stroke who participated in various randomized controlled trials and observational studies in the Montreal area and elsewhere in Canada.

Interventions: Not applicable.

Main Outcome Measure(s): Data were comprised of 17 tests or indices of upper extremity function and health-related quality of life, for a total of 99 items related to upper extremity function. Tests and indices included, among others, the Box and Block Test, the Nine-Hole Peg test, grip strength, the Barthel Index and the SF-36. Data were collected at various times post stroke from three days to one year.

Results: Once the data fit the model, a bank of items measuring upper extremity function with persons and items organized hierarchically by difficulty and ability, in log units was produced. An item bank, a collection of items measuring upper extremity function all calibrated on the same ruler resulted.

Conclusions: A bank containing 49 items capturing upper extremity function post-stroke was calibrated. This bank forms the basis for an eventual computer adaptive testing. The items should be further psychometrically tested.

Key words: stroke, cerebrovascular accident, Rasch analysis, arm, upper extremity.

Introduction

Choosing an appropriate outcome assessment to evaluate upper extremity function after stroke is important and serves many purposes, from monitoring change in individuals, to guiding clinical decisions and evaluating rehabilitation programs. The argument for opting for psychometrically sound tests and indices has been made and is now widely accepted. Most of the tests and indices that have been created for measuring upper extremity function have been developed and tested using Classical Test Theory (CTT) where items are scored on an ordinal scale and the total score is the sum obtained on each of the items, all of which have to be administered to obtain a total score. This can lead to serious misinterpretations of the results as ordinal scales are treated as if they were interval scales. Moreover, all items comprised in the test or index have to be administered in order to calculate a final score.

Over the past few years, Rasch analysis has been increasingly utilized in the field of health outcome measurement to create 'true' measures. In Rasch analysis, items measuring a same construct or broad concept are positioned hierarchically along that construct, creating an interval-like measure or a conceptual ruler that allows for the measurement of true difference between and within individuals. Once the items are calibrated on the continuum of the construct being assessed, a standard metric allowing for the measurement of that construct has been created where items represent marks of increasing difficulty in accomplishing a certain task or endorsing a statement. Rasch analysis is different from classical test theory in the way data are related to a measurement model. The data are imposed to the model and must conform to it to comply with the requirements of measurements.[179] In other words, the model is not used to explain the data. In the Rasch model, the only parameters of interest are the *ability* of the persons and the *difficulty* of the items. The model defines the probability of a correct response when a person is challenged by an item. If the ability of a person is greater than the item's level of difficulty, the probability of a correct response should be more than 50%. Conversely, if the ability of a person is less that the item's difficulty, the probability of a correct response should be less than 50%. If the two parameters are equal, the probability of a correct response is exactly 50%. The equation of the basic Rasch model, for dichotomously scored items is:

$$P = \frac{\exp(\beta - \delta)}{1 + \exp(\beta - \delta)}$$

When items 'fit' the Rasch model, adhering to its expectations, they meet the requirements for true measurement: linearity, additivity, equal distances and sample independence.[137] The items can then be used as a true measure of the construct under study.

Objective

The objective of this study is to develop, using Rasch analysis, a method of parsimoniously and hierarchically measuring upper extremity function in persons with stroke by the calibration of a pool of items into a bank.

Methods

Source of items and subjects

A secondary analysis of retrospective data at item level from over 4000 patients with stroke who participated in studies that included different tests and indices were used to create a new bank of items aimed at assessing upper extremity function after stroke. The upper extremity performance tests and indices from which the items were selected were administered to stroke patients in a wide variety of different settings over the past few years. The eight studies include longitudinal observational studies [118,366-368] as well as randomized controlled trials.[109,369-371] There is also data from a Canada-wide survey, the Canadian Stroke Registry (2 cohorts). Persons with hemorrhagic and non-hemorrhagic strokes were included. The hypothesis underlying this study is that it will be possible to assess the single construct of upper extremity function by combining items from diverse tests that were specifically designed to evaluate upper extremity function and to co-calibrate individual items relating to the upper extremity from indices of use of the upper extremity and of health-related quality of life. Every research project had ethical approval and there is no additional approval required for secondary analyses of the data.

Linking design

Because Rasch analyses are performed on existing data, linking analysis is the preferred procedure for developing a bank of items. In the RUMM program,[147] the pairwise algorithm is used for the calculation of the sufficient statistics for item parameters. This allows for missing data and the estimation of item parameters even when the data set, because of the specific study design, has systematically missing blocks of data. Because there is overlap between the subsets of data, parameters are estimated simultaneously without subsequent re-calibrations. Data from the different studies are thus automatically scaled to the same linear continuum and are directly comparable.[172] The linking design of the eight studies is presented in Table 6.1.

Items

Items representing the different domains of the International Classification of functioning, disability and Health (ICF) were chosen. Tests designed to assess upper extremity function, representing the *body functions and structures* domain are included in the pool. As well, items chosen from indices of *activity* and *participation* were chosen if upper extremity involvement was deemed to be important for that particular activity. For example, items such as driving, recreation and work were retained to be part of the item pool in order to remain as conservative as possible and not to delete items that would potentially provide useful information about upper extremity function. The ICF model was chosen because it goes beyond the pathology, providing a basis for understanding health-related states and a global language for disability.[372] It also provides a framework to develop outcome assessments that target important domains for persons living with disabilities related to upper extremity function.

Because some of the items assessing upper extremity function were measured in several studies, the distribution of items that were unique to each of the studies can be linked through the common items. Some of the indices used for the development of this new bank are not scored on an ordinal scale. The Box and Block Test, for example, is scored on a quasi-continuous scale consisting of the number of blocks a person can manipulate within a minute.[52] The Nine-Hole Peg Test is a timed test that also required transformation into an ordinal scale. Grip strength, measured in kilograms of force, was also categorized. Up to fifty different techniques to determine cut scores to divide data into categories have been used in the past, all of which rely on human judgment.[373] To reduce the potential threat of low precision, the number of categories was kept relatively high (5 categories for the Box and Block Test and 4 for grip strength and the Nine-Hole Peg Test) while making sure that the number of persons in each category was sufficient (at least 10).[365] The Rasch analysis that will be performed will indicate whether the response categories are adequate. If they are not, a collapsing of some of the categories will be warranted. The upper extremity tests and indices used are presented in Table 6.2 with a description of their original scoring or categorization, and the ICF component to which they belong. The scoring structure for all tests and indices and explanatory variables were transformed to be identical across all eight studies.

Personal Factors

Personal and health information as well as clinical features of the stroke were chosen to detect their impact on the way persons performed on the items. These factors were: age, gender, the type of stroke, the number of comorbid conditions, the side of the hemiplegia, whether the dominant hand was affected, and the time of assessment since the stroke. For the purpose of this study, the time of assessment for most studies was at the onset of stroke. For the Brain Capacity study, it was scheduled at three months and for the Walking Competency study and the Canadian Stroke Registry, at six months post-stroke. Table 6.3 presents the personal factors and their categorization.

Analysis

One of the requirements of Rasch analysis is unidimensionality: The items must measure one unique construct, upper extremity function.

Defining unidimensionality and local independence – Prior Tests

This step was undertaken in order to determine if the items formed a unidimensional measure of upper extremity function and to identify a priory if some items were not measuring the construct of upper extremity function.

A classical approach, principal component analysis was performed to test whether the pool of items from the different studies measured the single construct of upper extremity function. Because the data set contains blocks of missing responses, analyses were performed on each individual study separately. Each data set was evaluated to determine whether the pre-requisites for conducting principal component analysis were met. Although a normal distribution of item scores is preferable, is not absolutely required when statistical inferences are used for descriptive purposes only, such as in this study.[374] To assess normality, skewness and kurtosis coefficients are examined and must not exceed ±2.0 for the majority of variables (≥60%).[375] An evaluation of linearity using scatterplots of item pairs is recommended given that factor analysis is based on the correlation matrix underlying item scores and it is assumed that relationships between pairs of items are linear.[374,375] Factorability of the correlation matrix should also be assessed through the verification of: (1) correlation coefficients exceeding 0.30,[374], (2) numerous significant correlations between item pairs[374] and (3) a Kaiser-Meyer-Olkin (KMO) test statistic of 0.50 or above.[375]

There are several ways of selecting the final number of factors present or retained. Selection of the final number of factors can be guided by the Kaiser-Guttman rule (eigenvalue, $\epsilon>1$), the scree test, the percentage of the common variance explained by the retained factors, a residual analysis (minimal number of residual and partial correlations (≥0.10) and the interpretability of the factors with a minimum of 2 items loading on a factor[374] and parallel analysis (PA). In PA, eigenvalues from a data set prior to rotation are compared with those from a matrix of random values containing the same number of variables and same sample size. This technique has been shown to be accurate in determining the threshold for significant components and variable loadings. Only components from the Principal Component analysis that are greater than PA eigenvalues from the corresponding random data are be retained and components with eigenvalues below this threshold value are not considered as a component.[376]

All data sets were merged into one as a first step to estimate item and person parameters. Rasch analysis was performed using the Rasch Unidimensional Measurement Model program (RUMM2020 version).[147] In RUMM, all item difficulty and person ability parameters can be estimated in a single analysis and can be ordered on the same linear continuum.

Model Selection

Each test or index has a different number of items that are scored on different scales. The meaning of each response option for each of the items is also different among the different tests or indices. The measurement model that is most appropriate to use in a situation such as this one is the Partial Credit Model. Figure 6.1 presents a flow chart used for the selection of the appropriate model.

The first step of the analysis was to divide the sample into two random subsets of persons in order to verify the stability of the item difficulty hierarchy across the two subsets. Rasch analysis is an iterative process, decisions being made on a combination of statistical and clinical considerations. In order to accomplish this, items and persons need to conform to the model. Items and persons that do not fit the model were removed. Because a finite set of items measuring upper extremity function is used to create a bank of calibrated items, the main goal was to keep the greatest number of items possible as opposed to measuring and calibrating persons. This best strategy to use in the development of the measure of upper extremity function consists of a series of consecutive steps:[377] 1) evaluating person fit, 2) eliminating misfitting persons, 3) recalibrating, 4) evaluating item fit, and 5) evaluating the overall fit of data to the model.

Person fit

Participants were divided into ten groups (class intervals) based on their total item scores. Fit residuals are provided for each person to estimate the difference between the Rasch model's theoretical expectations and the observed person's performance. If some persons do not fit the model, and it is judged appropriate, they are removed from the analysis. Fit was determined by standardized residuals; they should be between -2.5 and 2.5. There are a multitude of reasons that may explain why some persons do not fit the Rasch measurement model and each possibility requires careful attention and consideration. Sometimes persons will not fit the model because they did not answer or did not perform in the expected manner. It may also be due to the fact that they did not understand the question they were asked. In the stroke population, this may be due, for example, to receptive aphasia. After careful consideration, misfitting persons were removed from the analyses. Elimination was done starting with the most misfitting persons based on the mean square residuals.

Item fit

By default, the mean of the item locations along the measurement continuum is located at zero. This is because only relative positions along the continuum are estimated as opposed to absolute positions. The location of the items is, however, independent of the location of the persons, as this is a key feature of Rasch measurement model. A Standard Error (SE) is calculated for each item to evaluate the precision of the location parameter.

The analysis of fit also indicates which items fit the model and which do not. Two fit statistics describe the closeness of the observed and predicted scores. Some authors have suggested that the misfitting items need to be closely examined before they are removed. For example, if the overall fit of the model is satisfactory and the item is judged to be an important part of the test, it should probably be kept. Conversely, Wise and colleague are in favour of removing misfitting items as soon as they are identified.[378] Fit of the items to the model can also be assessed graphically using the Item Characteristic Curves or ICCs, which are the expected curves based on the model specifications. Observed model fit for groups of participants across the measured construct (class intervals) are plotted against the ICCs and if each of the group plot is lying on the curve, the items fit the model well. Choosing to eliminate an item will have an impact for the face value of the measure and must be based not only on the numbers but also upon clinical reflection.[143] In RUMM, a fit statistic is presented for each of the items. It is presented as a standardized difference between the observed number of persons in the group who have the item correct and the expected number based on the model. In addition, an item-trait interaction is provided. It is the sum of the overall chi-square for the items and it is adjusted for multiple testing using Bonferroni corrections.[143]

When items do not fit the model, careful consideration of each misfitting item is also warranted. An item may not fit because it is not 'measuring' the same attribute as the other items. In other words, it may not be measuring upper extremity function and thus cannot be 'fitted' on to the ruler.

Ordered Rating Scale

Another reason why the item may misfit is if the categories of its response options are not optimal. Rasch analysis allows the verification and the optimization of response option categorization. Within each item, participants with less ability should endorse the lower scoring category and people with more ability should endorse a higher scoring category. When the opposite happens, *disordered thresholds* are observed. As well, investigation of the utility of the rating scale categories, both statistically and graphically, was undertaken. Some categories were under- or unused compared to the other categories and items needed to be rescored. After each potential rescoring combination of rating scale categories, the overall fit to the model and the fit of that item were verified and several strategies were attempted before reaching the best possible fit. Once all the items had been optimally categorized, persons and items closely scrutinized for fit to the model.

Targeting

The average mean person location indicates whether the items contained in the bank are well-targeted for the population under study. If the measure was perfectly targeted, this number would be zero logits to match the average mean item location, their respective standard error should be similar. A number below zero is an indication that the persons are at a lower level of upper extremity function than the average of the items contained in the bank (the items are too hard). Conversely, a number above zero indicates that the persons are at a higher level of upper extremity function than the average of the items in the bank (the items are too easy).

Differential Item Functioning (DIF)

Items displaying DIF change their location on the measurement continuum depending on the group of persons being assessed.[137] As the goal of this study is to construct a true measure that is invariant across the measurement continuum, these items have to be dealt with. In the RUMM program, it is possible to detect the presence of DIF both graphically and statistically.[143] For each item individually, scores across each level of influencing factor (also called persons factors) and across different levels of the construct (class intervals), are considered. Each person is identified by a person factor and by a class interval. Standardized residuals for each person attempting each item are analyzed through a standard analysis of variance (ANOVA).[143] If a significant main effect of the person factor is detected, it indicates the presence of uniform DIF. If a significant interaction effect between persons and class interval is detected, it indicates the presence of non-uniform DIF. [143]

Verifying the Unidimensionality of the New Measure – Post-Hoc Tests

Fitting the data to a Rasch model incorporates a test of the degree to which persons and items fit the linear continuum of upper extremity function. In RUMM, the overall fit to the model is ascertained first by the summary statistics. The mean item fit residual across all items should be close to 0 and its standard deviation should be close to 1. RUMM also provides reliability indices: the Person Separation Index (reliability of the hierarchy) and Cronbach's Alpha. The first is useful to understand the fit. If it is close to 0, the power of the test of fit will be low. The quality of the fit will also be determined using the ICC's and the interaction test of fit and the chi-square statistic. Unidimensionality can also be ascertained graphically by visualizing the ICC's. The ICC's for each of the items as well as for categories within the items should never cross, they should be parallel. [143]

Principal Component Analysis of the residuals is another method of ascertaining unidimensionality. It rests on the assumption that once the 'Rasch factor' is removed, no other factors should remain.[379]

Smith's independent t-test has also been suggested as a way to verify unidimensionality.[166] This test determines if the person estimates derived from specific subsets of items are significantly different from those derived from the complete set of items.[379] If they are different, unidimensionality cannot be ascertained.

Validity

Construct validity is concerned with the extent to which a test designed to measure a specific theoretical trait or proficiency actually does so. It is comprised of two levels.[137] The first level is that of the items and assessing the degree to which they are unidimensional. Therefore, if the items representing the different domains of upper extremity function each fit the model, construct validity is demonstrated. The second level is concerned with the constructs. The question to be addressed is: are the construct as measured by a battery of tests related as we expect them to be to the constructs measured by the items present in the bank.[380] The relationship between the new measure and the standard measures needs to be tested using correlation coefficients to demonstrate evidence of construct validity.

Content validity concerns the degree to which the item bank generally, and the items selected for each individual adaptive test specifically, are representative of the domain of ability.[380] According to the literature, experienced workers usually choose the same items when devising a scale for functional evaluation post-stroke.[381,382]

Results

Table 6.4 presents the socio-demographic and stroke characteristics of subjects from the eight studies. Subjects were sixty-eight years of age on average (SD 15). The time interval between the stroke and the time of the initial evaluation varied from the different studies from the onset of the stroke to six months post-stroke; 37% of the participants had their right side affected and 42% had their dominant side affected.

Unidimensionality

The initial analysis on the 99 items revealed that skewness and kurtosis coefficients did not exceed ±2.0 for more than 60% of the items, indicating that the level of normality was acceptable in the distribution of the items scores. As well, The Kaiser-Meyer-Olkin (KMO) test statistic was greater than 0.50 for each of the analyses. The principal component analyses followed by parallel analyses for each one of the individual studies revealed the presence of one important factor representing upper extremity function. In each of the individual studies, however, some of the items were loading on a second factor. As well, some of the items were identified as being redundant, having correlations ≥0.90. This exercise served to identify *a priori* items that may not be measuring upper extremity function. None of the items were removed following this first 'triage'. Items that loaded on a second factor for each of the individual studies are presented in Table 6.5.

Hierarchical structure

The next step consisted of a Rasch analysis of the 99 items. The initial fit of the data when all 99 items are considered produced a significant item-trait interaction. This is an indication that the data do not fit the model. Misfitting items, misfitting persons or both can be the cause. The next steps were undertaken for each of the two sub-samples of persons separately.

Ordered Rating Scale

Several items displaying disordered thresholds were found. The scores for these items were recoded by collapsing the responses and thus reordering the thresholds. Several collapsing strategies were checked for each one of the items until the best fit to the model was found. Even collapsing of the items that did not demonstrate threshold disordering was performed if this improved the fit of the model. Following the collapsing of the items, the fit of the individual items was rechecked revealing several misffiting items. Figure 6.2. presents the ordering of the thresholds for the 49 items in the bank.

Individual Person Fit

Several participants had residuals outside the recommended range (±2.5). Persons can misfit for a great variety of reasons and due to the large sample size, it was not possible to verify each case individually. The most misfitting persons, if any, were removed from the analyses after each item was rescored or removed.

Individual Item Fit

After the recoding of the items, several items showed misfit to the model expectations. Items were removed from the model based on fit residual values above or below 2.5, significant chi-square statistics and F-statistics. Deleted items and the reason for their deletion are presented in Table 6.6.

Differential Item Functioning

The presence of DIF was explored for each of the personal factors. DIF was deemed to be present if analyses of variance were significant (Bonferroni-corrected p value of 0.000510). Grip strength demonstrated DIF by gender and was split into 2 categories, one for women and one for men. This also makes clinical sense since women have less upper body strength and cannot be placed at the same level as men on an ability scale. After being split, grip strength for males demonstrated both uniform and non-uniform DIF by time of assessment since the stroke for both subsamples and was removed. One item from the Chedoke-McMaster Stroke Assessment (Shoulder abduction in 90° pronation) demonstrated uniform DIF by number of comorbid conditions and was deleted, its deletion improving the overall fit of the model. Moreover, several shoulder items remained in the bank. One item of the SIS (carrying heavy objects) demonstrated uniform DIF by gender in only one of the samples. It was kept in its original form until it was further tested for an adaptive measure using both samples simultaneously.

Targeting

Figure 6.3 shows the distributions of persons (top) and items (bottom) for the bank of upper extremity items. The mean person location value is 1.404 and this suggests that the bank of items is well-targeted to stroke patients as it is reasonably close to the mean location of the items (0).

Properties of the final banked items

The final 49 items in the bank cover a wide range of difficulty level from the most difficult located at 5.603 logits which represent moving 69 blocks or more on the Box and Block Test. The easiest item is a bilateral task of the TEMPA: *unlocking and opening a container*. The logit associated with being unable to perform this task is located at -8.179 on the upper extremity function continuum. Summaries of the global fit statistics for the two random subsamples of persons are presented in Tables 6.7 and 6.8 respectively. Table 6.9 presents the statistical characteristics of the items in the bank for sample number 1.

Reliability

The person separation index for the final model was 0.91 indicating that the items in the bank have good person separation reliability.

The subject measures and the item thresholds distribution are presented in Figure 6.4. Easy items are located towards the bottom of the graph while hard items are at the top. Likewise, persons with less upper extremity function are at the bottom of the graph while persons with more upper extremity function are located at the top.

Verifying the Unidimensionality of the New Measure – Post-Hoc Tests

Unidimensionality was then ascertained with Smith's independent t-test where person estimate from specific subsets of items differed from person estimate derived for the complete set of items. The specific subsets chosen were items assessing *function*, items assessing *activity*, each subset of items that emerged from a specific test or index. The t-tests were found to be not significant, ascertaining the unidimensionality of the items.

Construct Validity

Construct validity is demonstrated by unidimensionality and by the reliability of the item hierarchy across persons. Also, the consistency of the hierarchy of the participants' response patterns across the items was demonstrated by an adequate person reliability index of 0.90602, 0.89553 for random samples respectively. The persons fit residuals ranged from -1.936 to 2.452 for sample 1 and from -1.681 to 2.337 for the second sample. Moreover, the stability of the hierarchy of the item difficulties was demonstrated across the two random samples. A Friedman's test was performed and the associated p-value was 0.1161; not sufficiently low (<0.05) to indicate that the distributions of the locations of the items in the two samples are different.

Content Validity

Because most of the items in the pool of items originate from standardized measures of upper extremity function whose validity and reliability have already been tested, we can be satisfied that the items chosen will indeed measure upper extremity function. As well, the items in the bank are representative of the ICF domains *Body functions* and *Activity*. However, no items representing *Participation* were retained. The items cover a broad spectrum of difficulty of the thresholds ranging from -8.179 and -7.984 to 5.603 and 7.592 for the first and second samples respectively.

There are 35 items that require observation and rating on the part of a therapist and 14 self-report items pertaining to the level of difficulty or how the person feels they are able to carry out a specific task (the Barthel Index is included in the 14 items although sometimes it is scored through observation by the therapist). Items representing the ICF domain of *Body function* cover movements at the shoulder level as well as at the level of the hand. There are unilateral as well as bilateral tasks. Items representing the *Activity* domain cover self-care as well as domestic life. While some items may seem redundant, such as the Barthel 4 – *Dressing and Undressing* and the SIS5B – *In the past 2 weeks, how difficult was it to dress the top part of your body?*, in fact, the Barthel 4 evaluates if the person is able to do the tasks and the SIS item is concerned with how the persons feels about the level of difficulty in accomplishing the task. Both are very important to capture.

Generalizability

In some of the studies, only a few items were administered so in the item reduction process, some records became invalid because persons were only administered items that were deleted or persons became 'extreme' if they had a perfect score on the items that were kept. This caused a reduction in the sample size used for the estimation process as these invalid records and extreme scores are not used in the estimation of item and person parameters. Table 6.10 presents demographic and clinical characteristics of persons in the final sub-samples. Approximately 40% of the persons were aged under 69, 30% between 70 and 79 and about 30% were aged 80 and above. Males represented 64% and 62% of the first and second subsamples respectively. Approximately 50% had more than three comorbid conditions, over 60% had an ischemic stroke, and more than half had their dominant hand affected by the stroke. All of the studies were represented by the final subsamples. Six percent of persons were drawn from the Walking Competency study, 5% from No Place Like Home, 4% from Bridging the Gap, 4% from Recovery from Stroke, 7% from Brain Capacity, 22% from Quality of Life, 22% from Long Term Outcome of Stroke and 31% from both cohorts of the Canadian Stroke Registry.

The iterative process used to estimate the position of persons and items along the continuum of upper extremity function results in the elimination of persons because having eliminated items, some people no longer have data to contribute to the estimation process. Starting with the 99 items and 4058 individuals, the iterative process resulted in 49 items and 1636 individuals. Because of the large sampe size available to this study, this reduction does not affect estimation accuracy.

Conclusion

A unique bank of forty-nine upper extremity function items to assess upper extremity function post-stroke was created. The items contained in the bank span a wide range of upper extremity ability and are representative of the domains of the ICF *Body Structures and Function* and *Activities.* Included in the bank were both unilateral and bilateral tasks. The inclusion of bilateral tasks is important as they are an indication of the degree to which the affected limb contributes to the task and reintegration of the affected arm is important in the recovery process.[383,384] Some of the items evaluate the degree of recovery of movement at specific joints, others assess manipulation, reaching and grasping while others assess activities such as bathing, dressing, doing housework and carrying bags, a good indication of content validity. Rasch analysis also contributed to confirmation of the latent construct of upper extremity function. By fitting items from the two main domains of the ICF, both capacity and performance items as well as uni- and bilateral confirmed all these are important to capture when measuring upper extremity function and are really part of that construct. This is in agreement with the definition Barreca and colleagues developed:[38] *'The main purpose of the arm and hand is to move as an integrated unit in various directions so as to stabilize, reach, grasp, and manipulate objects of various sizes and weights repeatedly in order to perform basic life skills such as feeding, dressing personal care, domestic chores, mobility, and communication. Functional use of both arms allows the client optimal completion of various activities of living, work, and leisure.'*

After recoding of some items to achieve ordered response categories, removal of misffing persons and items, the global fit statistics revealed that the data fit the Rasch model. Initial psychometric evaluation of this bank indicates that construct validity was achieved through fit of the items and persons to the model as well as by an excellent person reliability index. Among the indices rescored, the SF-36 and the SIS had been rescored by other researchers as well in the same manner as done here, by collapsing the middle categories.[90,188] An interesting fact is that most of the test and indices had to be rescored by having most of the easiest categories collapsed. This illustrate how inadequate the simple summation of scores in Classical Test Theory is since the distances between each of the categories are unknown and the distance between them not really equal to 1 as most response options are using.

Despite the need to eliminate persons from the analyses, the final study samples (random samples 1 and 2) showed characteristic in terms of age and proportion of males and with ischemic strokes, similar to stroke populations described in the literature.[385]

This bank forms a basis for and opens the way for the future of measurement of upper extremity function in clinical settings and research. By a process of equating and anchoring, new items can be added and calibrated to expand the bank and improve the psychometric properties of the items. It also opens the way to Computer Adaptive Testing in which a special algorithm will determine which items are going to be administered based upon previous responses. Instead of being presented with all the items in one test, patients will only be presented items that are the most informative and this will shorten considerably the evaluation time required to determine the patient's level of ability in upper extremity function. Chances are that this will also decrease the amount of stress and fatigue for participants and therapists alike.

Another difference with the more commonly used standardized evaluations is that each patient is presented a unique test. Although the items come from a common bank, the different combinations of items presented to the patient will be chosen based on the patient's level of ability and by the computer through the adaptive algorithm.

We now have the potential to substantially advance the field of health status assessment by constructing and calibrating assessments based on Rasch measurement models and other latent trait models and administering those using computerized adaptive methods. The rapid emergence of new computer technologies will make the collection of data in the electronic medium increasingly feasible and cost-effective. Tailoring the instrument to the characteristics of the respondents should produce the best instrument. Furthermore, the use of scale-free measurement of upper extremity recovery in rehabilitation will allow the identification of the smallest measurable, observable, and meaningful differences in function, criteria essential for the evaluation of treatment strategies, interventions and programs.

Limitations

This study used indices that only contained few upper extremity items (e.g. SF-36). Factor analyses were performed on relatively small samples sizes that can render them unstable. Also, as item reduction proceeded, the were eliminated due to extreme scores. Another limitation is that the number of tests/indices of upper extremity function is limited to those used in the studies and even though the items capture a wide range of difficulty, some gaps appear in some locations, making the estimation of the ability of persons located near or within those gaps less precise. Also, no participation items are included in the final bank of items. Very few participation items were included in the initial pool of items (RNL4 and 6) and these items simply did not fit with the rest. It may have been because when participants are asked about their limitations in work and activities, they may not be considering only their upper extremity but also their ability to walk and get around or any other factor that may affect their ability to participate in life roles. Because of the nature of the linking design where there are blocks of missing responses, Principal Component Analysis of the residuals could not be calculated directly by the software. This also prevented selecting subsets of item combinations from the person residuals grid and estimating specific sets of correlations because the samples sizes that could be retrieved from the matrices were very small and no stable factor analyses can be performed on such small sample sizes.

Future work

This bank of items is the first step toward creating a Computer Adaptive version of a measure of upper extremity function. As the items will be presented to participants, they will be re-calibrated, their properties can be further examined (e.g. the presence of DIF) and this will make them more stable. As mentioned earlier, new items to fill the gaps can be added to make the bank more comprehensive. Adding participation items would greatly enhance the comprehensives and usefulness of this bank. Because very few participation items are specific to the use of the upper extremity, it might prove necessary to create new item "from scratch". Items that include 'upper extremity' in the statement for example: Does your affected upper extremity limit you in your participation in your work or activities?

Having a bank of items available makes it possible to add or modify items as subjects or patients are being assessed and re-calibrating these items on a continuous basis. This will produce an ideal measure of upper extremity function, tailored to each individual, comprehensive and psychometrically sound.

Table 6.1 Structure of the Combined Data Set Based on Eight Studies.*

Studies / Tests/ Indices	Walking Competency (N = 91)[109]	Brain Capacity (N=269) (CSN funded)	Recovery from Stroke (N = 59)[118]	Bridging the Gap (N = 195)[371]	No Place Like Home (N = 118)[369,370]	Long Term Outcome of Stroke (N = 606)[368]	Quality of Life (N = 682)[366,367]	Canadian Stroke Registry (2 cohorts) (N = 1372/683) (CSN funded)
BBT	■	■	■					
Barthel Index	■	■	■	■	■	■		
Grip Strength	■	■	■					
STREAM	■	■	■		■			
TEMPA	■	■						
EQ-5D	■	■				■	■	■
NHPT	■		■					
SF-36	■	■		■	■	■	■	
OARS-IADL	■			■	■	■		
CNS			■	■	■		■	
PBSI		■					■	■
RNL		■		■	■	■		
SIS		■					■	■
FRENCHAY			■					
CMSA		■						
HUI		■	■				■	■
RANKIN								■

* A Shaded Cell indicates that the specific test (row) was administered as part of the study (column).
CSN; Canadian Stroke Network, BBT; Box and Block Test, STREAM; STroke REhabilitation Assessment of Movement, TEMPA; Test Évaluant la Performance des Membres Supérieurs des Personnes Agées, EQ-5D; EuroQol-5d, NHPT; Nine Hole Peg Test, SF-36; Medical Outcomes Study 36-Item Short form Questionnaire, OARS-IADL; Older Americans Resources and Services Scale-

Instrumental Activities of Daily Living, CNS; Canadian Neurological Scale, PBSI; Preference Based Stroke Index, RNL; Reintegration to Normal Living Index, SIS; Stroke Impact Scale, FAT; Frenchay Arm Test, CMSA; Chedoke-McMaster Stroke Assessment, HUI; Health Utility Index, Rankin, Ranking Index.

Table 6.2 Items Chosen to Represent Upper Extremity Function.

ICF Component	Tests and Indices	Scoring and categorization
Function	**Box and Blocks Test (BBT)**	**Number of blocks** 0=0-29; 1=29-49; 2=49-59; 3=59-69; 4=69-79 5=79-100
Function	**Canadian Neurological Scale (CNS)** Arm Proximal Arm Distal	**Weakness** 0= none 1 mild, 2 significant, 3 total

Function	**Chedoke McMaster Stroke Assessment (CMSA) ARM**	**0** Unable 1 Able
CHE_3	Not yet stage 2	
C3_2a	Resistance to passive shoulder abduction or elbow extension	
C3_2b	Facilitated elbow extension	
C3_2c	Facilitated elbow flexion	
C3_3a	Touch opposite knee	
C3_3b	Touch chin	
C3_3c	Shoulder shrugging > ½ range	
C3_4a	Extension synergy, then flexion synergy	
C3_4b	Shoulder flexion to 90°	
C3_4c	Elbow at side, 90° flexion: supination, then pronation	
C3_5a	Flexion synergy, then extension synergy	
C3_5b	Shoulder abduction to 90° with pronation	
C3_5c	Shoulder flexion to 90°: pronation then supination	
C3_6a	Hand from knee to forehead 5X in 5 sec.	
C3_6b	Shoulder flexion to 90°: trace a figure 8	
C3_6c	Arm resting at side of body: raise arm overhead with full supination	
C3_7a	Clap hand overhead, then behind back 3X in 5 sec.	
C3_7b	Shoulder flexion to 90°: scissor in front 3X in 5 sec.	
C3_7c	Elbow at side, 90° flexion: resisted shoulder external rotation	
	HAND	
CHE_4	Not yet stage 2	
C4_2a	Positive Hoffman	
C4_2b	Resistance to passive wrist of finger extension	
C4_2c	Facilitated finger flexion	
C4_3a	Wrist extension >1/2 range	

Table 6.2 Items Chosen to Represent Upper Extremity Function. (cont.).

ICF Component	Tests and Indices	Scoring and categorization
C4_3b	Finger/wrist flexion >1/2 range	
C4_3c	Supination, thumb in extension: thumb to index finger	
C4_4a	Finger extension then flexion	
C4_4b	Thumb extension >1/2 range, then lateral prehension	
C4_4c	Finger flexion with lateral prehension	
C4_5a	Finger flexion, the extension	
C4_5b	Pronation: finger abduction	
C4_5c	Pronation: tap index finger 10X in 5 sec	
C4_6a	Pistol grip: pull trigger, then return	
C4_6b	Pronation: wrist and finger extension wit finger abduction	
C4_6c	Thumb to finger tips, then reverse 3X in 12 sec.	
C4_7a	Bounce a ball 4 times in succession, then catch	
C4_7b	Pour 250 mL from 1 L pitcher, the reverse	
C4_7c	Pronation: tap index finger 10X in 5 sec	
Function	**Grip Strength**	**Kilograms of Force:** **0**=0-27; **1**=27-40; **2**=40-50; **3** >50
Function	**Nine Hole Peg Test (NHPT)**	**Continuous (time in seconds)** **3**=33.6 >= time > 10.0 **2**=45.8100 >= time>33.6000 **1** > 45.8100 **0**=unable
Function	**Rankin Index (RI)**	**0** No symptoms **1** No disability **2** Slight disability **3** Moderate disability **4** Moderately severe disability **5** Severe disability, bedridden

Table 6.2 Items Chosen to R epresent Upper Extremity Function. (cont.).

ICF Component	Tests and Indices	Scoring and categorization
Function	**Stroke REhabilitation Assessment of Movement (STREAM)** 1 Protract Scapula 2 Extend elbow 3 Shrug shoulders 4 Raise hand to top of head 5 Hand on sacrum 6 Raise arm overhead 7 Supinate pronate 8 Close hand 9 Open hand 10 Opposition	**0** unable **1** abnormal **2** normal
Activity	**Barthel Index** 1. Feeding 2. Personal Hygiene 3. Bathing 4. Dressing and Undressing	**2** Independent **1** Assistance **0** Unable
Activity	**EuroQoL-5D (EQ-5D)** 2. Self-care 3. Usual Activities	**0** No Problems **1** Some Problems **2** Unable
Activity	**Frenchay Arm Test (FAT)** Ruler/pencil Grasp cylinder Pick up glass and drink Spring clothes peg Comb hair	**0** Unable **1** Able1
Activity	**Health Utilities Index (HUI)** 24. Full use of hands 25. Need help due to hand/fingers 26. Need help for tasks 27. Special equipment 28. Eat, bathe, dress, use toilet 29. Need help Eat bathe dress use toilet 30. Special Equipment Eat bathe dress use toilet	**1** Yes **0** No **3** Don't know **3** Refused

Table 6.2 Items Chosen to Represent Upper Extremity Function. (cont.).

ICF Component	Tests and Indices	Scoring and categorization
Activity	**Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36)** Sf3c Pick up/lift Grocery bags Sf3j Take bath or dress	**1** very limited **2** limited a little **3** not limited at all
Activity	**Older Americans Resources and Services Scale-Instrumental Activities of Daily Living (OARS-IADL)** 4. Prepare meals 5. Housework	**2** Without help **1** With help **0** Unable
Activity	**Preference-Based Stroke Index (PBSI)** 4. Recreational Activities 5. Work/Activities 6. Driving	**0** All **1** Some **2** Unable
Function	**Stroke Impact Scale (SIS)** 1A Strength arm 1B Strength hand	**1** No Strength **2** A little **3** Moderate **4** Quite a bit **5** A lot
Activity	**SIS (cont.)** 5a. Cut food 5b. Dress upper body 5c. Wash (bath/shower) 5d. Cut toe nails 5h. Housework (light) 5j. Housework (heavy) 7a. Carry heavy objects 7b. Turn doorknob 7c. Open can or jar 7d. Tie laces 7e. Pick up money	**1** Unable **2** Very difficult **3** Moderate **4** A little difficult **5** Not difficult
Activity	**Test d'évaluation des Membres Supérieurs chez les Personnes Âgées (TEMPA)** Pick up and move a jar Open jar and remove spoonful of coffee Pick up pitcher and pour water in glass Unlock a lock and open pill container Write on envelope and stick stamp Tie scarf around neck Shuffle and deal playing cards Handle coins Pick up and move small objects	**3** normal **2** hesitation **1** difficulty-more than 25% **0** unable less than 25%

Table 6.2 Items Chosen to Represent Upper Extremity Function (cont.)

ICF Component	Tests and Indices	Scoring and categorization
Participation	**Reintegration to Normal Living (RNL)** 4. Self-care needs 6. Recreational Activities	**0** Yes **1** Partially **2** No

BBT; Box and Block Test, STREAM; STroke REhabilitation Assessment of Movement, TEMPA; Test Évaluant la Performance des Membres Supérieurs des Personnes Agées, EQ-5D; EuroQol-5d, NHPT; Nine Hole Peg Test, SF-36; Medical Outcomes Study 36-Item Short form Questionnaire, OARS-IADL; Older Americans Resources and Services Scale-Instrumental Activities of Daily Living, CNS; Canadian Neurolocial Scale, PBSI; Preference Based Stroke Index, RNL; Reintegration to Normal Living Index, , SIS; Stroke Impact Scale, FAT; Frenchay Arm Test, CMSA; Chedoke-McMaster Stroke Assessment, HUI; Health Utility Index, Rankin, Ranking Index

Table 6.3 Categorization of Influencing (Personal) Factors.

Factors	Categorization
Age	**0** = 0 to 69 years of age **1**= 70 to 79 years of age **2** = over 80 years of age
Gender	**0** = Male **1**= Female
Type of Stroke	**0** = Ischemic **1** = Hemorrhagic **2** = Not defined or missing
Number of Comorbid Conditions	**0** = No comorbid conditions **1** = 1 or 2 comorbid conditions **2** = More than 3 comorbid conditions
Side of Hemiplegia	**0** = Left **1** = Right **2** = Bilateral **3** = Missing **4** = None noted
Time of Assessment Post-Stroke	**0** = Stoke Onset **1** = 1month post-stroke **2** = 3 months post-stroke **3** = 6 months post-stroke **4** = 7 months post-stroke **5** = 12 months post-stroke
Dominant Hand Affected	**0 = Yes** **1 = No** **2 = Missing**

Table 6.4 Demographic and Clinical Characteristics of Study Participants.

Subject Characteristic (N = 4058)	
Age Mean (SD)	68 (15)
Gender No. (%) male	2206 (54)
Number of Comorbid conditions No. (%)	
0	803 (20)
1-2	1470 (37)
3-4	1556 (40)
>4	110 (3)
Type of CVA No. (%)	
Ischemic	2328 (59)
Hemorrhagic	415 (11)
Missing	116 (2)
Not Noted/Determined	1090 (28)
Side of Hemiplegia No. (%)	
Right	1475 (37)
Left	1467 (37)
Bilateral	234 (6)
Missing	447 (11)
Not Noted	316 (8)
Dominant UE affected No. (%)	1375 (42)

Table 6.5 Items Loading on a Separate Factor for Each of the Studies.

Studies	Walking Competency	Brain Capacity	Recovery from Stroke	Bridging the Gap	No Place Like Home	Long Term Outcome of Stroke	Quality of Life	Canadian Stroke Registry Cohort 1/2
KMO	0.88	0.90	0.89	0.70	0.84	0.89	0.93	0.95/0.91
Variance explained by 1st factor (%)	60	58	80	75	60	89	66	83/84
Items	EQ2	EQ2	None	CNS5	BART2	RNL6	SF3C	HUI27/HUI27
	EQ3	EQ3		CNS6	BART3		PBSI6	HUI28/Rankin
	SF3J	PBSI4		NHPT	BART4		CNS5	PBSI4
	SF3C	PBSI5			SF3C		CNS6	
		PBSI6			SF3J		HUI24	
		HUI24					HUI25	
		HUI25					HUI26	
		HUI26					HUI27	
		HUI27					HUI28	
		HUI28					HUI29	
		HUI29					HUI30	
		HUI30						

BBT; Box and Block Test, STREAM; STroke REhabilitation Assessment of Movement, TEMPA; Test Évaluant la Performance des Membres Supérieurs des Personnes Agées, EQ-5D; EuroQol-5d, NHPT; Nine Hole Peg Test, SF-36; Medical Outcomes Study 36-Item Short form Questionnaire, OARS-IADL; Older Americans Resources and Services Scale-Instrumental Activities of Daily Living, CNS; Canadian Neurolocial Scale, PBSI; Preference Based Stroke Index, RNL; Reintegration to Normal Living Index, , SIS; Stroke

Impact Scale, FAT; Frenchay Arm Test, CMSA; Chedoke-McMaster Stroke Assessment, HUI; Health Utility Index, Rankin, Ranking Index

Table 6.6 Deleted Items and Reason for Deletion.

Reason for Deletion	Deleted Items*
Not loading on upper extremity factor for 1 or more studies and lack of fit to the model	EQ2, EQ3; CNS5&6 HUI24, HUI25, HUI26, HUI27, HUI28, HUI29, HUI30 PBSI4, PBSI5, PBSI6
Lack of fit to the model	NHPT SIS5J, SIS7C SIS5D, SIS1A, SIS1B, SIS5C RNL4 RNL6 RANKIN BARTHEL2 FRENCHAY STREAM10, STREAM11, STREAM14 CHE_3 CHE_4 C3_2a,b,c C3_3a,b,c C3_5a, b C3_6b C3_7b C4_2a,b,c C4_3a,b,c C4_4a,c C4_6c
Redundancy	SIS7C, STREAM10, STREAM11, STREAM14
Differential Item Functioning	C3_5b Grip strength for males

*Please refer to Table 6.2 for meaning of abbreviated item names

Table 6.7 Summary of Global Fit Statistics for the Item Bank for Sample No.1.

	ITEMS		PERSONS	
	Location	Fit Residual	Location	Fit Residual
Mean	0.000	-0.464	1.407	-0.285
Standard Deviation	1.818	0.835	2.130	0.801
Skewness		0.734		0.580
Kurtosis		-0.147		-0.104
Correlation		0.000		-0.094

ITEM-TRAIT INTERACTION		RELIABILITY INDICES	
Total Item Chi-Square	459.570	**Separation Index**	0.90602
Total Degrees of Freedom	440.000	**Cronbach Alpha**	N/A
Total Chi-Square Probability	0.250594		

POWER OF TEST-OF-FIT	**Power is EXCELLENT** Based on Separation Index of 0.90602

Table 6.8 Summary of Global Fit Statistics for the Item Bank for Sample No.2.

	ITEMS		PERSONS	
	Location	Fit Residual	Location	Fit Residual
Mean	0.000	-0.465	1.622	-0.347
Standard Deviation	1.982	0.758	2.138	0.792
Skewness		0.944		0.739
Kurtosis		0.350		0.087
Correlation		0.000		-0.169

ITEM-TRAIT INTERACTION		**RELIABILITY INDICES**	
Total Item Chi-Square	480.951	**Separation Index**	0.89553
Total Degrees of Freedom	437.000	**Cronbach Alpha**	N/A
Total Chi-Square Probability	0.071873		

POWER OF TEST-OF-FIT	**Power is EXCELLENT** Based on Separation Index of 0.89553

Table 6.9 Characteristics of the Items in the Upper Extremity Function Item Bank. Sample No1.

Item Code	Item description	Location	SE	Fit Residual	Chi-Square statistic*	F-statistic*
T4F	*Unlock a lock and open a pill container (bilateral task)*	-4.264	0.394	1.48	13.47	0.678
BART1	*Feeding*	-2.991	0.134	-1.098	9.381	1.049
T6F	*Tie a scarf around one's neck (bilateral task)*	-2.834	0.376	0.205	7.931	0.847
BART4	*Dressing and Undressing*	-2.282	0.125	-1.474	8.7	1.117
T2F	*Open a jar and remove a spoonful of coffee (bilateral task)*	-2.172	0.361	-0.133	4.067	0.44
T5F	*Write on an envelope and stick a stamp on it (bilateral task)*	-2.056	0.349	0.703	13.678	1.195
SF3J	*Does your health now limit you bathing or dressing yourself? If so, how much?*	-1.829	0.174	0.179	11.183	1.189
C4_5B	*Pronation: finger abduction*	-1.617	0.739	-0.747	2.729	1.373
C4_4B	*Thumb extension >1/2 range, then lateral prehension*	-1.466	0.693	-0.394	2.95	0.669
C3_4A	*Extension synergy then flexion synergy*	-1.367	0.665	0.092	9.062	0.928

Table 6.9 Characteristics of the Items in the Upper Extremity Function Item Bank. Sample No1. (cont.).

Item Code	Item description	Location	SE	Fit Residual	Chi-Square statistic	F-statistic
C4_5A	*Finger flexion then extension*	-1.316	0.651	0.305	14.538	0.979
C3_4B	*Shoulder flexion to 90°*	-1.284	0.643	-0.946	4.926	2.749
STR13	*Open hand from fully closed position*	-1.138	0.234	-0.987	8.928	2.255
STR2	*Extends elbow in supine (starting with elbow fully flexed)*	-1.056	0.231	-1.12	6.862	1.782
STR12	*Closes hand from fully opened position*	-0.978	0.228	-1.339	7.974	2.233
T3F	*Pick up a pitcher and pour water into a glass (unilateral task)*	-0.863	0.402	-1.072	7.621	2.327
T7F	*Shuffle and deal playing cards (bilateral task)*	-0.771	0.358	-0.425	6.07	0.821
C3_4C	*Elbow at side 90° flexion: supination then pronation*	-0.744	0.521	-0.763	2.392	0.788
T9F	*Pick up and move small objects (unilateral tasks)*	-0.635	0.398	-1.169	6.65	1.849
T8F	*Handle coins (unilateral tasks)*	-0.612	0.398	-1.151	10.446	3.996
STR9	*Place hand on sacrum*	-0.417	0.213	-0.674	6.876	1.08

Table 6.9 Characteristics of the Items in the Upper Extremity Function Item Bank. Sample No1. (cont.).

Item Code	Item description	Location	SE	Fit Residual	Chi-Square statistic	F-statistic
STR1	*Protract scapula in supine*	-0.321	0.211	-0.91	13.801	2.626
SIS5B	*In the past 2 weeks, how difficult was it to dress the top part of your body?*	-0.307	0.122	-1.138	16.493	2.297
OARS4	*Can you prepare your own meals?*	-0.286	0.15	-0.634	5.835	0.745
T1F	*Pick up and move a jar (unilateral task)*	-0.238	0.395	-1.004	8.764	2.175
STR7	*Shrugs shoulders (scapular elevation)*	-0.197	0.208	1.255	19.873	1.687
STR8	*Raises hand to touch top of head*	-0.174	0.207	0.728	19.819	1.5
SIS5A	*In the past 2 weeks, how difficult was it to cut your food with a knife and fork?*	-0.029	0.132	-0.082	16.695	2.096
C4_5C	*Hand unsupported: opposition of thumb to little finger*	0.254	0.381	-0.163	6.835	1.042
BART3	*Bathing*	0.283	0.141	-0.672	6.453	0.926

Table 6.9 Characteristics of the Items in the Upper Extremity Function Item Bank. Sample No1. (cont.).

Item Code	Item description	Location	SE	Fit Residual	Chi-Square statistic	F-statistic
SIS7E	*In the past 2 weeks, how difficult was it to use your hand that was most affected by your stroke to pick up a dime?*	0.303	0.124	-1.196	11.9	1.452
SIS5H	*In the past 2 weeks, how difficult was it to do light household tasks/chores (e.g. dust, make a bed, take out garbage, do the dishes)?*	0.341	0.113	-0.133	19.726	2.245
C3_5C	*Shoulder abduction to 90^o: pronation then supination*	0.415	0.366	-1.14	2.711	0.597
C4_6B	*Pistol grip: pull trigger then return*	0.423	0.365	-1.151	2.708	0.584
SIS7B	*In the past 2 weeks, how difficult was it to use your hand that was most affected by your stroke to turn a doorknob ?*	0.624	0.122	-0.312	13.495	1.628
C4_6A	*Pronation: tap index finger 10X in 5 sec.*	0.943	0.326	-0.9	6.772	1.381
C3_6A	*Hand from knee to forehead 5X in 5 sec.*	1.003	0.322	-1.307	3.53	0.72

Table 6.9 Characteristics of the Items in the Upper Extremity Function Item Bank. Sample No1. (cont.).

Item Code	Item description	Location	SE	Fit Residual	Chi-Square statistic	F-statistic
SF3C	*Does your health now limit you lifting or carrying groceries? If so, how much?*	1.073	0.124	1.564	20.455	1.421
OARS5	*Can you do your own housework?*	1.096	0.155	0.427	8.647	0.668
C3_6C	*Arm resting at side of body: raise arm overhead with full supination*	1.291	0.307	-1.01	4.521	0.799
SIS7A	*In the past 2 weeks, how difficult was it to use your hand that was most affected by your stroke to carry heavy objects (e.g. bag of groceries)?*	1.634	0.11	1.296	14.001	1.463
SIS5D	*In the past 2 weeks, how difficult was it to clip your toenails?*	2.186	0.123	0.232	8.503	0.903
C4_7C	*Pour 250mL from 1 L pitcher then reverse*	2.607	0.268	-2.033	9.408	2.02
C3_7C	*Elbow at side, 90o flexion: resisted shoulder external rotation*	2.634	0.268	0.302	4.465	0.5

Table 6.9 Characteristics of the Items in the Upper Extremity Function Item Bank. Sample No1. (cont.).

Item Code	Item description	Location	SE	Fit Residual	Chi-Square statistic	F-statistic
BBT	*Number of blocks transferred in 60 seconds*	2.714	0.209	-1.624	10.837	1.864
C4_7A	*Thumb to finger tips, then reverse 3X in 12 sec*	3.092	0.264	-0.78	4.365	0.591
C3_7A	*Clap hands overhead then behind back 3X in 5 sec.*	3.209	0.263	-0.688	11.304	1.681
FEMALE	*Grip Strength (females)*	3.466	0.379	-0.202	9.047	1.727
C4_7B	*Bounce ball 4 times in succession then catch*	4.65	0.285	-0.925	12.171	2.479

* All chi-square and F-Statistics were statistically non-significant after Bonferroni correction.

Table 6.10 Demographic and Clinical Characteristics of Persons.

Subject Characteristic (N= 843)	1st Sample (N=812)	2nd Sample (N = 824)
Age (%) 0 to 69/70 to 79 /over 80 /missing	42/32/25/2	40/29/29/2
Gender No. (%) male	64	62
Number of Comorbid conditions. (%) 0/1-2/>3	19/35/47	18/30/51
Type of CVA (%) Ischemic/Hemorrhagic/Other, not defined or missing	62/6/32	60/8/32
Side of Hemiplegia (%) Right / Left /Bilateral / None	37/39/8/9	37/38/9/8/6
Dominant Hand affected (%)	57	56

Figure 6.1 Simplified Rasch Model Overview (adapted from Wright et al.,[386])

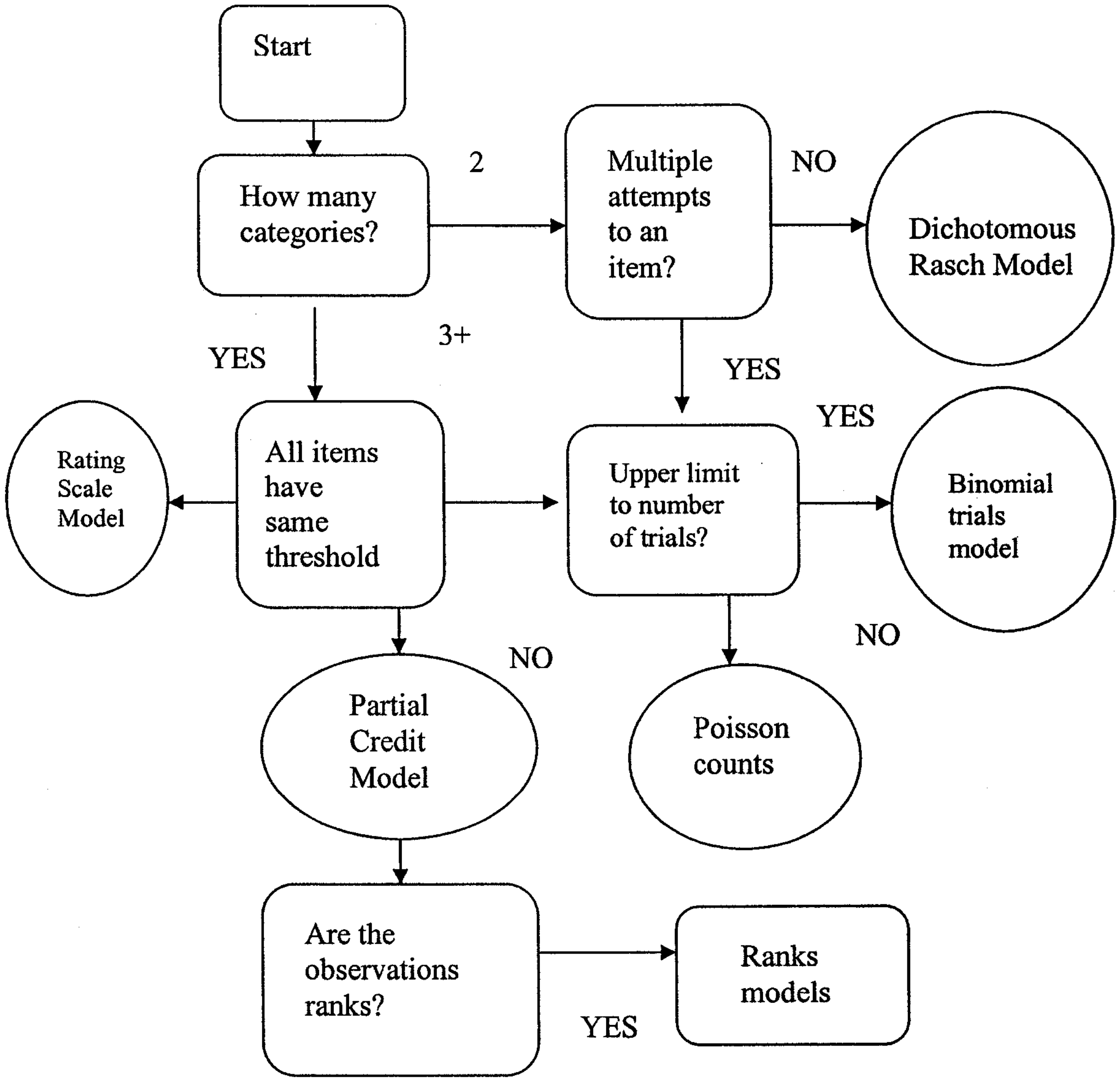


Figure 6.2 Final Threshold Ordering

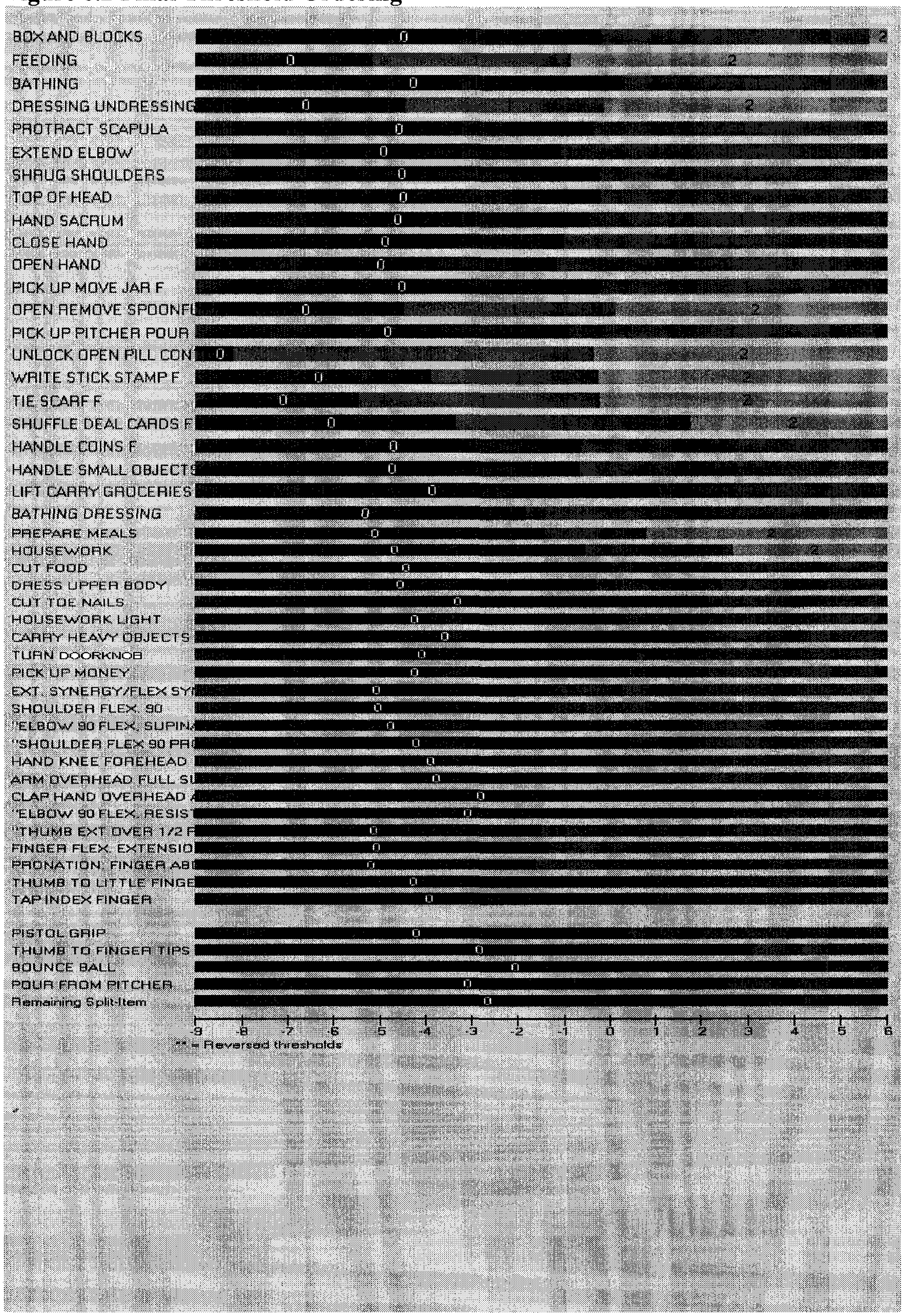
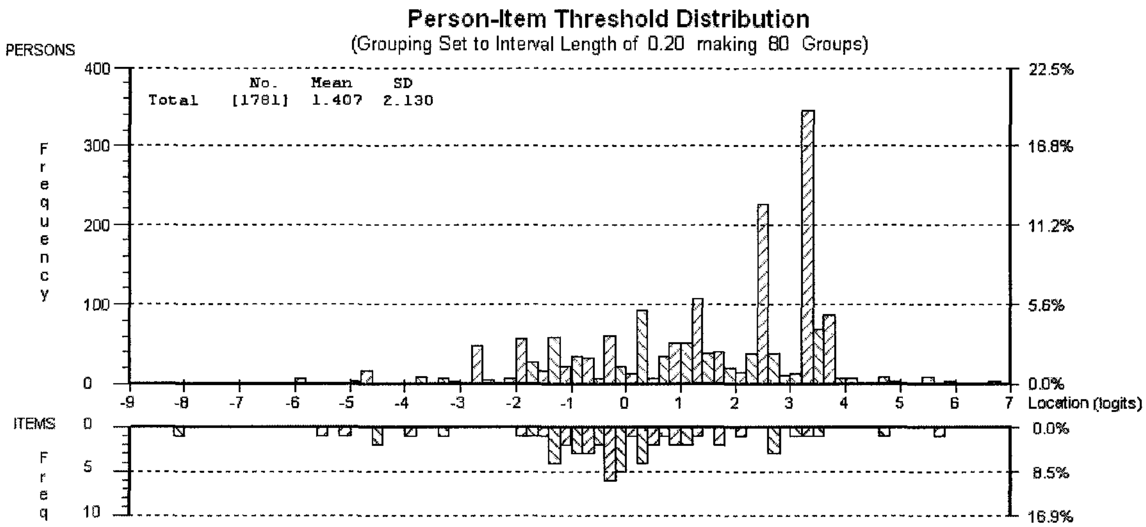


Figure 6.3 Person and Item Distribution for Sample No.1.

```
PERSONS              ITEMS [uncentralised thresholds]
 6.0                  |
                      |
                      | BBT.2
 5.0                  |
                      |
                     o| C4_7B.1
 4.0                  |
                      |
                 ooooo|
           .      oooo| EMALE.1
.ooooooooooooooooooooo| C3_7A.1
 3.0                 o| C4_7A.1
                     o|
                    oo| C4_7C.1  C3_7C.1  OARS5.2
         ooooooooooooo|
                    oo|
 2.0                 o| SIS5D.1
                     o|
                 .  oo| SIS7A.1   T7F.2
                    oo|
                oooooo| C36_C.1
 1.0               ooo| C3_6A.1  SF3C.1
                   ooo| OARS4.2  C4_6A.1
                    oo| SIS7B.1
                      | C3_5C.1  C4_6B.1
                 ooooo| C4_5C.1  BART3.1  SIS7E.1  SIS5H.1
 0.0                 o| T2F.2
                     o| STR7.1   BBT.1  STR8.1  BART4.2  SIS5A.1
                   ooo| T4F.2  STR1.1  SIS5B.1   T1F.1   T5F.2   T6F.2
                      | OARS5.1  STR9.1
                    oo| C3_4C.1   T9F.1   T8F.1
-1.0                oo| STR12.1   T3F.1  BART1.2
                     o| STR13.1  STR2.1
                   ooo| OARS4.1  C3_4A.1  C4_5A.1  C3_4B.1
                     o| C4_B.1
                    oo| C4_5B.1
-2.0               ooo| SF3J.1
                      |
                   ooo|
-3.0                  |
                      | T7F.1
                      |
                     o|
-4.0                  | T5F.1
                      |
                      | T2F.1  BART4.1
                     o|
-5.0                  |
                      | BART1.1
                      |
                      | T6F.1
-6.0                  |
                      |
                      |
-7.0                  |
                      |
-8.0                  |
                      | T4F.1
```

o = 18 Persons

Figure 6.4 Upper Extremity Function and Corresponding Subjects' Distribution (top panel)



CHAPTER 7 Manuscript 4

Preface to Manuscript 4

Development and Psychometric Evaluation of a "Paper-CAT" Measure of Upper Extremity Function Post-Stroke

The previous manuscript of this thesis described the development, through Rasch modeling, of a bank of items designed to measure upper extremity function in persons with stroke. The development of an item bank is the first step towards adaptive testing. When a validated item bank is available, items can be specifically chosen to target the participant's level of ability. Modern psychometric approaches, such as Rasch analysis, are perfectly suited to produce adaptive scales. In adaptive testing, items are usually presented on a computer screen and the score on each item entered directly. Based on the participants answer to the last item performed or answered, the computer, through a pre-programmed algorithm, will choose the next best item to present to the person. Within relatively few items and a short amount of time, a precise ability estimate is obtained. This is possible because calibrated items are invariant and thus their difficulty level is known. Computers are not always readily available in the clinics and both researchers and clinicians may be reluctant to using complex computer programs to administer and score the assessment. The second manuscript of the thesis revealed that very few studies in the field of rehabilitation made use of Rasch-developed measures. There may be a tendency for rehabilitation professionals to use their conventional tests and indices because they are familiar and trusted in terms of their psychometric properties and method of scoring. Developing a paper version of an adaptive test may be a very useful intermediary step between conventional tests and indices and computer adaptive testing. Because it stems from a bank of calibrated items, it offers all the advantages of adaptive testing without the sometimes intimidating aspect of computers and algorithms. As well, the paper format allows the clinician or researcher to view all the items in the bank, and this may increase their level of confidence in the measure.[236]

In this fourth manuscript, a paper version of an adaptive test of upper extremity function was created from the item bank produced in the previous manuscript.

Title Page

Development and Psychometric Evaluation of a "Paper-CAT" Measure of Upper Extremity Function Post-Stroke

Authors: Johanne Higgins, MSc, O.T., Lois E. Finch, MSc, P.T., Jacek Kopec, MD, Ph.D, Nancy E. Mayo, BSc(PT), Ph.D.

Posts of the Authors at the Time of the Study:

Higgins J: PhD candidate, School of Physical and Occupational Therapy, Faculty of Medicine, McGill University.

Finch LE: PhD candidate, School of Physical and Occupational Therapy, Faculty of Medicine, McGill University.

Kopek J: Assistant Professor Health Care and Epidemiology Faculty of Medicine, Mather Building University of British Columbia.

Mayo NE: Associate Professor, School of Physical and Occupational Therapy, Faculty of Medicine, McGill University, Associate member, Department of Epidemiology and Biostatistics and Occupational Health, Faculty of Medicine, McGill University

Work conducted in: Montreal, Quebec, Canada, and Quebec City, Quebec, Canada

Address correspondence to: Nancy Mayo, Royal Victoria Hospital, Division of Clinical Epidemiology, 687 Pine Avenue West, Ross 4.29, Montreal, Quebec, H3A 1A1 CANADA, Telephone: (514) 934-1934 ext. 36922, Fax: (514) 843-1493, e-mail: nancy.mayo@mcgill.ca

Introduction

A bank of items has been created to measure upper extremity function in persons with stroke. The next logical step was to create a format for testing persons using a subset of items best suited to their function. An adaptive measure consists of only the items that are targeted to the individual's level of ability, thus, decreasing burden and discouragement while capturing a precise estimation of ability.

Objective

The objective of this study was to develop a paper version of an adaptive test of upper extremity function.

Methods

Participants

Over 4000 patients with stroke who participated in various randomized controlled trials and observational studies in the Montreal area. Patients were assessed on several different occasions post-stroke depending on the study.

Instruments

A total of 17 tests and indices of upper extremity function were considered for this study. Together, they comprise a total of 99 items. Some of the tests were especially designed to assess upper extremity function. Some items from indices such as the Stroke Impact Scale and the Barthel Index that assess activity which were considered to involve the upper extremity were also included in the pool. As well some items included in indices capturing global motor recovery such as the STroke REhabiliation Assessment of Movement were initially chosen. A calibration of the items was previously performed creating a bank of items measuring upper extremity function. This bank contains a total of 49 items and only those items were used for the development of the present adaptive measure.

Analysis

Data were analyzed to test whether they adhered to the expectations of the Rasch partial credit model. Rasch models are mathematical functions that describe the relation between an individual ability and the probability of responding to a question or performing a task in a category. The Rasch Unidimensional Measurement Models (RUMM 2020)[147] was used to carry out the analyses.

Conclusion: A scale-free adaptive measure of upper extremity function in a paper version has been created. The measure demonstrates adequate initial psychometric properties. Further testing of the measure in its adaptive format is necessary in order to allow the identification of the smallest measurable, observable, and meaningful differences in upper extremity function.

Introduction

At the present time, there are no objective and accurate measures of upper extremity function that assess the range of upper extremity function observed clinically: the most severe to almost complete recovery. Thus, to assess the effectiveness of rehabilitation interventions or for research purposes, rehabilitation professionals need several different evaluations to ensure comprehensive assessment of the construct.

The standard evaluations used at the present time are composed of several test items. When assessing a patient using a specific measure, all the items comprised within this measure need to be presented to the patient. This makes the process of evaluation long and tiring, especially for persons who have suffered a stroke. Furthermore, most existing measures are comprised of item scored on an ordinal scale and thus the simple sum of the scores obtained on the items yields a meaningless metric. Thus, to advance the measurement of upper extremity function, two challenges are evident: (i) covering the complete construct; and (ii) producing a measurement scale with mathematical properties.

The World Health Organization's International Classification of Functioning, Disability, and Health (ICF) has outlined the concepts of functioning and disability. Functioning includes body functions and structures, and *activities and participation*. Items used to create this new measure of upper extremity function are pooled from tests and indices that encompass these two components.

The ideal upper extremity function measure should contain items spanning the whole range of abilities and that is 'custom tailored' to each individual patient so that a relatively small number of items will yield maximum information about the patient's level of upper extremity function. This process avoids having the more recovered patients perform items that are too easy for them and avoids the severely affected patients having to try too many items that are too difficult for them and get discouraged by the successive failures. This is where Rasch analysis and adaptive testing come into play.

Rasch analysis is a method that aligns items along a calibrated hierarchy and as such the units have mathematical properties. The main goal of using an adaptive measure in our situation is to create a measure of upper extremity function using items from a common bank. Because computers are not widely used and accessible in most clinical setting, a 'paper-and-pencil' adaptive scale: a flexilevel scale of upper extremity function was devised. The article describes the development and initial psychometric evaluation of an adaptive measure of upper extremity function and its psychometric properties.

Methods

Study Sample

Data from over 4000 stroke survivors who participated in eight research projects including observational studies and randomized controlled trials were used to construct this new measure and constituted the initial pool of items. Each of the eight studies had different evaluation times post-stroke depending on the design of the particular study. These evaluation schedules are presented in Table 7.1.

Items

Subjects' ratings of upper extremity function on 49 items were available for analysis. Items were originally chosen from different existing tests and indices. Some of the tests are designed to evaluate upper extremity function while some of the indices are meant to assess patients' ability to perform activities using their upper extremity. As well, two items from the SF-36 *"Does your health now limit you lifting or carrying groceries? If so, how much?"* and *"Does your health now limit you Bathing or dressing yourself? If so, how much?"*, a health-related quality of life index were included because they tapped into the person's feeling about how limited they are in performing a task that involves the upper extremity. Items from each of the World Health Organization's International Classification of Functioning, Disability, and Health (ICF), *Body Functions* and Activ*ity and Participation* were included. An effort was made to include as many items as possible, even with evident redundancies in order keep the ones demonstrating the best psychometric properties. A list of all items that are part of the bank is provided in Table 7.2 along with their original scoring or categorization if they were assessed on a continuous scale (Box and Block test, Grip strength and Nine-Hole Peg Test). Also, each item was classified into one of two domains of the ICF.

Personal Factors

Several personal factors, collected as part of the original studies, were considered. They were the person's age, their gender, the type of stroke, the number of comorbid conditions, the side of their hemiplegia, whether their dominant hand was affected by the stroke as well as the time of evaluation post-stroke. The severity of the stroke at onset was also examined. Severity was based on the patient's score on the Canadian Neurological Scale (CNS) at the onset of stroke. The CNS, [387,399] scored from 1.5 (most severe) to 11.5 (least severe), was categorized into four groups: mild if the score was greater than 6.5; mild-moderate with a score between 5 and 6; moderate with a score between 4 and 4.5 and severe with if the score was below 3.5.[53,387] All personal factors and their categorizations are presented in Table 7.3. These factors were chosen because they may have an impact on upper extremity function and how patients perform or respond to certain items.

Analysis

Unidimensionality

Scores on indices of function and activity and participation from 4058 stroke survivors were included in the analyses. One of the key features of Rasch models is that only one unique construct or trait is being measured. In this case 'upper extremity function' is the construct of interest. The item pool was tested through a series of Principal Component Analyses (PCA) for each one of the individual studies using the FACTOR procedure in the Statistical analysis software SAS version 9.1 (SAS institute, 100 SAS Campus Drive, Cary NC 27513) to examine factor loadings and confirm that the items contained in the item pool are all measuring upper extremity function. Parallel analyses were subsequently performed to confirm the number of factors present.

Calibration of the item pool

Items were fit to the partial-credit Rasch model using RUMM2020[147] computer software. The nature of the raw data, items scored on different ordinal scales, led to the selection of the partial credit model. The partial credit model is an extension of the basic Rasch model for dichotomous scoring. It is used when points are awarded for intermediate levels of performance on an item. As all Rasch models, probability functions are based on two parameters: 1) the difficulty of the items and 2) the ability of the person. After removal of the misfitting items and persons, items are placed hierarchically in order of difficulty from easy items to difficult items. In the same manner, persons are place on the same continuum of upper extremity function, from most able to least able. The choice of which items to retain was based mostly on their fit to the model but also on 'clinical decisions', care being taken to ensure that items adequately covered as many upper extremity movements and activities as possible.

Fit to the model can be assessed using different criterion: the fit statistics, the item characteristic curves (ICCs), A PCA of the Rasch model item residuals, and category characteristic curves. Fit of persons must also be assessed through their fit statistics. In the presence of adequate targeting of the items to the sample, the sample size required to perform a Rasch analysis yielding stable person and item estimates (±0.5 logit at the 95% confidence level) and based on an expected standard error level of ±0.1 is 200. [388,389]

Criteria used for the fit of the items and the persons were as follows: standardized fit residuals between +2.5 and -2.5. For the items a non-significant χ^2 (chi-square) and F-statistic were also required. As for the global model fit, a non significant item-trait interaction is necessary. Items and persons are removed in an iterative process using the Total-Item-Person Strategy for Analyzing Fit (TIP). This particular strategy consists of a series of consecutive steps:[377] 1) evaluating person fit, 2) eliminating misfitting persons, 3) recalibrating, 4) evaluating item fit, 5) evaluating the overall fit of data to the model. This strategy is thus the best one to use in the development of the measures.

Ordered Thresholds

Thresholds are those points along a theoretical continuum of item difficulty where the probability of a person responding either 0 or 1, and 1 or 2 respectively, are equally likely.[143] In the case where there are more than two answers or scoring options for the items, the fit of each category must be assessed. Rasch analysis is well suited to investigate the usefulness of the categories. All the indices, including those originally scored on an ordinal scale were scrupulously examined for the quality of the categories in their rating scale. According to Guilford, [163] categorizations should be well-defined as well as mutually exhaustive. In cases where disordered thresholds or un- or uner-utilized categories are present, rescoring by collapsoing categories is necessary.

Differential Item Functioning (DIF)

Fit to the Rasch model not only confirms the unidimensionality of the items but also their invariance. Invariance refers to the situation when estimates of item difficulty vary in their position on the linear continuum according to different populations being assessed (e.g. males vs. females).[167] DIF is a phenomenon that can occur when there is a loss of invariance of item estimates across different groups of individuals or testing occasions.[137] The standard residual of an observed score from the score predicted by the model was calculated for each person (the person-item deviation residual). The participants were divided into ten groups (class intervals) based on their total item scores. DIF was evaluated for time post-stroke. In order to proceed with this analysis, baseline evaluations as well as subsequent evaluations from one month to 7 months within eight different studies were pooled. Each participant at each evaluation time post-stroke was treated as a distinct individual. As well, several additional persons factors were examined for the presence of differential item functioning. They were, the person's age, their gender, the type of stroke, the number of comorbid conditions, the side of their hemiplegia as well as whether their dominant hand was affected by the stroke. The severity of the stroke at onset was also examined.

Psychometric Properties

Several psychometric properties examined or tested within each of the traditional or Rasch framework.

1) Traditional framework

a) Content

This type of validity is present when there is evidence that the test or index is composed of a comprehensive sample of items that completely assess the domain of interest.[390,391]

b) Construct

Content validity is the extent to which the test or index provide results that are consistent with theoretically driven relationships.[27] As no 'gold standard' exists for upper extremity function, convergent, divergent and discriminative approaches were used. Specific hypotheses were developed on the correlation between the new adaptive measure of upper extremity function and other tests or indices.

(1) Low correlation with the mental health and emotional subscale of the SF-36 ($r < 0.2$);

(2) Moderate correlation (<r>0.6) with the STREAM total score.

Additionally, to test whether the new measure of upper extremity function could discriminate between subjects across the four level of stroke severity as measured by the CNS, a General Linear model was performed with the Tukey post-hoc test (significance set at $p<0.05$) using SAS version 9.1 (SAS institute, 100 SAS Campus Drive, Cary NC 27513) to compare functioning across the four levels of initial stroke severity described earlier.

The data were also divided into two random subsets of items and the correlation between the person locations on each subset was estimated. The presence of internal consistency is indicated by the level of agreement between the two person ability estimates.[392]

2) Rasch framework [382]

a) Content is concerned with *relevance of* the items, based on expert opinions, *representativeness of* the items, examining the empirical hierarchy and spread of the item calibrations along the variable and *technical quality*, assessed via item fit statistics.

b) Substantive refers to *"theoretical rationales for the observed consistencies in test responses along with empirical evidence that the theoretical processes are actually engaged by respondents in the assessment tasks."* This type of validity may be addressed by verifying the definition of the variable intended by the researchers (confirmation of the intended item hierarchy) and examination of person fit statistics.

c) Structural is ascertained when the chosen measurement model's requirements, such as unidimensionality, are satisfied.

d) Generalizability is concerned with the degree to which inferences based on person measures or item calibrations are invariant, across different tasks, time, groups, or contexts.

d) External represents the correspondence between different measures of the same construct, and discriminant evidence, the lack of correspondence from measures of distinct constructs. It is the equivalent of the traditional contruct validity.

f) Interpretability is the degree to which qualitative meaning can be assigned to quantitative measure and corresponds to the extent to which the meaning of a score can be communicated and interpreted.

Reliability

When using RUMM, there are two reliability indices provided. The first is the Person Reliability Index and the second is Cronbach's Alpha. These two statistics are very similar to the ones seen in Classical Test Theory. They range between 0 and 1 and a higher value is considered to represent a higher level of reliability. A person separation index can also be calculated. The STRATA, computed by:

STRATA = (4G +1)/3)[148] indicates the number of distinct ability level separated by three errors of measurement. Rasch item reliability is determined in the same way and it represents how dispersed are the person measures along the construct being measured.[393]

Results

The baseline characteristics of the 4058 participants in each of the eight studies are presented in Table 7.4. Participants were 68 years of age on average and 54% were males; 77% had between one and four comorbid conditions and 60% had an ischemic stroke. The dominant hand was affected by the stroke in 42% of the sample.

Unidimensionality of the items

Unidimensionality was examined through factor analyses for each of the eight individual studies. Although the presence of one main factor was evident, several items loaded on a second factor, indicating that they may not be measuring upper extremity function. None of the items, however, were discarded prior to the Rasch analysis as the assumptions underlying the PCA, such as multicollinearity, were not met and the type of data (ordinal) and the samples sizes were, in some instances, inadequate. Kaiser-Meyer-Olkin (KMO) test statistics (KMO) were between 0.77 and 0.93 and the first factor explained between 59% and 92% of the item variance which is an indication that the set of items is measuring a single dimension.[394] Parallel analyses confirmed the presence of a single factor for each of the individual studies.

Calibration of the item pool

A first calibration of the items revealed a significant global fit statistic (chi-square) indicating that the items and/or the persons did not meet the expectation of the partial credit Rasch model.

Fit of the persons, fit of the items, the ordering of the item thresholds, as well as the presence of differential item functioning were all examined in the process of item reduction towards the creation of a true measure of upper extremity function. Because the goal was to create a short test that measured efficiently upper extremity function with few items, redundancies among the items was also examined. Persons and items having standardized residuals outside of the range (-2 to +2) and significant χ^2 and F-statistics were removed.

Ordered Thresholds

Upon the first calibration, several of the items presented with disordered thresholds. The categories were thus collapsed until the situation was rectified and all items presented ordered thresholds. An example of an item, the Box and Block Test that presented with disordered thresholds and how the collapsing of categories corrected the situation is presented in Figures 7.1a and b. After all items were examined and all their thresholds were ordered, all fit statistics, the standardized residuals, χ^2 and F-statistics, were re-examined.

Differential Item Functioning (DIF)

DIF was examined for each of the personal factors presented. One of the items, the Stroke Impact Scale item 7a (*In the past 2 weeks, how difficult was it to use your hand that was most affected by your stroke to carry heavy objects (e.g. bag of groceries?)* had different measurement characteristics for males and females. The presence of uniform DIF was detected. The item was split and two different difficulty estimates were produced for the two genders. None of the items demonstrated DIF for the evaluation time post-stroke.

Structure of the Measure

Of the 99 items that were regarded as suitable for inclusion in the upper extremity paper CAT, 64 were removed and 35 were retained. The 64 items deleted along with the reason for deletion are presented in Table 7.5. Most of the items did not fit the model, as shown by residuals outside of the acceptable range, significant χ^2 or significant F-statistics. The summary of the global fit statistics is presented in Table 7.6. The description of the 35 final items and their thresholds is given in Table 7.7 along their location, chi-square and F statistics.

Targeting

The targeting of the items to the subjects in the sample was good. The average person measure 1.149 (SD 1.747) which is slightly above zero (the mean item location by default). This means that the items only slightly too easy for the stroke survivors in this sample. The person-item threshold distribution map is presented in Figure 7.2. The persons are represented at the top of the graph while the items are at the bottom.

Selection of Starting Item for the adaptive measure

In order to create an adaptive measure and to avoid having to have the participant perform and answer each of the items retained in the bank, it was divided into two "testlets"; an easy "testlet" and a difficult "testlet". The starting item selected is part of the Chedoke McMaster Stroke Assessment Scale. It is an item assessing a gross hand movement: "*Pistol grip: pull trigger then return*". The subject must be able to close their hand (finger flexion), and extend and flex their index finger. This item was chosen because it is located at 0.103 logits and is of average difficulty. It is also quick and easy to administer. If the subject is unable to perform the pistol grip, the easy "testlet" is administered, starting with the item located at the midway point between the easiest and the most difficult item of this easy "testlet". Success or failure on this average item will determine if easier or more difficult items will be administered thereafter.

If the person is able to perform the pistol grip item as prescribed, then the therapist will proceed with the difficult "testlet", starting with the middle item in the difficult "testlet". Again, success or failure on this average item will determine if easier or more difficult items will be administered to complete the test (Table 7.8). Groups of items were formed that were two standard deviations above or below the starting item. These groups of items are demarcated by thick lines. The evaluator can choose to administer only one of the items in a group before skipping to an easier or harder group. Once the best group of items is located for subject, the evaluator may then want to administer all items in that particular group for a more precise score.

For ease of interpretation, item difficulty estimates were then transformed into percentages using the formula presented by Smith (Table 7.8).[395]

Unidimensionality – Post-Hoc Test

Unidimensionality was ascertained with Smith's independent t-test where person estimate from specific subsets of items are compared to person estimate derived for the complete set of items. The specific subsets chosen were items assessing *function*, items assessing *activity*, each subset of items that emerged from a specific test or index. The t-tests were found to be not significant, ascertaining the unidimensionality of the items.

Also carried out was a confirmatory factor analysis (MPlus software [396]) using weighted least square methods for categorical data, where possible, for each study. Only those studies whose measurement approach included enough retained items could be factor analyzed.

In the Brain capacity study, items from the Chedoke McMaster Stroke Assessment, the STroke REhabilitation Assessment of Movement, the Barthel Index as well as the Stroke Impact Scale were retained. Almost all factor loadings were near 1 on the main factor. The Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI) were 0.99 (values greater than .0.96 indicate good model fit[397]) and the Root-Mean-square error of Approximation (RMSEA) was 0.160, (values less than or equal to 0.06 indicate good model fit[397]). Only 3% of the residual correlations were greater than 0.1 so only a few items were locally dependent (the items from the Chedoke McMaster Stroke Assessement: ***Hand from knee to forehead 5X in 5 seconds*** and ***Elbow at side 90° flexion: supination then pronation***). Although these items were locally dependent, because these items evaluate different types of movement, they were kept as part of the measure.

In the Walking Competency Study, items remaining in the bank were pulled from the Test Évaluant la Performance des Membres Supérieurs des Personnes Agées (TEMPA), the Barthel Index, the OARS-IADL and the STREAM. Results of the CFA for the Walking competency study revealed CFI, TLI and RMSEA values of 0.99,0.98 and 0.12

respectively and only 4% of the residual correlations were greater than 0.1. Although RMSEA values are high, the overall fit statistics were considered acceptable to confirm unidimensionality given that the small sample sizes may have affected the fit estimate.[397] The Barthel Index item #3 (bathing) is correlated with other items (from the STREAM) but because it is the only activity item dealing with this particular activity, it was kept. As well, items from the TEMPA (#6 and #8) are correlated. However because one of the items assesses bilateral gross motor function and the other item assesses fine motor dexterity, two important constructs, both were kept as part of the final measure.

Psychometric characteristics of the measure

Content Validity

An important number of items included in the measure were from tests and indices that were initially developed to assess upper extremity function (the Box and Blocks Test, the Test Évaluant la Performance des Membres Supérieurs des Personnes Agées, the Chedoke McMaster Stroke Assessment, the Stroke Impact Scale (upper extremity items), the Stroke REhabilitation Assessment of Movement (upper extremity items)) so we were confident that the items well represented the construct being measured. In addition, the items comprised in the new adaptive measure of upper extremity function contain items representing the *body function* (15 items) and *activity and participation* (20 items) domains of the ICF. A number of items assess the upper extremity at the level of the shoulder, the level of the elbow and the level of the hand. Five items are bilateral tasks, involving all joints and two items assess how the person feels they are able to accomplish certain activities. The items cover a wide range of difficulty levels (Figures 7.2 and 7.3). In the Rasch Framework, content validity was also confirmed. The placement of the items hierarchically along the upper extremity continuum made intuitive sense. The items from the Chedoke McMaster Stroke Assessment were placed in the exact order of the original test which was based on Brunnstrom's theory on how the upper extremity recovers. All the persons in the final sample had fit residuals between -2.0 and +2.0. All the measurement model's requirements are met, including unidimensionality. The global fit statistics and all item and person fit statistics confirm this type of validity

Constuct Validity

As expected, the correlation between the total scores on the new measure of upper extremity function and the index of global functional recovery (Total score on the STREAM) are higher ($r = 0.6$; $p < .0001$) than those between the mental and emotional subcores of the SF-36 ($r = 0.2$ $p < .0001$). This is a confirmation of convergent and divergent validity. Also, internal consistency was further supported by a Cronbach's alpha of 0.97 for the persons' location on each of the two random subsets of items.

The new measure was able to differentiate between two different levels of stroke severity, mild and moderate-severe. Persons classified as having had moderate and severe strokes had scores that were similar; 40.1 (95%CI: 38.7 to 41.4) and 42.8 (95% CI: 40.9 to 41.5) respectively. Persons classified as having had a mild stroke scored an average of 50.4 (95% CI: 49.5 to 51.4) which was statistically different from the other classification groups.

Good correlation with the total score on the STroke REhabilitation Assessment of Movement ($r = 0.6$) and poor correlation with the mental and emotional subscores of the SF-36 ($r = 0.2$) also confirmed construct validity.

Generalizability

Differential item functioning was examined across a variety of different person factors and after splitting one of the Stroke Impact Scale items, none of them demonstrated any DIF. The final sample being much smaller than the original sample, it was warranted to examine the characteristics of the persons that make up this final sample on which the final person and item estimates are based (see Table 7.8). They do not differ significantly from the original sample. There are persons in each of the age groups, 57% of them are males, the distribution of types of stroke is similar to the original sample and to a typical stroke population. However, no persons evaluated at 12 months post-stroke remained in the final sample.

Reliability

The person separation index of 2.2 indicates the subjects separated into 2 distinct strata. The reliability of the hierarchy of person ability and item difficulty was good at 0.83. The item reliability index is 4.4, indicatiting a good dispersion, the measure distinguishes between 4 distinct difficulty levels of the items separated by 3 errors of measurement.

Ceiling and Floor Effects

The new measure demonstrates no floor or ceiling effect as no persons' ability levels are located below the easiest item or above the hardest item. The Test Information Function is presented in Figure 7.3. This graph provides an indication of the precision of the measure or the Standard Error of the measure. SEM= $(TIF)^{1/2}$. It indicates the amount of information the measure provides about the ability level at each maximum likelihood estimate[398] and can be used to compare the amount of information in different measures or subsets of items within a measure. Our measure of upper extremity measure is most useful around a logit of -1, so for persons close to average ability. The curves start tapering off around -7 and +7, beyond which point the persons ability levels cannot be measured as precisely. The range is still reasonably wide where information is available.

Discussion

An adaptive test of upper extremity function after stroke was created from a bank of previously calibrated. The adaptive test, in a paper version, contains 35 items, divided into 2 testlets; an easy and a hard one. The items all fit the Rasch partial credit model, tailored to the particular dataset at hand. The item thresholds spanned a wide range of difficulty from -7.377 to 6.167 logits and target well the sample with persons located close to 0 at 1.149 logits.

The hardest item was the Box and Block Test and, to be able to pass this item, the subject must be able to move at least 69 blocks from one side of the box to the other. This corresponds to a normal score for both men and women;[50] by scoring 100% on the adaptive test a person is considered to have fully recovered.

The easiest items are from the Test Évaluant la Performance des Membres Supérieurs des Personnes Agées. They are bilateral tasks. Although at first glance the position of these items at the bottom of the hierarchy may seem rather counter intuitive, it can be explained by the fact that the tasks are performed in part by the unaffected upper extremity and the affected upper extremity serves as support. For example, to hold a container while the unaffected arm unscrews the lid. If a person is unable to use their affected arm for support, it is an indication that there is no movement possible likely due to a high level of spasticity or flaccidity. To improve the evaluation process and the scoring, the instructions to the evaluator should specify that the affected arm should be used. As expected, the items originating from the Chedoke McMaster Stroke Assessment are located exactly in the order that they were intended.

For person and item calibration, the final sample included 942 subjects. In addition person ability estimates are available for 1014 persons with extreme scores (scored perfectly on all items or unable to do any item)

An important number of subjects have been deleted due to the nature of the data set. Because some of the studies included only a few items, a good number of which have been removed in the Rasch analysis process, they became either invalid (if persons did not respond to any of the final items in the measure). The characteristics of the final sample are similar in all aspects to the original sample except for the time of evaluation post-stroke. In the final sample no evaluation at 12 months post-stroke was retained and, thus, the paper adaptive test of upper extremity function is not generalizable to patients that are 1 year post stroke onset and should only really be used up to 7 months post-stroke.

Future work

Future work with this new measure involves extensive psychometric evaluation and concomitant re-calibration of the items when administered as an adaptive test to subject after stroke. Feasibility, efficiency, and validity under adaptive conditions need to be addressed. Responsiveness needs to be assessed, as well as all types of validity. For example, in the framework of Rasch analysis, discriminant validity is assessed by administering the new measure to a different group of subjects that may have impairments in their upper extremities but whose pattern of recovery is not the same as that of patients with stroke (e.g. patients with multiple sclerosis). If the ordering of the items is different from that obtained with the stroke population, it is evidence for discriminant validity.

It would be beneficial to add new items to fill the existing gaps (Figure 7.3). New items representing the participation domain of the ICF are needed. The existing items from the Reintegration to Normal Living index did not fit the model. This may be because there was not mention of the upper extremity in the question or statement and patients do not necessarily relate their limitations to their upper extremity but rather to their limited mobility, for example. I believe new items such as *"Are you limited in the type or number of social activities that you participate to **because of your affected upper extremity?**"* should be written in collaboration with health care professionals and stroke survivors and these items should be added to the bank through an equating process.

There is no doubt that the new measure as it stands will require thorough psychometric evaluation before it can be used by clinicians and researchers but they form a good basis. The fact that the items have been calibrated through Rasch analysis makes it an ideal setting to alter, delete, or add new items as the measure is being tested.

Table 7.1 Schedule of Evaluation Times Post-Stroke for Each of the Eight Studies.

Study \ Time	Stroke Onset	1 month	3 months	6 months	7 months	12 months
Walking Competency				■	■	
Brain Capacity	■		■			
Recovery from Stroke	■	■	■			
Bridging the Gap	■	■		■		
No Place Like Home	■	■	■	■		
Long Term Outcome of Stroke	■			■		■
Quality of Life	■	■	■	■		■
Canadian Stroke Registry				■		

Table 7.2. Item Bank of Upper Extremity Function Items.

ICF Component	Tests and Indices	Original Scoring and Categorization
Function	**Box and Block Test (BBT)**	**Number of blocks** 0=0-29; 1=29-49; 2=49-59; 3=59-69; 4=69-79 5=79-100
Function	**Chedoke McMaster Stroke Assessment (CMSA)** **ARM**	**0** Unable **1** Able
	Extension synergy, then flexion synergy	
	Shoulder flexion to 90°	
	Elbow at side, 90° flexion: supination then pronation	
	Shoulder flexion to 90°: pronation then supination	
	Hand from knee to forehead 5X in 5 sec.	
	Arm resting at side of body: raise arm overhead with full supination	
	Clap hand overhead, then behind back 3X in 5 sec.	
	Elbow at side, 90° flexion: resisted shoulder external rotation	
	HAND	
	Thumb extension >1/2 range, then lateral prehension	
	Finger flexion, the extension	
	Pronation: finger abduction	
	Hand unsupported: opposition of thumb to little finger	
	Pronation: tap index finger 10X in 5 sec	
	Pistol grip: pull trigger, then return	
	Thumb to finger tips, then reverse 3X in 12 sec.	
	Bounce a ball 4 times in succession, then catch	
	Pour 250 mL from 1 L pitcher, then reverse	
Function	**Grip Strength**	**Kilograms of Force:** **0**=0-27; **1**=27-40; **2**=40-50; **3** >50
Function	**STroke REhabilitation Assessment of Movement (STREAM)** 1 Protract Scapula 2 Extend elbow 7 Shrug shoulders 8 Raise hand to top of head 9 Hand on sacrum 12 Close hand 13 Open hand 14 Thumb to index finger	**0** unable **1** abnormal **2** normal

Table 7.2. Item Bank of Upper Extremity Function Items. (cont.).

ICF Component	Tests and Indices	Original Scoring and categorization
Activity	**Barthel Index** Feeding Bathing Dressing and Undressing	2 Independent 1 Assistance 0 Unable
Activity	**Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36)** Pick up/lift Grocery bags Take bath or dress	1 very limited 2 limited a little 3 not limited at all
Activity	**Older Americans Resources and Services Scale-Instrumental Activities of Daily Living (OARS-IADL)** Prepare meals Housework	2 Without help 1 With help 0 Unable
Activity	**SIS** Cut food Dress upper body Cut toe nails Housework (light) Carry heavy objects Turn doorknob Pick up money	1 Unable 2 Very difficult 3 Moderate 4 A little difficult 5 Not difficult
Activity	**Test d'évaluation des Membres Supérieurs chez les Personnes Âgées (TEMPA)** Pick up and move a jar Open jar and remove spoonful of coffee Pick up pitcher and pour water in glass Unlock a lock and open pill container Write on envelope and stick stamp Tie scarf around neck Shuffle and deal playing cards Handle coins Pick up and move small objects	3 normal 2 hesitation 1 difficulty-more than 25% 0 unable less than 25%

Table 7.3 Categorization of Influencing (Personal) Factors.

Factors	Categorization
Age	0 = 0 to 69 years of age 1= 70 to 79 years of age 2 = over 80 years of age
Gender	0 = Male 1= Female
Type of Stroke	0 = Ischemic 1 = Hemorrhagic 2 = Not defined or missing
Number of Comorbid Conditions	0 = No comorbid conditions 1 = 1 or 2 comorbid conditions 2 = More than 3 comorbid conditions
Side of Hemiplegia	0 = Left 1 = Right 2 = Bilateral 3 = Missing 4 = None noted
Time of Assessment Post-Stroke	0 = Stoke Onset 1 = 1month post-stroke 2 = 3 months post-stroke 3 = 6 months post-stroke 4 = 7 months post-stroke 5 = 12 months post-stroke
Dominant Hand Affected	**0 = Yes** **1 = No** **2 = Missing**
Stroke severity*	**1 = < 5** **2 =5 - 9.5** **3 = 9.5 - 11** **4 = > 11**

*Severity was based on the patient's score on the Canadian Neurological Scale (CNS) at the onset of stroke. The CNS, [387,399] was categorized into four groups: mild if the score was greater than 6.5; mild-moderate with a score between 5 and 6; moderate with a score between 4 and 4.5 and severe with if the score was below 3.5.[53,387]

Table 7.4. Demographic and Clinical Characteristics of Study Participants (n = 4058).

Subject Characteristic	
Age Mean (SD)	68 (15)
Gender No. (%) male	2206 (54)
Number of Comorbid conditions No. (%)	
0	803 (20)
1-2	1470 (37)
3-4	1556 (40)
>4	110 (3)
Type of CVA No. (%)	
Ischemic	2328 (59)
Hemorrhagic	415 (11)
Missing	116 (2)
Not Noted/Determined	1090 (28)
Side of Hemiplegia No. (%)	
Right	1475 (37)
Left	1467 (37)
Bilateral	234 (6)
Missing	447 (11)
Not Noted	316 (8)
Dominant UE affected No. (%)	1375 (42)
Stroke severity (%)	
Severe/Moderate/Mild-Moderate/Mild	29/6/7/58

*Severity was based on the patient's score on the Canadian Neurological Scale (CNS) at the onset of stroke. The CNS, [387,399] was categorized into four groups: mild if the score was greater than 6.5; mild-moderate with a score between 5 and 6; moderate with a score between 4 and 4.5 and severe with if the score was below 3.5.[53,387]

Figure 7.1a Example of Disordered Thresholds – The Box and Block Test.

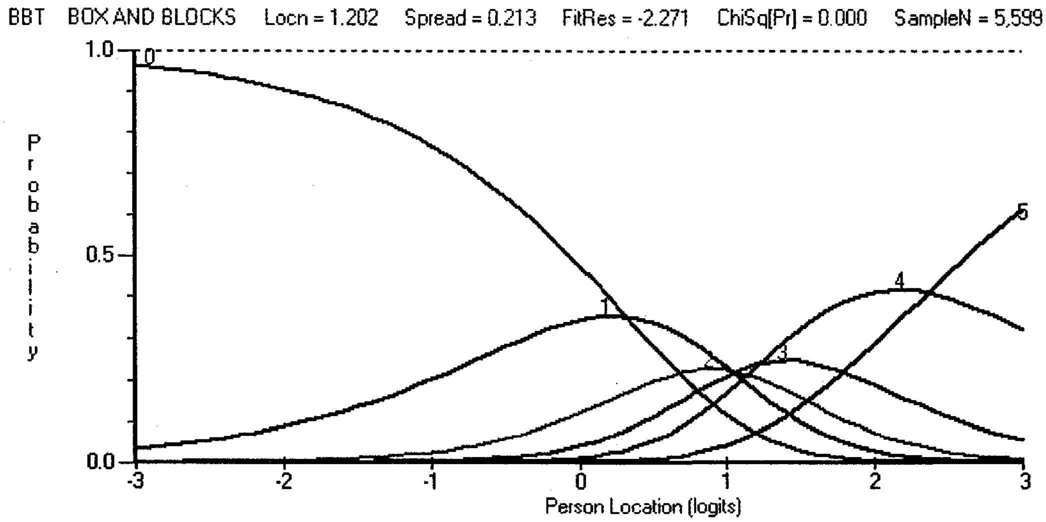


Figure 7.1b Example of Ordered Thresholds – The Box and Block Test After Collapsing of the Categories.

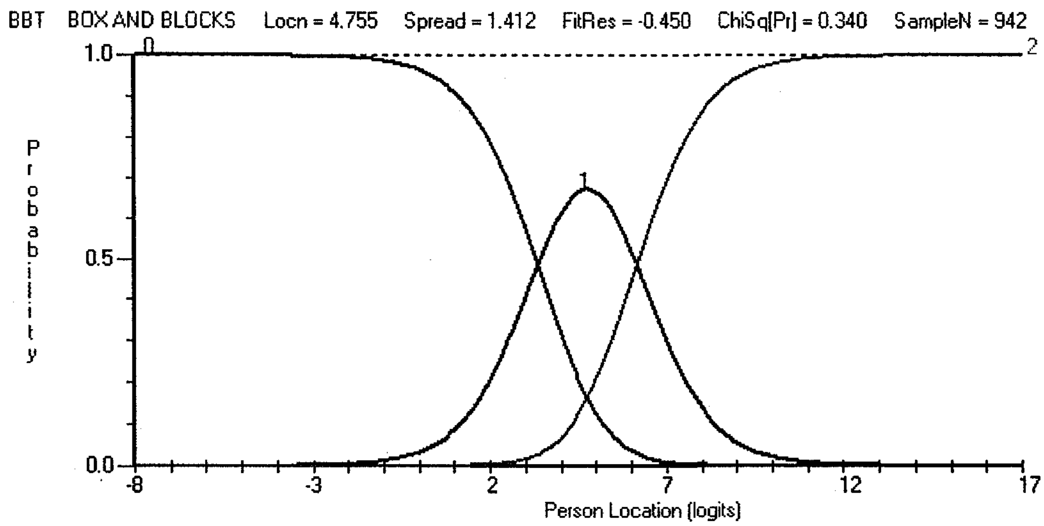


Table 7.5 Deleted Items and Reason for Deletion.

Reason for Deletion*	Deleted Items[§]
Lack of fit to the model	STR8, 12, 13 T3F SF3C, SF3J SIS5B SIS7E SIS5D SIS5C C3_4a,b C3_5c C4_4b C4_5b C4_7c Grip strength
Redundancy	SF3J STR8,12,13
Differential Item Functioning (person factors)	SF3J (observation) T3F (limb)

*Items may be deleted for more than 1 reason

[§]Please refer to Table 6.2 for meaning of abbreviated item names

Table 7.6 Summary of Global Fit Statistics for the Item Bank.

	ITEMS		PERSONS	
	Location	Fit Residual	Location	Fit Residual
Mean	0.000	-0.373	1.149	-0.275
Standard Deviation	2.109	0.575	1.747	0.724
Skewness		0.488		0.581
Kurtosis		-0.689		-0.153
Correlation		0.000		-0.041
ITEM-TRAIT INTERACTION			**RELIABILITY INDICES**	
Total Item Chi-Square	331.136		**Separation Index**	0.83191
Total Degrees of Freedom	309.000		**Cronbach Alpha**	N/A
Total Chi-Square Probability	0.184910			
POWER OF TEST-OF-FIT			**Power is GOOD** Based on Separation Index of 0.83191	

Table 7.7 The 35 Items and Their Measurement Properties by Location Order.

Description/Item Set	Location	SE	Fit Residual	Chi Square*	F-stat*
Open a jar and remove a spoonful of coffee (bilateral task)/TEMPA	-3.437	0.356	-0.352	8.75	1.305
Unlock a lock and open a pill container (bilateral task)/TEMPA	-3.421	0.355	-0.085	6.001	0.709
Tie a scarf around one's neck (bilateral task)/TEMPA	-3.313	0.359	0.176	10.739	1.176
Feeding/Barthel Index	-2.835	0.108	-0.252	4.895	0.614
Write on an envelope and stick a stamp on it (bilateral task)/TEMPA	-1.997	0.323	0.122	8.942	1.045
Elbow at side 90° flexion: supination then pronation/CMSA	-1.824	0.278	-0.688	3.197	0.929
Dressing and Undressing/Barthel Index	-1.818	0.096	0.722	23.145	1.714
Finger flexion then extension/CMSA	-1.737	0.274	0.171	12.364	1.293
Extends elbow in supine (starting with elbow fully flexed)/STREAM	-1.507	0.169	-1.132	11.501	2.742
Protract scapula in supine/STREAM	-1.216	0.163	-0.978	13.931	2.789
Can you prepare your own meals?/OARS-IADL	-1.165	0.186	0.123	9	1.029
Hand unsupported: opposition of thumb to little finger/CMSA	-0.898	0.247	-0.741	6.108	1.309
Handle coins (unilateral tasks)/TEMPA	-0.852	0.367	-0.784	7.551	2.23
Place hand on sacrum/STREAM	-0.712	0.155	-1.199	10.971	2.048
Shuffle and deal playing cards (bilateral task)/TEMPA	-0.652	0.325	-0.983	5.821	0.853
Shrugs shoulders (scapular elevation)/STREAM	-0.615	0.154	-0.292	5.23	0.716
Can you do your own housework?/OARS-IADL	-0.297	0.173	0.619	11.917	1.05
Pick up and move small objects (unilateral tasks)/TEMPA	-0.229	0.356	-0.57	11.289	2.547
In the past 2 weeks, how difficult was it to cut your food with a knife and fork?/SIS	-0.015	0.12	-1.18	17.556	2.55
Pistol grip: pull trigger then return/CMSA	0.103	0.225	-1.135	6.268	1.413
In the past 2 weeks, how difficult was it to use your hand that was most affected by your stroke to turn a doorknob?/SIS	0.212	0.117	0.667	9.09	1.035

Table 7.7 The 35 Items and Their Measurement Properties by Location Order. (cont.).

Description/Item Set	Location	SE	Fit Residual	Chi Square*	F-stat*
Pick up and move a jar (unilateral task)/TEMPA	0.301	0.356	-0.703	2.77	0.432
In the past 2 weeks, how difficult was it to do light household tasks/chores (e.g. dust, make a bed, take out garbage, do the dishes)?/SIS	0.449	0.108	0.91	11.105	1.184
Bathing/Barthel	0.729	0.119	-0.133	18.965	2.388
Pronation: tap index finger 10X in 5 sec./CMSA	0.87	0.219	-0.726	5.318	0.88
Arm resting at side of body: raise arm overhead with full supination/CMSA	0.913	0.219	-1.029	9.828	1.973
Pronation: tap index finger 10X in 5 sec/CMSA.	1.089	0.219	-0.674	4.199	0.725
In the past 2 weeks, how difficult was it to use your hand that was most affected by your stroke to carry heavy objects (e.g. bag of groceries)?/ SIS(MALES)	1.138	0.145	0.179	11.625	1.512
SIS (FEMALES)	2.114	0.162	-0.428	7.561	1.014
In the past 2 weeks, how difficult was it to clip your toenails?/SIS	2.167	0.118	-0.172	13.401	1.322
Elbow at side, 90° flexion: resisted shoulder external rotation/CMSA	2.603	0.234	-0.793	7.68	1.782
Thumb to finger tips, then reverse 3X in 12 sec/CMSA	2.793	0.238	-0.229	9.085	1.313
Clap hands overhead then behind back 3X in 5 sec./CMSA	3.378	0.252	-0.695	7.156	2.046
Number of blocks transferred in 60 seconds/BBT	4.755	0.183	-0.45	10.128	1.51
Bounce ball 4 times in succession then catch/CMSA	4.925	0.323	-0.345	8.049	2.486

BBT; Box and Block Test, CMSA; Chedoke McMaster Stroke Assessment, OARS_IADL; Older American Resources and Services Scale-Instrumental Activities of Daily Living, SE; Standard Error, SIS; Stroke Impact Scale, STREAM; STroke REhabilitation Assessment of Movement, TEMPA; Test Évaluant la Performance des Membres Supérieurs des Personnes Agées

* All chi-square and F-Statistics were statistically non-significant after Bonferroni correction.

Figure 7.2 Person-Item Threshold Distribution for the Final 35-Item Measure

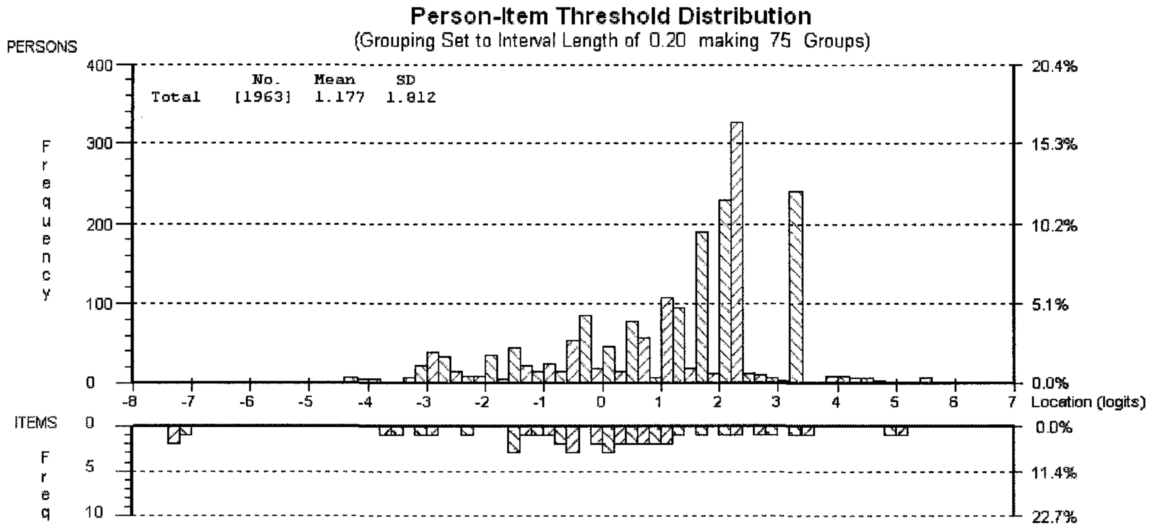


Table 7.8 Final 35-Item Adaptive Measure of Upper Extremity Function Post-Stroke

Starting Item: Pistol grip: pull trigger then return

Unable Easy testlet Start with #7		**Able DIFFICULT testlet Start with #36**	
Able Move down until patient is unable to meet the criteria for the specific task	**Unable Move up until patient is able to meet the criteria for the specific task**	**Able Move down until patient is unable to meet the criteria for the specific task**	**Unable Move up until patient is able to meet the criteria for the specific task**

EASY Testlet Items	**Score /100**
*1-**Tie a scarf around one's neck (bilateral task)**. The task is partially executed (more than 25%) or certain steps are executed with major difficulties necessitating repeated efforts. Part of the task may have had to be modified or needed assistance to make it achievable.*	3
*2-**Open a jar and remove a spoonful of coffee (bilateral task)** The task is partially executed (more than 25%) or certain steps are executed with major difficulties necessitating repeated efforts. Part of the task may have had to be modified or needed assistance to make it achievable.*	4
*3-**Unlock a lock and open a pill container (bilateral task)** The task is partially executed (more than 25%) or certain steps are executed with major difficulties necessitating repeated efforts. Part of the task may have had to be modified or needed assistance to make it achievable.*	5
*4-**Feeding independently**. The patient **needs some assistance** to feed himself a meal from a tray or table when someone places the food within his reach. He needs assistance to put on an assistive device if required, cut up his food, use salt and pepper, spread butter, etc. He needs assistance to be able to accomplish this in a reasonable time.*	23
*5-**Write on an envelope and stick a stamp on it (bilateral task)** The task is partially executed (more than 25%) or certain steps are executed with major difficulties necessitating repeated efforts. Part of the task may have had to be modified or needed assistance to make it achievable*	29

*6-**Dressing and Undressing** Patient needs some assistance: to put on, remove and fasten all clothing and tie shoe laces (unless it is necessary to used adaptive aids for this). This includes putting on, removing and fastening corsets or braces when they are prescribed.*	33
7-Shuffle and deal playing cards (bilateral task) The task is partially executed (more than 25%) or certain steps are executed with major difficulties necessitating repeated efforts. Part of the task may have had to be modified or needed assistance to make it achievable***	34
*8-**Elbow at side 90^0 flexion: supination then pronation***	44
*9-**Finger flexion then extension***	45
*10-**Extends elbow** **in** **supine** (starting with elbow fully flexed) Able to complete the movement in a manner that is comparable to the unaffected side.*	46
*11-**Protract scapula in supine** Able to complete the movement in a manner that is comparable to the unaffected side.*	48
*12-**Can you prepare your own meals? Cook meals yourself***	49
*13-**Feeding independently:** The patient can feed himself a meal from a tray or table when someone places the food within his reach. He is able to put on an assistive device if required, cut up his food, use salt and pepper, spread butter, etc. He must be able to accomplish this in a reasonable time*	51
*14-**Hand unsupported: opposition of thumb to little finger***	51
15-Handle coins (unilateral task) The task is partially executed (more than 25%) or certain steps are executed with major difficulties necessitating repeated efforts. Part of the task may have had to be modified or needed assistance to make it achievable.	52
*16-**Place hand on sacrum** Able to complete the movement in a manner that is comparable to the unaffected side.*	53
*17-**Shrugs shoulders (scapular elevation)** Able to complete the movement in a manner that is comparable to the unaffected side.*	55
*18-**Can you do your own housework?** Without help*	56
*19-**Dressing and Undressing Independently**. Patient is able to put on, remove and fasten all clothing and tie shoe laces (unless it is necessary to used adaptive aids for this). This includes putting on, removing and fastening corsets or braces when they are prescribed.*	56
*20-**Pick up and move small objects (unilateral task)** The task is partially executed (more than 25%) or certain steps are executed with major difficulties necessitating repeated efforts. Part of the task may have had to be modified or needed assistance to make it achievable.*	57
*21-**Write on an envelope and stick a stamp on it (bilateral task)** The task is partially executed (more than 25%) or certain steps are executed with major difficulties necessitating repeated efforts. Part of the task may have had to be modified or needed assistance to make it achievable.*	57

DIFFICULT "testlet" Items	**Score /100**
*22-**In the past 2 weeks, how difficult was it to cut your food with a knife and fork?***	58
*23-**In the past 2 weeks, how difficult was it to use your hand that was most affected by your stroke to turn a doorknob?***	59
*24-**Pick up and move a jar** (unilateral task) The task is successfully completed without hesitation or difficulty, as instructed or demonstrated.*	59
*25-**Unlock a lock and open a pill container** (bilateral task) The task is successfully completed without hesitation or difficulty, as instructed or demonstrated.*	60
*26-**In the past 2 weeks, how difficult was it to do light household tasks/chores (e.g. dust, make a bed, take out garbage, do the dishes)?** Just a little or not difficult at all.*	61
*27-**Bathing Independently** The patient must be able to use a bathtub, a shower or take a complete sponge bath. He must be able to perform all the steps involved in any one of these tasks without another person being present.*	**63**
*28-**Tie a scarf around one's neck** (bilateral task) The task is successfully completed without hesitation or difficulty, as instructed or demonstrated.*	63
*29-**Hand from knee to forehead 5X in 5 seconds***	64
*30-**Arm resting at side of body: raise arm overhead with full supination***	**64**
*31-**Pronation: tap index finger 10X in 5 seconds***	65
*32-**In the past 2 weeks, how difficult was it to use your hand that was most affected by your stroke to carry heavy objects (e.g. bag of groceries)?(Men)** Just a little or not difficult at all.*	66
33-Open a jar and remove a spoonful of coffee (bilateral task) The task is successfully completed without hesitation or difficulty, as instructed or demonstrated.	71
34-In the past 2 weeks, how difficult was it to clip your toenails? Just a little or not difficult at all.	**73**
*35-**In the past 2 weeks, how difficult was it to use your hand that was most affected by your stroke to carry heavy objects (e.g. bag of groceries)?(Women)** Just a little or not difficult at all.*	73
§*36-**Elbow at side, 90^0 flexion: resisted shoulder external rotation***	**76**
*37-**Thumb to finger tips, then reverse 3X in 12 sec***	78
*38-**Number of blocks transferred in 60 seconds >30***	82
*39-**Clap hands overhead then behind back 3X in 5 sec.***	82
*40-**Bounce ball 4 times in succession then catch***	93
*41-**Number of blocks transferred in 60 seconds >60***	100

*Starting item for EASY "testlet"

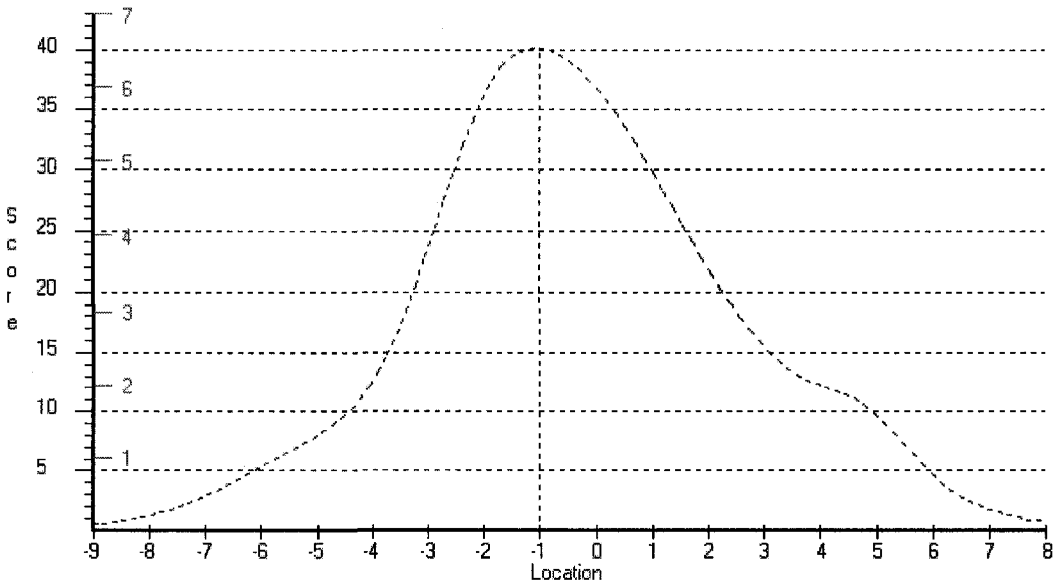
§Starting Item for DIFFICULT "testlet"

Table 7.9 Characteristics of the persons in the original and final samples

Subject Characteristic (N= 843)	Final Sample (N=942)	Full Sample (N = 6239)	Chi-Square *p-value*
Age (%) 0 to 69/70 to 79 /over 80 /missing	41/26/23/10	40/24/20/17	1.99 0.57
Gender No. (%) male	54	57	0.081 *0.78*
Type of CVA (%) Ischemic/Hemorrhagic/Other, not define or missing	85/7/7	88/6/6	0.20 *0.90*
Side of Hemiplegia (%) Right / Left /Bilateral / None or none noted	40/37/8/15	39/36/10/15	0.25 *0.97*
Dominant Hand affected (%)	63	66	0.07 *0.79*
Stroke severity at onset of stroke	10/21/5/7/44	7/23/7/8/50	1.05 *0.90*
Time post-stroke Onset-1 month/3months/6months/7months/12 months	27/16/18/37/12/0	20/16/12/36/1/15	26.15 *0.0001**

* There were no persons in the 12 months category in the final measure.

Figure 7.3 Test Information Function for the Paper Adaptive Test of Upper Extremity Function.



CHAPTER 8 Summary and Conclusions

Upper extremity recovery after a stroke is a complex phenomenon. Researchers worldwide are actively searching for ways to treat upper extremity deficits in order for persons who have suffered a stroke to return to a level of function that will allow them to pursue their activities, fulfill their life roles and attain an ideal quality of life. Several new and important leads have arisen as far as rehabilitative approaches, recently such as Constraint Induced Movement Therapy. Although this type of treatment seems to be making important headway and be very promising, more quality randomized controlled trials need to be performed in different sub-populations to be able to generalize its applicability.

When determining the effectiveness of a particular treatment or program of intervention, choosing an appropriate outcome assessment is as important as it is difficult to do. Deciding which test or index to use involves a multitude of decisions. Both researchers and clinicians are faced with difficult choices and assessing function of the upper extremity is not exception. Ideally, a test or index should be psychometrically sound and should cover the whole spectrum of upper extremity function from the most basic movements to the most complex but should also be able to capture the activities and the participation in life roles associated with a level of upper extremity function.

As exemplified by the first manuscript of this thesis, a study aimed at evaluating the effectiveness of a task-oriented intervention on upper extremity function, existing tests and indices tend to be very narrowly focused, assessing only a particular level of function. Traditionally, the strategy employed by researchers and rehabilitation professionals has been to administer an entire arsenal of tests and indices, making the processes very tiring and bothersome for both participants and evaluator alike. Another inconvenience is making sense of all these tests and indices that are scored differently, and whose total score does not necessarily have an inherent meaning. Moreover, the ordinal scales on which most of the existing scales are based are routinely treated as if they were interval making the quantification of change misleading.

The second chapter introduced new psychometric models that have been used extensively in the fields of education and psychology. These 'modern' models have the potential to overcome the difficulties encountered with the classical approach of measurement and the development of measures. Rasch analysis and IRT, by transforming ordinal scores into linear measures make it possible to objectively assess latent construct, such as how the patient feels it is difficult for them to accomplish a certain activities or how satisfied they are with certain aspects of their lives. Rasch analysis and Item Response Theory have made their debut into the world of rehabilitation. The first report of their use date to the 1990's. Since then, there has been a steady increase in the use of these modern psychometric models in the development of new measures, and in the assessment of psychometric properties of these measures. The use of measures that have been developed through these models, however, is lagging. This may be due to the fact researchers and clinicians are not always aware of these new trends and the new measures that are being created. The publication of this manuscript will hopefully inform rehabilitation professionals about these new measurement methods and also assist them in choosing an appropriate measure for the patients or research study. From the results of this manuscript, it is evident that very few measures were developed to assess upper extremity function in persons with stroke. None of them included items that represent each of the domains of the International Classification of Functioning, Disability and Health.

The third manuscript describes the development of a bank of items. The goal was to create a method to parsimoniously and hierarchically assess upper extremity function in persons with stroke. This bank contains a total of fifty calibrated items that can be used to assess clients with upper extremity impairments due to a stroke. The advantage and ultimate goal of having a bank of items is the possibility of administering a 'tailored' measure. The items are chosen to target the person's ability level and a true score can be obtained with the administration of only a few items. This decreases patient burden and frustration linked to administering items that are too easy or too hard. The items cover the domains of the ICF: *Body structures and function* and *Activity and participation*.

A new paper format of an adaptive test was created in manuscript 4. In order to assess whether the items performed differently across time, different evaluation periods were included. The new measure contains 35 items and can be administered in a few minutes.

When assessing patients with a measured developed through Rasch analysis, true change can be estimated. This is of utmost importance when evaluating the effectiveness of an intervention. Finding effective interventions is a challenge but until rehabilitation professionals are using true measures to assess their clients, results will not be conclusive.

Capturing the client's perception on the difficulties there are facing when accomplishing their tasks, or how able they are in participating in their usual activities is an indispensable part of healthcare provision. The use of Rasch models makes true measurement possible and is necessary for the interpretation of change. Ultimately, this measure will help determine if the interventions aimed at improving upper extremity function are effective and this will translate into optimal patient care.

A few years ago a new chapter in the evaluation of rehabilitation outcomes was opened. Since then, a number of clinicians and researchers have recognized the immense benefits of using new, contemporary psychometric approaches to develop true measures. New measures have been developed; existing ones have been further tested and evaluated in diverse populations such as stroke, traumatic brain injury and multiple sclerosis. Studies using previously Rasch-developed measures are still scarce and the livelihood of these psychometric models in the field of rehabilitation rests in the hands of the researchers that are developing and refining these measures.

Translating the fruits of research to other researchers as well as to clinicians is a challenge but is a necessary step to the widespread use of these techniques. The calibration of item banks and the psychometric refinement of measures is an ongoing process. As researchers use these measures as outcome evaluations as part of their research project, as clinicians use them to assess change in individuals patients or as part of program evaluations, items need to be harvested and calibrated on an on-going basis as patient populations change and new rehabilitative techniques are explored and tested. I believe the developers of item banks have the responsibility of maintaining the bank, adding new items and calibrating them and working in collaboration with other researchers to join forces for launching and maintaining large scale items banks. It is only by a collaborative process that these goals are attainable and the field of rehabilitation can be recognized to its full potential and rehabilitation professionals can deliver therapy based on evidence.

Future Directions

1. Maintain data bank of upper extremity items
2. Calibrate new items as data become available
3. Develop novel items and calibrate
4. Validate "paper-CAT" through classical psychometric evaluation
5. Use effective knowledge translation strategies to facilitate incorporation into research and clinical practice
6. Develop curricula at undergraduate and professional levels to teach modern psychometric approaches to assessment of patients

Major advancements in science are preceded by breakthroughs in measurement methods

(Nunnaly 1978).

REFERENCES

1. Fisher WP Jr., Harvey RF, Taylor P, Kilgore KM, Kelly CK. Rehabits: a common language of functional assessment. Arch Phys Med Rehabil 1995;76:113-122.

2. Rasch G. Introduction. Probabilistic models for some intelligence and attainment tests. Chicago: The University of Chicago Press, 1980:ix-xix.

3. Heart and Stroke Foundation of Canada. http://ww2.heartandstroke.ca/Page.asp?PageID=24.2007. 5-5-0007.

4. Duncan PW. Measuring Recovery of Function After Stroke: Clinical and Measurement Issues in Selecting Stroke Outcome Measures in Clinical Trials. In Goldstein LB, editor. Restorative Neurology: Advances in Pharmacotherapy for Recovery After Stroke. Armonk, NY: Futura Publishing Co., 1998:225-239.

5. Granger CV, Deutsch A, Linn RT. Rasch analysis of the Functional Independence Measure (FIM) Mastery Test. Arch Phys Med Rehabil 1998;79:52-57.

6. Parker VM, Wade DT, Langton Hewer R. Loss of arm function after stroke: measurement, frequency, and recovery. Int Rehabil Med 1986;8:69-73.

7. Mayo NE, Wood-Dauphinee S, Ahmed S, Gordon C, Higgins J, McEwen S, Salbach N. Disablement following stroke. Disabil Rehabil 1999;21:258-268.

8. McEwen S, Mayo N, Wood-Dauphinee S. Inferring quality of life from performance-based assessments. Disabil Rehabil 2000;22:456-463.

9. Wyller TB, Sveen U, Sodring KM, Pettersen AM, Bautz-Holter E. Subjective well-being one year after stroke. Clin Rehabil 1997;11:139-145.

10. Kelsey JL, Pastides H, Krieger N. Upper Extremity Disorders: A Survey of their Frequency and Cost in the United States. In C.V.Mosby Co., editor. St. Louis (MO): 1980.

11. Harvey RL. Motor recovery after stroke: new directions in scientific inquiry. Phys Med Rehabil Clin N Am 2003;14:S1-S5.

12. Ernst E. A review of stroke rehabilitation and physiotherapy. Stroke 1990;21:1081-1085.

13. Duncan PW. Synthesis of intervention trials to improve motor recovery following. stroke. Top Stroke Rehabil Winter 1997;3:1-20.

14. Gelber DA, Josefczyk PB, Herrman D, Good DC, Verhulst SJ. Comparison of two therapy approaches in the rehabilitation of the pure motor hemiparetic stroke patient. J Neurol Rehabil 1995; 9:191-196.

15. Wagenaar RC, Meijer OG, van Wieringen PC, Kuik DJ, Hazenberg GJ, Lindeboom J, Wichers F, Rijswijk H. The functional recovery of stroke: a comparison between neuro-developmental treatment and the Brunnstrom method. Scand J Rehabil Med 1990;22:1-8.

16. Kwakkel G, Kollen BJ, Wagenaar RC. Therapy impact on functional recovery in stroke rehabilitation: a critical review of the literature. Physiotherapy 1999;85:377-391.

17. Muir KW, Grosset DG, Lees KR. Interconversion of stroke scales. Implications for therapeutic trials. Stroke 1994;25:1366-1370.

18. Basmajian JV, Gowland CA, Finlayson MA, Hall AL, Swanson LR, Stratford PW, Trotter JE, Brandstater ME. Stroke treatment: comparison of integrated behavioral-physical therapy vs traditional physical therapy programs. Arch Phys Med Rehab 1987;68:267-272.

19. Kannel WB, Sorlie P. Some health benefits of physical activity. The Framingham Study. Arch Intern Med 1979;139:857-861.

20. Taub E, Crago JE, Uswatte G. Constraint-Induced Movement Therapy: A new approach to treatment in physical rehabilitation. Rehabil Psychol 1998;43:152-170.

21. Taub E, Miller NE, Novack TA, Cook EW, III, Fleming WC, Nepomuceno CS, Connell JS, Crago JE. Technique to improve chronic motor deficit after stroke. Arch Phys Med Rehabil 1993;74:347-354.

22. Kunkel A, Kopp B, Muller G, Villringer K, Villringer A, Taub E, Flor H. Constraint-induced movement therapy for motor recovery in chronic stroke patients. Arch Phys Med Rehabil 1999;80:624-628.

23. van der Lee JH, Wagenaar RC, Lankhorst GJ, Vogelaar TW, Deville WL, Bouter LM. Forced use of the upper extremity in chronic stroke patients: results from a single-blind randomized clinical trial. Stroke 1999;30:2369-2375.

24. Miltner WH, Bauder H, Sommer M, Dettmers C, Taub E. Effects of constraint-induced movement therapy on patients with chronic motor deficits after stroke: a replication. Stroke 1999;30:586-592.

25. Dromerick AW, Edwards DF, Hahn M. Does the application of constraint-induced movement therapy during acute rehabilitation reduce arm impairment after ischemic stroke? Stroke 2000;31:2984-2988.

26. Wolf SL, Winstein CJ, Miller JP, Taub E, Uswatte G, Morris D, Giuliani C, Light KE, Nichols-Larsen D. Effect of constraint-induced movement therapy on upper extremity function 3 to 9 months after stroke: the EXCITE randomized clinical trial. JAMA 2006;296:2095-2104.

27. Andresen EM. Criteria for assessing the tools of disability outcomes research. Arch Phys Med Rehabil 2000;81:S15-S20.

28. Sloan JA, Aaronson N, Cappelleri JC, Fairclough DL, Varricchio C. Assessing the clinical significance of single items relative to summated scores. Mayo Clin Proc 2002;77:479-487.

29. Lawton MP. The functional assessment of elderly people. J Am Geriatr Soc. 1971;19:465-481.

30. Bode RK, Lai JS, Cella D, Heinemann AW. Issues in the development of an item bank. Arch Phys Med Rehabil 2003;84:S52-S60.

31. Higgins J, Salbach NM, Wood-Dauphinee S, Richards CL, Cote R, Mayo NE. The effect of a task-oriented intervention on arm function in people with stroke: a randomized controlled trial. Clin Rehabil 2006;20:296-310.

32. Dean CM, Richards CL, Malouin F. Task-related circuit training improves performance of locomotor tasks in chronic stroke: a randomized, controlled pilot trial. Arch Phys Med Rehab 2000;81:409-417.

33. Richards L, Stoker-Yates J, Pohl P, Wallace D, Duncan P. Reliability and validity of two test of upper extremity motor function post-stroke. OTJR 2001;21:201-219.

34. Granger CV, Kelly-Hayes M, Johnston M, Deutsch A, Braun S, Fiedler R. Quality and 6utcome 0easures for 0edical rehabilitation. Rasch Measurement in Health Sciences. Maple Grove: JAM Press, 2000:82-113.

35. Rudman D, Hannah S. An instrument evaluation framework: description and application to assessments of hand function. J Hand Ther 1998;11:266-277.

36. Unsworth CA. The Concept of Function. Br J Occup Ther 1993;56:287-292.

37. World Health Organization. International Classification of Functioning, Disability and Health. Geneva: 2001.

38. Barreca S, Gowland CK, Stratford P, Huijbregts M, Griffiths J, Torresin W, Dunkley M, Miller P, Masters L. Development of the Chedoke Arm and Hand Activity Inventory: theoretical constructs, item generation, and selection. Top Stroke Rehabil 2004;11:31-42.

39. Brunnstrom S. Movement therapy in hemiplegia. New York: Harper and Row, 1970.

40. Baker AB. Clinical Neurology. New York: Hoeber-Harper, 1962.

41. Wechsler IS. Textbook of Clinical Neurology. Philadelphia: W.B. Saunders Co.1952.

42. Brunnstrum S. Movement therapy in Hemiplegia: A Neurophysiological approach. New York: Harper and Row, 1970.

43. Patrick DL, Kinne S, Engelberg RA, Pearlman RA. Functional status and perceived quality of life in adults with and without chronic conditions. J Clin Epidemiol 2000;53:779-785.

44. Shumaker SA, Naughton MJ. The international assessment of health-related quality of life: A theoretical perspective. In: Shumaker SA, Berzon RA, editors. The International Assessment of Health-Related Quality of Life. New York: Rapid Communications, 2000:3-10.

45. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. The EuroQol Group. Health Policy 1990;16:199-208.

46. Feeny D, Torrance G, Furlong W. Health Utilities Index. In Spilker B, editor. Quality of Life and Pharmacoeconmics in Clinical Trials. Philadelphia: Lippincott-Raven, 1996:239-252.

47. Torrance GW, Furlong W, Feeny D, Boyle M. Multi-Attribute Preference Functions - Health Utilities Index. Pharmacoeconomics 1995;7:503-520.

48. Ware Jr. JE, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): I. Conceptual framework and item selection. Med Care 1992;30:473-481.

49. Cromwell FS. Occupational Therapist's Manual for Basic Skill Assessment; Primary Prevocational Evaluation. Altadena: Fair Oaks Printing, 1976.

50. Desrosiers J, Bravo G, Hebert R, Dutil E, Mercier L. Validation of the Box and Block Test as a measure of dexterity of elderly people: reliability, validity, and norms studies. Arch Phys Med Rehab 1994;75:751-755.

51. Hebert R, Carrier R, Bilodeau A. The Functional Autonomy Measurement System (SMAF): description and validation of an instrument for the measurement of handicaps. Age Ageing 1988;17:293-302.

52. Mathiowetz V, Volland G, Kashman N, Weber K. Adult norms for the Box and Block Test of manual dexterity. Am J Occup Ther 1985;39:386-391.

53. Cote R, Battista RN, Wolfson C, Boucher J, Adam J, Hachinski V. The Canadian Neurological Scale: validation and reliability assessment. Neuro 1989;39:638-643.

54. Cote R, Battista RN, Wolfson CM, Hachinski V. Stroke assessment scales: guidelines for development, validation, and reliability assessment. Can J Neurol Sci 1988;15:261-265.

55. Gowland C, VanHullenaar S, Torresin W, Moreland J, Vanspall B, Barrecca S. Administering the Chedoke Assessment. Chedoke-McMaster Stroke Assessment. Chedoke-McMaster Hospital and McMaster University, 1995:6-1-6-7.

56. Gowland C, Stratford P, Ward M, Moreland J, Torresin W, Van Hullenaar S, Sanford J, Barreca S, Vanspall B, Plews N. Measuring physical impairment and disability with the Chedoke- McMaster Stroke Assessment. Stroke 1993;24:58-63.

57. Mathiowetz V, Weber K, Volland G, Kashman N. Reliability and validity of grip and pinch strength evaluations. J Hand Surg [Am] 1984;9:222-226.

58. Mathiowetz V, Kashman N, Volland G, Weber K, Dowe M, Rogers S. Grip and pinch strength: normative data for adults. Arch Phys Med Rehab 1985;66:69-74.

59. Kellor, M, Frost, J, Silberberg, N, Iversen, I, and Cummings, R. Hand strength and dexterity: norms for clinical use. Am J OccTher 1971;25:77-83.

60. Mathiowetz, V, Weber, K, Kashman, N, and Volland, G. Adult norms for the Nine Hole Peg test of finger dexterity. American Journal of Occupational Therapy 1985;39:25-38.

61. Sunderland A, Tinson DJ, Bradley EL, Fletcher D, Langton Hewer R, Wade DT. Enhanced physical therapy improves recovery of arm function after stroke. A randomised controlled trial. J Neurol Neurosurg Psychiatry 1992;55:530-535.

62. Rankin J. Cerebral vascular accidents in patients over the age of 60. II. Prognosis. Scott Med J 1957;2:200-215.

63. van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. Stroke 1988;19:604-607.

64. Daley K, Mayo N, Wood Dauphinee S, Danys I, Cabot R. Verification of the STroke REhabilitation Assessment of Movement (STREAM). Physiother Can 1997;49:269-278.

65. Visintin M, Barbeau H, Korner-Bitensky N, Mayo NE. A new approach to retrain gait in stroke patients through body weight support and treadmill stimulation. Stroke 1998;29:1122-1128.

66. De Souza LH, Hewer RL, Miller S. Assessment of recovery of arm control in hemiplegic stroke patients. Arm function tests. Int Rehabil Med 1980;2:3-9.

67. Heller A, Wade DT, Wood VA, Sunderland A, Hewer RL, Ward E. Arm function after stroke: measurement and recovery over the first three months. J Neurol Neurosurg Psychiatry 1987;50:714-719.

68. Desrosiers, J, Hebert, R, Dutil, E, and Bravo, G. Development and Reliability of an upper extremity function test for the elderly: The TEMPA. Can J Occup Ther 1993;60:9-16.

69. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. Int J Rehabil Res 1998;4:483-492.

70. Desrosiers J, Hebert R, Bravo G, Dutil E. Upper extremity performance test for the elderly (TEMPA): normative data and correlates with sensorimotor parameters. Test d'Evaluation des Membres Superieurs de Personnes Agees. Arch Phys Med Rehab 1995;76:1125-1129.

71. Platz T, Winter T, Muller N, Pinkowski C, Eickhof C, Mauritz KH. Arm ability training for stroke and traumatic brain injury patients with mild arm paresis: a single-blind, randomized, controlled trial. Arch Phys Med Rehabil 2001;82:961-968.

72. Mahoney FI, Barthel DW. Functional evaluation: The Barthel Index. Md St Med J 1965;14:61-65.

73. Roy CW, Togneri J, Hay E, Pentland B. An inter-rater reliability study of the Barthel Index. Int J Rehabil Res 1988;11:67-70.

74. Wolfe CDA, Taub NA, Woodrow EJ, Burney PGJ. Assessment of Scales of Disability and Handicap for Stroke Patients. Stroke 1991;22:1242-1244.

75. Fillenbaum GG, Smyer MA. The development, validity, and reliability of the OARS multidimensional functional assessment questionnaire. J Gerontol 1981;36:428-434.

76. Reuben DB, Valle LA, Hays RD, Siu AL. Measuring physical function in community-dwelling older persons: A comparison of self-administered, interviewer-administered, and performance-based measures. J Am Geriatr Soc 1995;43:17-23.

77. Poissant L, Mayo E, Wood-Dauphinee S, Clarke A. The Development and Preliminary Validation of a Preference-Based Stroke Index (PBSI). Health Qual Life Outcomes 2003;1:43.

78. Wood-Dauphinee S, Williams JI. Reintegration to Normal Living as a proxy to quality of life. J Chronic Dis 1987;40:491-502.

79. Wood-Dauphinee SL, Opzoomer MA, Williams JI, Marchand B, Spitzer WO. Assessment of global function: The Reintegration to Normal Living Index. Arch Phys Med Rehab 1988;69:583-590.

80. Steiner A, Raube K, Stuck AE, Aronow HU, Draper D, Rubenstein LZ, Beck JC. Measuring psychosocial aspects of well-being in older community residents: performance of four short scales. Gerontologist 1996;36:54-62.

81. McColl MA, Paterson M, Davies D, Doubt L, Law M. Validity and community utility of the Canadian Occupational Performance Measure. Can J Occup Ther 2000;67:22-30.

82. Dorman PJ, Waddell F, Slattery J, Dennis M, Sandercock P. Are Proxy Assessments of Health Status After Stroke with the EuroQol Questionnaire Feasible, Accurate, and Unbiased? Stroke 1997;28:1883-1887.

83. Dorman P, Slattery J, Farrell B, Dennis M, Sandercock P. Qualitative comparison of the reliability of health status assessments with the EuroQol and SF-36

questionnaires after stroke. United Kingdom Collaborators in the International Stroke Trial. Stroke 1998;29:63-68.

84. Dorman PJ, Dennis M, Sandercock P. How do scores on the EuroQol relate to scores on the SF-36 after stroke? Stroke 1999;30:2146-2151.

85. Dorman PJ, Waddell F, Slattery J, Dennis M, Sandercock P. Is the EuroQol a Valid Measure of Health-Related Quality of Life After Stroke? Stroke 1997;28:1876-1882.

86. Mathias SD, Bates MM, Pasta DJ, Cisternas MG, Feeny D, Patrick DL. Use of the Health Utilities Index With Stroke Patients and Their Caregivers. Stroke 1997;28:1888-1893.

87. Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. J Clin Epidemiol 1997;50:79-93.

88. Segal ME, Schall RR. Determining functional/health status and its relation to disability in stroke survivors. Stroke 1994;25:2391-2397.

89. Duncan PW, Wallace D, Lai SM, Johnson D, Embretson S, Laster LJ. The Stroke Impact Scale Version 2.0. Evaluation of reliability, validity, and sensitivity to change. Stroke 1999;30:2131-2140.

90. Duncan PW, Bode RK, Lai SM, Perera S. Rasch analysis of a new stroke-specific outcome scale: The Stroke Impact Scale. Arch Phys Med Rehab 2003;84:950-963.

91. World Health Organization. Global Burden of Stroke. The Atlas of Heart Disease and Stroke. www.who.int/cardiovascular_diseases/resources/atlas/en. Accessed May 2007.

92. Nakayama H, Jorgensen HS, Raaschou HO, Olsen TS. Recovery of Upper Extremity Function in Stroke Patients - the Copenhagen Stroke Study. Arch Phys Med Rehab 1994;75:394-398.

93. Nakayama H, Jorgensen HS, Raaschou HO, Olsen TS. Compensation in recovery of upper extremity function after stroke: the Copenhagen Stroke Study. Arch Phys Med Rehab 1994;75:852-857.

94. Duncan PW, Goldstein LB, Horner RD, Landsman PB, Samsa GP, Matchar DB. Similar motor recovery of upper and lower extremities after stroke. Stroke 1994;25:1181-1188.

95. Langhammer B, Stanghelle JK. Bobath or motor relearning programme? A comparison of two different approaches of physiotherapy in stroke rehabilitation: a randomized controlled study. Clin Rehabil 2000;14:361-369.

96. Jongbloed L, Stacey S, Brighton C. Stroke rehabilitation: sensorimotor integrative treatment versus functional treatment. Am J Occup Ther 1989;43:391-397.

97. Fasoli SE, Krebs HI, Stein J, Frontera WR, Hogan N. Effects of robotic therapy on motor impairment and recovery in chronic stroke. Arch Phys Med Rehabil 2003;84:477-482.

98. Feys HM, De Weerdt WJ, Selz BE, Cox Steck GA, Spichiger R, Vereeck LE, Putman KD, Van Hoydonck GA. Effect of a therapeutic intervention for the hemiplegic upper limb in the acute phase after stroke: a single-blind, randomized, controlled multicenter trial. Stroke 1998;29:785-792.

99. Jang SH, Kim YH, Cho SH, Lee JH, Park JW, Kwon YH. Cortical reorganization induced by task-oriented training in chronic hemiplegic stroke patients. Neuroreport 2003;14:137-141.

100. Lum PS, Burgar CG, Shor PC, Majmundar M, Van der LM. Robot-assisted movement training compared with conventional therapy techniques for the rehabilitation of upper-limb motor function after stroke. Arch Phys Med Rehabil 2002;83:952-959.

101. Chae J, Bethoux F, Bohine T, Dobos L, Davis T, Friedl A. Neuromuscular stimulation for upper extremity motor and functional recovery in acute hemiplegia. Stroke 1998;29:975-979.

102. Francisco G, Chae J, Chawla H, Kirshblum S, Zorowitz R, Lewis G, Pang S. Electromyogram-triggered neuromuscular stimulation for improving the arm function of acute stroke survivors: a randomized pilot study. Arch Phys Med Rehabil 1998;79:570-575.

103. Cauraugh J, Light K, Kim S, Thigpen M, Behrman A. Chronic motor dysfunction after stroke: recovering wrist and finger extension by electromyography-triggered neuromuscular stimulation. Stroke 2000;31:1360-1364.

104. Winstein CJ, Rose DK, Tan SM, Lewthwaite R, Chui HC, Azen SP. A randomized controlled comparison of upper-extremity rehabilitation strategies in acute stroke: A pilot study of immediate and long-term outcomes. Arch Phys Med Rehabil 2004;85:620-628.

105. Liepert J, Bauder H, Wolfgang HR, Miltner WH, Taub E, Weiller C. Treatment-induced cortical reorganization after stroke in humans. Stroke 2000;31:1210-1216.

106. Taub E, Uswatte G, Elbert T. New treatments in neurorehabilitation founded on basic research. Nat Rev Neurosci 2002;3:228-236.

107. Kwakkel G, Wagenaar RC, Twisk JW, Lankhorst GJ, Koetsier JC. Intensity of leg and arm training after primary middle-cerebral-artery stroke: a randomised trial. Lancet 1999;354:191-196.

108. Duncan P, Studenski S, Richards L, Gollub S, Lai SM, Reker D, Perera S, Yates J, Koch V, Rigler S, Johnson D. Randomized clinical trial of therapeutic exercise in subacute stroke. Stroke 2003;34:2173-2180.

109. Salbach NM, Mayo NE, Wood-Dauphinee S, Hanley JA, Richards CL, Cote R. A task-oriented intervention enhances walking distance and speed in the first year post stroke: a randomized controlled trial. Clin Rehabil 2004;18:509-519.

110. Roccaforte WH, Burke WJ, Bayer BL, Wengel SP. Validation of a telephone version of the mini-mental state examination. J Am Geriatr Soc 1992;40:697-702.

111. Gibbons WJ, Fruchter N, Sloan S, Levy RD. Reference values for a multiple repetition 6-minute walk test in healthy adults older than 20 years. J Cardiopulm Rehabil 2001;21:87-93.

112. Butland RJ, Pang J, Gross ER, Woodcock AA, Geddes DM. Two-, six-, and 12-minute walking tests in respiratory disease. Br Med J 1982;284:1607-1608.

113. World Health Organization. International Classification of Impairments, Disabilities, and Handicaps. Geneva:, 1980.

114. Daley, KA The STroke REhabilitation Assessment of Movement (STREAM): content validity and preliminary reliability. 1994. McGill University.

115. Daley KA, Mayo N, Danys I, Cabot R, Wood-Dauphinee S. The STroke REhabili1tation Assessment of Movement (STREAM): refining and validating the content. Physiother Can 1997;49:269-278.

116. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, Leirer VO. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1983;17:37-49.

117. Salbach NM, Mayo NE, Robichaud-Ekstrand S, Hanley JA, Richards CL, Wood-Dauphinee S. The effect of a task-oriented walking intervention on improving balance self-efficacy poststroke: a randomized, controlled trial. J Am Geriatr Soc 2005;53:576-582.

118. Higgins J, Mayo NE, Desrosiers J, Salbach NM, Ahmed, S. Upper extremity function and recovery in the acute phase post stroke. J Rehabil Res Dev 2004;42:65-76.

119. Fairclough DL. Workshop on imputation of missing data in QOL research. 1998. Baltimore. At International Society for Quality of Life Research (ISOQOL) Annual Conference.

120. McEwen, S. Performance-based Correlates of Health-related Quality of Life in Community-dwelling Persons with Stroke. 1995. McGill University.

121. van der Lee JH, Snels IA, Beckerman H, Lankhorst GJ, Wagenaar RC, Bouter LM. Exercise therapy for arm function in stroke patients: a systematic review of randomized controlled trials. Clin Rehabil 2001;15:20-31.

122. van der Lee JH. Constraint-induced therapy for stroke: more of the same or something completely different? Curr Opin Neurol 2001;14:741-744.

123. Sterr A, Elbert T, Berthold I, Kolbel S, Rockstroh B, Taub E. Longer versus shorter daily constraint-induced movement therapy of chronic hemiparesis: an exploratory study. Arch Phys Med Rehabil 2002;83:1374-1377.

124. Wu C, Trombly CA, Lin K, Tickle-Degnen L. Effects of object affordances on reaching performance in persons with and without cerebrovascular accident. Am J Occup Ther 1998;52:447-456.

125. Nunnally J. Psychometric Theory. New York: Mc Graw-Hill, 1978.

126. Wade DT. Measurement in Neurological Rehabilitation. New York: Oxford University Press, 1996.

127. Embretson SE. The New Rules of Measurement. Psychological Assessment 1996;8:341-349.

128. Reise, S. P. IRT Workshop. Advances in Health Outcomes Measurement. Exploring the current Stat and the Future Applications of Item Response Theory, Item Banks, and Computer-adaptive Testing. 2004. Bethesda, MD.

129. Hambleton RK. Fundamentals of Item Response Theory. Sage Production Editor: Diane S.Foster. Newbury Park CA, Sage publications, 1991.

130. Stucki G, Daltroy L, Katz JN, Johannesson M, Liang MH. Interpretation of change scores in ordinal clinical scales and health status measures: The whole may not equal the sum of the parts. J Clin Epidemiol 1996;49:711-717.

131. Streiner DL, Norman G. Health Measurement Scales: A Practical Guide to their Development and Use. Oxford: Oxford University Press, 1995.

132. Messick S. Validity of Psychological Assessment. Am Psychol 1995;50:741-749.

133. Messick S. Meaning and values in test validation. EducRes 1989;18:5-11.

134. Messick S. The Interplay of Evidence and Consequences in the Validation of Performance Assessments. Educ Researcher 1994;23:13-23.

135. Finch E, Brooks D, Stratford PW, Mayo NE. Physical rehabilitation outcome measures. Hamilton: BC Decker Inc., 2002.

136. Steven SS. On the Theory of Scales of Measurement. Science 1946;103:667-680.

137. Bond TG, Fox C. Applying The Rasch Model. Mahwah, New Jersey: Lawrence Erlbaum Associates, 2001.

138. Wright BD, Stone MH. The measurement model. Best Test Design: Rasch Measurement. Chicago: Mesa Press, 1979:1-17.

139. Velozo CA, Kielhofner G, Lai JS. The use of Rasch analysis to produce scale-free measurement of functional ability. Am J Occup Ther 1999;53:83-90.

140. Andersen EB, Olsen LW. The life of Georg Rasch as a Mathematician and as a Statistician. In Boomsma A, van Duijn MAJ, Sniders TAB, editors. Essays in item response theory. New York: Springer, 2000.

141. Guttman. The Basis for Scalogram Analysis. In S.A.Stouffer, editor. Measurement and Prediction. New York: Wiley, 1950.

142. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: The University of Chicago Press, 1980.

143. Andrich D, Sheridan B. Interpreting RUMM 2020. Rumm Laboratory Pty Ltd, 2005.

144. Masters, G. N. The Key to Objective Measurement in the Psychosocial Sciences. Camberwell. Australian Council for Educational Research, 2001.

145. Andrich D. Rasch Models for measurement. Sage, 1988:9-94.

146. Linacre JM. Estimation Methods for Rasch Measures. In Smith EV, Smith RM, editors. Introduction To Rasch Measurements. Maple Grove: JAM press, 2004:25-47.

147. Andrich, D, Sheridan, B., and Luo, G. RUMM2020: A Windows interactive program for analysising data with Rasch Unidimensional Models for Measurement. RUMM. [RUMM2020]. 2006. Perth Western Australia, RUMM laboratory.

148. Smith EV, Jr. Evidence for the reliability of measures and validity of measure interpretation: a Rasch measurement perspective. J Appl Meas 2001;2:281-311.

149. Smith RM. Fit analysis in latent trait measurement models. J Appl Meas 2000;1:199-218.

150. Stone MH. Substantive scale construction. J Appl Meas 2003;4:282-297.

151. Wright BD, Panchapakesan N. A procedure for sample free item analysis. Educ Psychol Meas 1969;29:23-48.

152. Andersen EB. Goodness of Fit Test for Rasch Model. Psychometrika 1973;38:123-140.

153. Gustafason J.E. Testing and obtaining fit of data to the Rasch model. Br J Math Stat Psychol 1980;33:205-233.

154. van den Wollenberg A.L. Two new test statistics for the Rasch model. Psychometrika 1982;47:123-140.

155. Smith R.M., Hedges LV. A comparison of likelihood ratio $\chi 2$ and Pearsonian $\chi 2$ tests of fit in the Rasch model. Educ Res Perspect 1982;9:44-54.

156. Karabatsos G. A critique of Rasch residual fit statistics. J Appl Meas 2000;1:152-176.

157. Rost J, von Davier M. A conditional item fit index for Rasch model. Applied Psychological Measurement 1994;18:171-182.

158. Levine MV, Drasgow F. Optimal appropriateness measurement. Psychometrika 1988;53:161-176.

159. Detecting inappropriate test scores with optimal and practical appropriateness indices. Appl Psychol Measurement 1987;11:59-79.

160. Klauer KC. The Assessment of Persons fit. In GH Fischer and IW Molenaar, editor. Rasch Models: Foundations, Recent developments, and Applications. New York: Springer-Verlag, 1995.

161. Smith RM et al. Using Item Mean Squares to Evaluate Fit to the Rasch Model. 1-17. 1995.

162. Wright BD, Linacre JM. Reasonable Mean-Square fit values. Rasch Measurement Transactions 1994;8:370.

163. Guilford JP. Fundamental Statistics in Psychology and Education. New York: 1965.

164. Linacre JM. Investigating rating scale category utility. J Outcome Meas 1999;3:103-122.

165. Smith R.M. A Comparison of Methods for Determining Dimensionality in Rasch Measurement. Struct Equ Model 1996;3:25-40.

166. Tennant A, Pallant JF. Unidimensionality Matters! (A Tale of Two Smiths?). Rasch Measurement Transactions 2006;20:1048-1051.

167. Wright B, Masters GN. Rating Scale Analysis. Chicago: MESA Press, 1982.

168. Scott SC, Goldberg MS, Mayo NE. Statistical assessment of ordinal outcomes in comparative studies. J Clin Epidemiol 1997;50:45-55.

169. Williams NJ, Beretvas SN. DIF identification using HGLM for polytomous items. Appl Psychol Meas 2006;30:22-42.

170. Health outcomes methodology. Med Care 2000;38:II7-13.

171. Flaugher R. Item Pools. In Wainer H., editor. Computerized Adaptive Testing: A primer. Hillsdale, NJ: Lawrence Erlbaum Associates, 2000:38.

172. Luo, Seow and Chin. Linking and Anchoring Techniques in Test Equating Using the Rasch Model. 2006. Loughborough University. Proceedings of the 5th CAA Conference.

173. Wolfe EW, Smith EV, Jr. Instrument development tools and activities for measure validation using Rasch models: part II--validation activities. J Appl Meas 2007;8:204-234.

174. Fisher WP. Reliability Statistic. Rasch Measurement Transactions 1992;6:288.

175. Mokken R.J. Nonparametric models for dichotomous responses. In W.J.van der Linden and R.K.Hambleton, editor. Handbood of modern item response theory. New York: Springer-Verlag, 1997:351-368.

176. Hobart J. Measuring disease impact in disabling neurological conditions: are patients' perspectives and scientific rigor compatible? Curr Opin Neurol 2002;15:721-724.

177. Massof RW. The measurement of vision disability. Optome Vis Sci 2002;79:516-552.

178. Lindeboom R, Vermeulen M, Holman R, De Haan RJ. Activities of daily living instruments:optimizing scales for neurologic assessments. Neurology 2003;60:738-742.

179. Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. J Rehabil Med 2003;35:105-115.

180. Hart DL, Mioduski JE, Werneke MW, Stratford PW. Simulated computerized adaptive test for patients with lumbar spine impairments was efficient and produced valid measures of function. J Clin Epidemiol 2006;59:947-956.

181. Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, Williams JI. The Quebec Back Pain Disability Scale: conceptualization and development. J Clin Epidemiol 1996;49:151-161.

182. Young NL, Williams JI, Yoshida KK, Wright JG. Measurement properties of the activities scale for kids. J Clin Epidemiol 2000;53:125-137.

183. Greenlund KJ, Giles WH, Keenan NL, Croft JB, Mensah GA. Physician advice, patient actions, and health-related quality of life in secondary prevention of stroke through diet and exercise. Stroke 2002;33:565-570.

184. Turano KA, Massof RW, Quigley HA. A self-assessment instrument designed for measuring independent mobility in RP patients: generalizability to glaucoma patients. Invest Ophthalmol Vis Sci 2002;43:2874-2881.

185. Antonucci G, Aprile T, Paolucci S. Rasch analysis of the Rivermead Mobility Index: A study using mobility measures of first-stroke inpatients. Arch Phys Med Rehab 2002;83:1442-1449.

186. Azouvi P, Olivier S, de Montety G, Samuel C, Louis-Dreyfus A, Tesio L. Behavioral assessment of unilateral neglect: study of the psychometric properties of the Catherine Bergego Scale. Arch Phys Med Rehabil 2003;84:51-57.

187. Brock KA, Goldie PA, Greenwood KM. Evaluating the effectiveness of stroke rehabilitation: choosing a discriminative measure. Arch Phys Med Rehabil 2002;83:92-99.

188. Duncan PW, Lai SM, Bode RK, Perera S, DeRosa J. Stroke Impact Scale-16: A brief assessment of physical function. Neurology 2003;60:291-296.

189. Engberg A, Garde B, Kreiner S. Rasch analysis in the development of a rating scale for assessment of mobility after stroke. Acta Neurol Scand 1995;91:118-127.

190. Fujiwara T, Liu M, Tsuji T, Sonoda S, Mizuno K, Akaboshi K, Hase K, Masakado Y, Chino N. Development of a new measure to assess trunk impairment after stroke (trunk impairment scale): its psychometric properties. Am J Phys Med Rehabil 2004;83:681-688.

191. Larson EB, Kirschner K, Bode RK, Heinemann AW, Clorfene J, Goodman R. Brief cognitive assessment and prediction of functional outcome in stroke. Top Stroke Rehabil 2003;9:10-21.

192. Tsuji T, Liu M, Sonoda S, Domen K, Chino N. The stroke impairment assessment set: its internal consistency and predictive validity. Arch Phys Med Rehabil 2000;81:863-868.

193. Liu M, Chino N, Tuji T, Masakado Y, Hase K, Kimura A. Psychometric properties of the Stroke Impairment Assessment Set (SIAS). Neurorehabil Neural Repair 2002;16:339-351.

194. Patomella AH, Tham K, Kottorp A. P-drive: assessment of driving performance after stroke. J Rehabil Med 2006;38:273-279.

195. Penta M, Tesio L, Arnould C, Zancan A, Thonnard JL. The ABILHAND questionnaire as a measure of manual ability in chronic stroke patients: Rasch-based validation and relationship to upper limb impairment. Stroke 2001;32:1627-1634.

196. Hsueh IP, Wang WC, Sheu CF, Hsieh CL. Rasch analysis of combining two indices to assess comprehensive ADL function in stroke patients. Stroke 2004;35:721-726.

197. Bode RK, Heinemann AW, Semik P. Measurement properties of the Galveston Orientation and Amnesia Test (GOAT) and improvement patterns during inpatient rehabilitation. J Head Trauma Rehabil 2000;15:637-655.

198. Malec JF. Comparability of Mayo-Portland Adaptability Inventory ratings by staff, significant others and people with acquired brain injury. Brain Inj 2004;18:563-575.

199. Doninger NA, Ehde DM, Bode RK, Knight K, Bombardier CH, Heinemann AW. Measurement properties of the Neurobehavioral Cognitive Status Examination (Cognistat) in traumatic brain injury rehabilitation. Rehabil Psychol 2006;51:281-288.

200. Malec JF. The Mayo-Portland Participation Index: A brief and psychometrically sound measure of brain injury outcome. Arch Phys Med Rehabil 2004;85:1989-1996.

201. Kutlay S, Kucukdeveci AA, Elhan AH, Yavuzer G, Tennant A. Validation of the Middlesex Elderly Assessment of Mental State (MEAMS) as a cognitive screening test in patients with acquired brain injury in Turkey. Disabil Rehabil 2007;29:315-321.

202. Tennant A, Kucukdeveci AA, Kutlay S, Elhan AH. Assessing normative cut points through differential item functioning analysis: an example from the adaptation of the Middlesex Elderly Assessment of Mental State (MEAMS) for use as a cognitive screening test in Turkey. Health Qual Life Outcomes 2006;4:18.

203. Arnould C, Penta M, Renders A, Thonnard JL. ABILHAND-Kids: a measure of manual ability in children with cerebral palsy. Neurology 2004;63:1045-1052.

204. Avery LM, Russell DJ, Raina PS, Walter SD, Rosenbaum PL. Rasch analysis of the Gross Motor Function Measure: validating the assumptions of the Rasch

model to create an interval-level measure. Arch Phys Med Rehabil 2003;84:697-705.

205. Resnick B, Orwig D, Zimmerman S, Hawkes W, Golden J, Werner-Bronzert M, Magaziner J. Testing of the SEE and OEE post-hip fracture. West J Nurs Res 2006;28:586-601.

206. Shi W, Wang S, Liao Y, Yang H, Xu X, Shao X. Reliability and validity of the GMFM-66 in 0- to 3-year-old children with cerebral palsy. Am J Phys Med Rehabil 2006;85:141-147.

207. Chen CC, Bode RK, Granger CV, Heinemann AW. Psychometric properties and developmental differences in children's ADL item hierarchy: a study of the WeeFIM instrument. Am J Phys Med Rehabil 2005;84:671-679.

208. Hobart JC, Riazi A, Thompson AJ, Styles IM, Ingram W, Vickery PJ, Warner M, Fox PJ, Zajicek JP. Getting the measure of spasticity in multiple sclerosis: the Multiple Sclerosis Spasticity Scale (MSSS-88). Brain 2006;129:224-234.

209. Tesio L, Perucca L, Franchignoni FP, Battaglia MA. A short measure of balance in multiple sclerosis: validation through Rasch analysis. Funct Neurol 1997;12:255-265.

210. Cattaneo D, Marazzini F, Crippa A, Cardini R. Do static or dynamic AFOs improve balance? Clin Rehabil 2002;16:894-899.

211. Franchignoni F, Ferriero G, Giordano A, Guglielmi V, Picco D. Rasch psychometric validation of the Impact on Participation and Autonomy questionnaire in people with Parkinson's disease. Eura Medicophys 2007.

212. Altangerel U, Spaeth GL, Steinmann WC. Assessment of function related to vision (AFREV). Ophthalmic Epidemiol 2006;13:67-80.

213. Haymes SA, Johnston AW, Heyes AD. The development of the Melbourne low-vision ADL index: a measure of vision disability. Invest Ophthalmol Vis Sci 2001;42:1215-1225.

214. Lamoureux EL, Pallant JF, Pesudovs K, Hassell JB, Keeffe JE. The Impact of Vision Impairment Questionnaire: an evaluation of its measurement properties using Rasch analysis. Invest Ophthalmol Vis Sci 2006;47:4732-4741.

215. Lamoureux EL, Pallant JF, Pesudovs K, Rees G, Hassell JB, Keeffe JE. The impact of vision impairment questionnaire: an assessment of its domain structure

using confirmatory factor analysis and rasch analysis. Invest Ophthalmol Vis Sci 2007;48:1001-1006.

216. Massof RW, Hsu CT, Baker FH, Barnett GD, Park WL, Deremeik JT, Rainey C, Epstein C. Visual disability variables. I: the importance and difficulty of activity goals for a sample of low-vision patients. Arch Phys Med Rehabil 2005;86:946-953.

217. Massof RW, Hsu CT, Baker FH, Barnett GD, Park WL, Deremeik JT, Rainey C, Epstein C. Visual disability variables. II: The difficulty of tasks for a sample of low-vision patients. Arch Phys Med Rehabil 2005;86:954-967.

218. Misajon R, Hawthorne G, Richardson J, Barton J, Peacock S, Iezzi A, Keeffe J. Vision and quality of life: the development of a utility measure. Invest Ophthalmol Vis Sci 2005;46:4007-4015.

219. Stelmack J, Szlyk JP, Stelmack T, Babcock-Parziale J, Demers-Turco P, Williams RT, Massof RW. Use of Rasch person-item map in exploratory data analysis: A clinical perspective. J Rehabil Res Dev 2004;41:233-241.

220. Stelmack JA, Szlyk JP, Stelmack TR, Demers-Turco P, Williams RT, Moran D, Massof RW. Psychometric properties of the Veterans Affairs Low-Vision Visual Functioning Questionnaire. Invest Ophthalmol Vis Sci 2004;45:3919-3928.

221. Szlyk JP, Stelmack J, Massof RW, Stelmack TR, Demers-Turco P, Williams RT, Wright BD. Performance of the Veterans Affairs Low Vision Visual Functioning Questionnaire. J Vis Impair Blind 5;98:261-275.

222. Velozo CA, Lai JS, Mallinson T, Hauselman E. Maintaining instrument quality while reducing items: application of Rasch analysis to a self-report of visual function. J Outcome Meas 2000;4:667-680.

223. Cook CE, Richardson JK, Pietrobon R. Dimensionality, internal consistency, and item analysis of the National Health and Nutrition Examination Surveys activities of daily living instrument among patients with report of low back pain. J Manipulative Physiol Ther 2006;29:183-189.

224. Davidson M, Keating JL, Eyres S. A low back-specific version of the SF-36 Physical Functioning scale. Spine 2004;29:586-594.

225. Kucukdeveci AA, Tennant A, Elhan AH, Niyazoglu H. Validation of the Turkish version of the Roland-Morris Disability Questionnaire for use in low back pain. Spine 2001;26:2738-2743.

226. Page SJ, Shawaryn MA, Cernich AN, Linacre JM. Scaling of the revised Oswestry low back pain questionnaire. Arch Phys Med Rehabil 2002;83:1579-1584.

227. White LJ, Velozo CA. The use of Rasch measurement to improve the Oswestry classification scheme. Arch Phys Med Rehab 2002;83:822-831.

228. Chiou CF, Suarez-Almazor ME, Sherbourne CD, Chang CH, Reyes C, Dylan M, Ofman J, Wallace DJ, Mizutani W, Weisman M. Development and validation of a preference weight multiattribute health outcome measure for rheumatoid arthritis. J Rheumatol 2006;33:2409-2411.

229. Penta M, Thonnard JL, Tesio L. ABILHAND: a Rasch-built measure of manual ability. Arch Phys Med Rehabil 1998;79:1038-1042.

230. Wolfe F, Michaud K, Kahler K, Omar M. The Short Arthritis Assessment Scale: a brief assessment questionnaire for rapid evaluation of arthritis severity in research and clinical practice. J Rheumatol 2004;31:2472-2479.

231. Helliwell P, Reay N, Gilworth G, Redmond A, Slade A, Tennant A, Woodburn J. Development of a foot impact scale for rheumatoid arthritis. Arthritis Rheum/Arthritis Care Res 2005;53:418-422.

232. Kucukdeveci AA, Sahin H, Ataman S, Griffiths B, Tennant A. Issues in cross-cultural validity: example from the adaptation, reliability, and validity testing of a Turkish version of the Stanford Health Assessment Questionnaire. Arthritis Rheum 2004;51:14-19.

233. O'Malley KJ, Suarez-Almazor M, Aniol J, Richardson P, Kuykendall DH, Moseley JB, Jr., Wray NP. Joint-specific multidimensional assessment of pain (J-MAP): factor structure, reliability, validity, and responsiveness in patients with knee osteoarthritis. J Rheumatol 2003;30:534-543.

234. Wolfe F, Michaud K, Pincus T. Development and validation of the health assessment questionnaire II: a revised version of the health assessment questionnaire. Arthritis Rheum 2004;50:3296-3305.

235. Wolfe F. Pain extent and diagnosis: development and validation of the regional pain scale in 12,799 patients with rheumatic disease. J of Rheumatol 2003;30:369-378.

236. Kopec JA, Sayre EC, Davis AM, Badley EM, Abrahamowicz M, Sherlock L, Williams JI, Anis AH, Esdaile JM. Assessment of health-related quality of life in

arthritis: conceptualization and development of five item banks using item response theory. Health Qual Life Outcomes 2006;4:33.

237. Aamodt G, Kjendahl A, Jahnsen R. Dimensionality and scalability of the Motor Assessment Scale (MAS). Disabil Rehabil 2006;28:1007-1013.

238. Bjorkdahl A, Nilsson AL, Sunnerhagen KS. The structural properties of the European Brain Injury Questionnaire. J Stroke Cerebrovasc Dis 2004;13:122-128.

239. Doyle PJ, Hula WD, McNeil MR, Mikolic JM, Matthews C. An application of Rasch analysis to the measurement of communicative functioning. J Speech Lang Hear Res 2005;48:1412-1428.

240. Lance CE, Butts MM, Michels LC. The sources of four commonly reported cutoff criteria - What did they really say? Organizational Research Methods 2006;9:202-220.

241. Heinemann AW, Harvey RL, McGuire JR, Ingberman D, Lovell L, Semik P, Roth EJ. Measurement properties of the NIH Stroke Scale during acute rehabilitation. Stroke 1997;28:1174-1180.

242. Hsueh IP, Wang WC, Wang CH, Sheu CF, Lo SK, Lin JH, Hsieh CL. A simplified stroke rehabilitation assessment of movement instrument. Phys Ther 2006;86:936-943.

243. Hula W, Doyle PJ, McNeil MR, Mikolic JM. Rasch modeling of revised token test performance: validity and sensitivity to change. J Speech Lang Hear Res 2006;49:27-46.

244. Jette AM, Haley SM, Ni P. Comparison of functional status tools used in post-acute care. Health Care Financ Rev 2003;24:13-24.

245. Koh CL, Hsueh IP, Wang WC, Sheu CF, Yu TY, Wang CH, Hsieh CL. Validation of the action research arm test using item response theory in patients after stroke. J Rehabil Med 2006;38:375-380.

246. Lai SM, Perera S, Duncan PW, Bode R. Physical and social functioning after stroke: comparison of the Stroke Impact Scale and Short Form-36. Stroke 2003;34:488-493.

247. Linn RT, Blair RS, Granger CV, Harper DW, O'Hara PA, Maciura E. Does the functional assessment measure (FAM) extend the functional independence

measure (FIM) instrument? A rasch analysis of stroke inpatients. J Outcome Meas 1999;3:339-359.

248. Lundgren-Nilsson A, Grimby G, Ring H, Tesio L, Lawton G, Slade A, Penta M, Tripolski M, Biering-Sorensen F, Carter J, Marincek C, Phillips S, Simone A, Tennant A. Cross-cultural validity of functional independence measure items in stroke: a study using Rasch analysis. J Rehabil Med 2005;37:23-31.

249. Millis SR, Straube D, Iramaneerat C, Smith EV, Jr., Lyden P. Measurement properties of the National Institutes of Health Stroke Scale for people with right- and left-hemisphere lesions: further analysis of the clomethiazole for acute stroke study-ischemic (class-I) trial. Arch Phys Med Rehabil 2007;88:302-308.

250. Sabari JS, Lim AL, Velozo CA, Lehman L, Kieran O, Lai JS. Assessing arm and hand function after stroke: a validity test of the hierarchical scoring system used in the motor assessment scale for stroke. Arch Phys Med Rehabil 2005;86:1609-1615.

251. Segal ME, Heinemann AW, Schall RR, Wright BD. Rasch analysis of a brief physical ability scale for long-term outcomes of stroke. First International Outcome Measurement Conference, co-sponsored by Rehabilitation Foundation, Inc., and the MESA Psychometric Laboratory at the University of Chicago. Physical Medicine & Rehabilitation 1997;11:385-396.

252. Tennant A, Penta M, Tesio L, Grimby G, Thonnard JL, Slade A, Lawton G, Simone A, Carter J, Lundgren-Nilsson A, Tripolski M, Ring H, Biering-Sorensen F, Marincek C, Burger H, Phillips S. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care 2004;42:I37-I48.

253. Tennant A, Geddes JML, Chamberlain MA. The Barthel Index: an ordinal score or interval level measure? Clin Rehabil 1996;10:301-308.

254. Tham K, Bernspang B, Fisher AG. Development of the assessment of awareness of disability. Scand J Occup Ther 12;6:184-190.

255. Tsuji T, Sonoda S, Domen K, Saitoh E, Liu M, Chino N. ADL structure for stroke patients in Japan based on the Functional Independence Measure. Am J Phys Med Rehabil 1995;74:432-438.

256. Van de WA, Feys H, van der KS, Messerli R, Baronti F, Lehmann R, Van Hemelrijk B, Pante F, Perfetti C, De Weerdt W. Can quality of movement be measured? Rasch analysis and inter-rater reliability of the Motor Evaluation Scale

for Upper Extremity in Stroke Patients (MESUPES). Clin Rehabil 2006;20:871-884.

257. van Hartingsveld F, Lucas C, Kwakkel G, Lindeboom R. Improved interpretation of stroke trial results using empirical Barthel item weights. Stroke 2006;37:162-166.

258. Dallmeijer AJ, de G, V, Roorda LD, Schepers VP, Lindeman E, van den Berg LH, Beelen A, Dekker J. Cross-diagnostic validity of the SF-36 physical functioning scale in patients with stroke, multiple sclerosis and amyotrophic lateral sclerosis: a study using Rasch analysis. J Rehabil Med 2007;39:163-169.

259. Dallmeijer AJ, Dekker J, Roorda LD, Knol DL, van Baalen B, de G, V, Schepers VP, Lankhorst GJ. Differential item functioning of the Functional Independence Measure in higher performing neurological patients. J Rehabil Med 2005;37:346-352.

260. Gandek B, Sinclair SJ, Jette AM, Ware JE, Jr. Development and Initial Psychometric Evaluation of the Participation Measure for Post-Acute Care (PM-PAC). Am J Phys Med Rehabil 2006.

261. Hart DL, Wright BD. Development of an index of physical functional health status in rehabilitation. Arch Phys Med Rehabil 2002;83:655-665.

262. Heinemann AW, Linacre JM, Wright BD, Hamilton BB, Granger C. Relationships between impairment and physical disability as measured by the Functional Independence Measure. Arch Phys Med Rehabil 1993;74:566-573.

263. Holman R, Weisscher N, Glas CA, Dijkgraaf MG, Vermeulen M, De Haan RJ, Lindeboom R. The Academic Medical Center Linear Disability Score (ALDS) item bank: item response theory analysis in a mixed patient population. Health Qual Life Outcomes 2005;3:83.

264. Kirby RL, Swuste J, Dupuis DJ, MacLeod DA, Monroe R. The Wheelchair Skills Test: a pilot study of a new outcome measure. Arch Phys Med Rehabil 2002;83:10-18.

265. Kucukdeveci AA, Yavuzer G, Elhan AH, Sonel B, Tennant A. Adaptation of the Functional Independence Measure for use in Turkey. Clin Rehabil 2001;15:311-319.

266. Kucukdeveci AA, Yavuzer G, Tennant A, Suldur N, Sonel B, Arasil T. Adaptation of the modified Barthel Index for use in physical medicine and rehabilitation in Turkey. Scand J Rehabil Med 2000;32:87-92.

267. Lundgren-Nilsson A, Tennant A, Grimby G, Sunnerhagen KS. Cross-diagnostic validity in a generic instrument: an example from the Functional Independence Measure in Scandinavia. Health Qual Life Outcomes 2006;4:55.

268. Nilsson AL, Sunnerhagen KS, Grimby G. Scoring alternatives for FIM in neurological disorders applying Rasch analysis. Acta Neurol Scand 2005;111:264-273.

269. Ostir GV, Granger CV, Black T, Roberts P, Burgos L, Martinkewiz P, Ottenbacher KJ. Preliminary results for the PAR-PRO: a measure of home and community participation. Arch Phys Med Rehabil 2006;87:1043-1051.

270. Silverstein B, Fisher WP, Kilgore KM, Harley JP, Harvey RF. Applying psychometric criteria to functional assessment in medical rehabilitation: II. Defining interval measures. Arch Phys Med Rehabil 1992;73:507-518.

271. Smith RM, Taylor PA. Equating rehabilitation outcome scales: developing common metrics. J Appl Meas 2004;5:229-242.

272. Bedell GM. Developing a follow-up survey focused on participation of children and youth with acquired brain injuries after discharge from inpatient rehabilitation. Neurorehabilitation 2004;19:191-205.

273. Bedell GM, Haley SM, Coster WJ, Smith KW. Developing a responsive measure of change for paediatric brain injury inpatient rehabilitation. Brain Inj 2002;16:659-671.

274. Cantagallo A, Carli S, Simone A, Tesio L. MINDFIM: a measure of disability in high-functioning traumatic brain injury outpatients. Brain Inj 2006;20:913-925.

275. Elhan AH, Kutlay S, Kucukdeveci AA, Cotuk C, Ozturk G, Tesio L, Tennant A. Psychometric properties of the Mini-Mental State Examination in patients with acquired brain injury in Turkey. J Rehabil Med 2005;37:306-311.

276. Hart T, Whyte J, Millis S, Bode R, Malec J, Richardson RN, Hammond F. Dimensions of disordered attention in traumatic brain injury: further validation of the Moss Attention Rating Scale. Arch Phys Med Rehabil 2006;87:647-655.

277. Johnston MV, Shawaryn MA, Malec J, Kreutzer J, Hammond FM. The structure of functional and community outcomes following traumatic brain injury. Brain Inj 2006;20:391-407.

278. Kothari DH, Haley SM, Gill-Body KM, Dumas HM. Measuring functional change in children with acquired brain injury (ABI): comparison of generic and ABI-specific scales using the Pediatric Evaluation of Disability Inventory (PEDI). Phys Ther 2003;83:776-785.

279. Malec JF, Moessner AM, Kragness M, Lezak MD. Refining a measure of brain injury sequelae to predict postacute rehabilitation outcome: rating scale analysis of the Mayo-Portland Adaptability Inventory. J Head Trauma Rehabil 2000;15:670-682.

280. Malec JF, Kragness M, Evans RW, Finlay KL, Kent A, Lezak MD. Further psychometric evaluation and revision of the Mayo-Portland Adaptability Inventory in a national sample. J Head Trauma Rehabil 2003;18:479-492.

281. Tesio L, Cantagallo A. The functional assessment measure (FAM) in closed traumatic brain injury outpatients: a Rasch-based psychometric study. J Outcome Meas 1998;2:79-96.

282. Whyte J, Hart T, Bode RK, Malec JF. The Moss Attention Rating Scale for traumatic brain injury: initial psychometric assessment. Arch Phys Med Rehabil 2003;84:268-276.

283. Williams GP, Robertson V, Greenwood KM, Goldie PA, Morris ME. The high-level mobility assessment tool (HiMAT) for traumatic brain injury. Part 2: content validity and discriminability. Brain Inj 2005;19:833-843.

284. Grimby G, Andren E, Holmgren E, Wright B, Linacre JM, Sundh V. Structure of a combination of Functional Independence Measure and Instrumental Activity Measure items in community-living persons: a study of individuals with cerebral palsy and spina bifida. Arch Phys Med Rehabil 1996;77:1109-1114.

285. Kottorp A, Bernspang B, Fisher AG. Validity of a performance assessment of activities of daily living for people with developmental disabilities. J Intellect Disabil Res 2003;47:597-605.

286. Russell DJ, Avery LM, Rosenbaum PL, Raina PS, Walter SD, Palisano RJ. Improved scaling of the gross motor function measure for children with cerebral palsy: evidence of reliability and validity. Phys Ther 2000;80:873-885.

287. Hagell P. WDMSPLO. Health status Measurement in Parkinson's Disease: Validity of the PDQ-39 and Nottingham Health Profile. Movement Disorders 2003;18:773-783.

288. Babcock-Parziale J, McKnight PE, Head DN. Evaluating psychometric properties of a clinical and a self-report blind rehabilitation outcome measure. J Rehabil Res Dev 2005;42:487-498.

289. Massof RW. A systems model for low vision rehabilitation. II. Measurement of vision disabilities. Optom Vis Sci 1998;75:349-373.

290. Pesudovs K, Garamendi E, Keeves JP, Elliott DB. The Activities of Daily Vision Scale for cataract surgery outcomes: re-evaluating validity with Rasch analysis. Invest Ophthalmol Vis Sci 2003;44:2892-2899.

291. Garratt AM. Rasch analysis of the Roland Disability Questionnaire. Spine 2003;28:79-84.

292. Granger CV, Lackner JM, Kulas M, Russell CF. Outpatients with low back pain: an analysis of the rate per day of pain improvement that may be expected and factors affecting improvement. Am J Phys Med Rehabil 2003;82:253-260.

293. McArthur DL, Cohen MJ, Schandler SL. Rasch analysis of functional assessment scales: an example using pain behaviors. Arch Phys Med Rehabil 1991;72:296-304.

294. Velozo CA, Kielhofner G, Gern A, Lin F, Azhar F, Lai J, Fisher G. Worker role interview: toward validation of a psychosocial work-related measure. J Occup Rehabil 1999;9:153-168.

295. Velozo CA, Choi B, Zylstra SE, Santopoalo R. Measurement qualities of a self-report and therapist-scored functional capacity instrument based on the dictionary of occupational titles. J Occup Rehabil 2006;16:109-122.

296. Chiou C, Sherbourne CD, Ofman J, Lee M, Lubeck DP, Paulus HE, Weisman M. Development and validation of Cedars-Sinai Health-Related Quality of Life in Rheumatoid Arthritis (CSHQ-RA) short form instrument. Arthritis Rheum/Arthritis Care Res 2004;51:358-364.

297. Dawson J, Linsell L, Doll H, Zondervan K, Rose P, Carr A, Randall T, Fitzpatrick R. Assessment of the Lequesne index of severity for osteoarthritis of the hip in an elderly population. Osteoarthr Cartil 2005;13:854-860.

298. Gilworth G, Chamberlain MA, Harvey A, Woodhouse A, Smith J, Smyth MG, Tennant A. Development of a work instability scale for rheumatoid arthritis. Arthritis Rheum/Arthritis Care Res 2003;49:349-354.

299. Ryser L, Wright BD, Aeschlimann A, Mariacher-Gehler S, Stucki G. A new look at the Western Ontario and McMaster Universities Osteoarthritis Index using Rasch analysis. Arthritis Care Res 1999;12:331-335.

300. Sheehan TJ, DeChello LM, Garcia R, Fifield J, Rothfield N, Reisine S. Measuring disability: application of the Rasch model to activities of daily living (ADL/IADL). J Outcome Meas 2000;4:681-705.

301. Haley SM, Coster WJ, Andres PL, Ludlow LH, Ni P, Bond TL, Sinclair SJ, Jette AM. Activity outcome measurement for postacute care. Med Care 2004;42:I49-I61.

302. Haley SM, Andres PL, Coster WJ, Kosinski M, Ni P, Jette AM. Short-form activity measure for post-acute care. Arch Phys Med Rehabil 2004;85:649-660.

303. Stahl, J., Shumway, R., Bergstrom, B., & Fisher, A.G. On-line performance assessment using rating scales. J Outcome Meas 1997;1, 173-191..

304. Fisher AG. The assessment of IADL motor skills: an application of many-faceted Rasch analysis. Am J Occup Ther 1993;47:319-329.

305. Krumlinde-Sundholm L, Eliasson A. Development of the Assisting Hand Assessment: a Rasch-built measure intended for children with unilateral upper limb impairments. Scand J Occup Ther 2;10:16-26.

306. Hahn EA, Cella D. Health outcomes assessment in vulnerable populations: measurement challenges and recommendations. Arch Phys Med Rehabil 2003;84:S35-S42.

307. Gilworth G, Chamberlain MA, Bhakta B, Haskard D, Silman A, Tennant A. Development of the BD-QoL: a quality of life measure specific to Behçet's disease. J Rheumatol 2004;31:931-937.

308. White LJ, Straube D, Keehn MT. Using compensations to assess physical performance for ambulatory outpatients. Arch Phys Med Rehabil 2004;85:1519-1524.

309. Yamada S, Liu M, Hase K, Tanaka N, Fujiwara T, Tsuji T, Ushiba J. Development of a short version of the motor FIM for use in long-term care settings. J Rehabil Med 2006;38:50-56.

310. Hart DL, Cook KF, Mioduski JE, Teal CR, Crane PK. Simulated computerized adaptive test for patients with shoulder impairments was efficient and produced valid measures of function. J Clin Epidemiol 2006;59:290-298.

311. Hart DL, Mioduski JE, Stratford PW. Simulated computerized adaptive tests for measuring functional status were efficient with good discriminant validity in patients with hip, knee, or foot/ankle impairments. J Clin Epidemiol 2005;58:629-638.

312. Beaton DE, Wright JG, Katz JN. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Joint Surg Am 2005;87:1038-1046.

313. Tesio L, Alpini D, Cesarani A, Perucca L. Short form of the Dizziness Handicap Inventory: construction and validation through Rasch analysis. Am J Phys Med Rehabil 1999;78:233-241.

314. Marchetti GF, Whitney SL. Construction and validation of the 4-item dynamic gait index. Phys Ther 2006;86:1651-1660.

315. Lusardi MM, Smith EV, Jr. Development of a scale to assess concern about falling and applications to treatment programs. J Outcome Meas 1997;1:34-55.

316. Lai JS, Cella D, Chang CH, Bode RK, Heinemann AW. Item banking to improve, shorten and computerize self-reported fatigue: an illustration of steps to create a core item bank from the FACIT-Fatigue Scale. Qual Life Res 2003;12:485-501.

317. Cook KF, Roddey TS, Gartsman GM, Olson SL. Development and psychometric evaluation of the Flexilevel Scale of Shoulder Function. Med Care 2003;41:823-835.

318. Zelinski EM, Gilewski MJ. A 10-item Rasch modeled memory self-efficacy scale. Aging Ment Health 2004;8:293-306.

319. Chakravarty EF, Bjorner JB, Fries JF. Improving patient reported outcomes using item response theory and computerized adaptive testing. J Rheumatol 2007;34:1426-1431.

320. Haley SM, Jette AM, Coster WJ, Kooyoomjian JT, Levenson S, Heeren T, Ashba J. Late Life Function and Disability Instrument: II. Development and evaluation of the function component. J Gerontol A Biol Sci Med Sci 2002;57:M217-M222.

321. Jette AM, Haley SM, Coster WJ, Kooyoomjian JT, Levenson S, Heeren T, Ashba J. Late life function and disability instrument: I. Development and evaluation of the disability component. J Gerontol A Biol Sci Med Sci 2002;57:M209-M216.

322. Chen CC, Granger CV, Peimer CA, Moy OJ, Wald S. Manual Ability Measure (MAM-16): a preliminary report on a new patient-centred and task-oriented outcome measure of hand function. J Hand Surg [Br] 2005;30:207-216.

323. Mungas D, Reed BR, Kramer JH. Psychometrically matched measures of global cognition, memory, and executive function for assessment of cognitive decline in older persons. Neuropsychology 2003;17:380-392.

324. Avlund K, Schultz-Larsen K, Kreiner S. The measurement of instrumental ADL: content validity and construct validity. Aging (Milano) 1993;5:371-383.

325. Haley SM, Fragala Pinkham MA, Dumas HM, Ni P, Skrinar AM, Cox GF. A physical performance measure for individuals with mucopolysaccharidosis type I. Dev Med Child Neurol 2006;48:576-581.

326. Mungas D, Reed BR. Application of item response theory for development of a global functioning measure of dementia with linear measurement properties. Stat Med 2000;19:1631-1644.

327. Prieto L, Alonso J, Lamarca R, Wright BD. Rasch measurement for reducing the items of the Nottingham Health Profile. J Outcome Meas 1998;2:285-301.

328. Rejeski WJ, Ip EH, Katula JA, White L. Older adults' desire for physical competence. Med Sci Sports Exerc 2006;38:100-105.

329. Haley SM, Ni P, Hambleton RK, Slavin MD, Jette AM. Computer adaptive testing improved accuracy and precision of scores over random item selection in a physical functioning item bank. J Clin Epidemiol 2006;59:1174-1182.

330. Lautamo T, Kottorp A, Salminen A. Play assessment for group settings: a pilot study to construct an assessment tool. Scand J Occup Ther 9;12:136-144.

331. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. Health Serv Res 2005;40:1918-1930.

332. McHorney CA, Cohen AS. Equating health status measures with item response theory: illustrations with functional status items. Med Care 2000;38:II43-II59.

333. Luther SL, Kromrey J, Powell-Cope G, Rosenberg D, Nelson A, Ahmed S, Quigley P. A pilot study to modify the SF-36V physical functioning scale for use with veterans with spinal cord injury. Arch Phys Med Rehabil 2006;87:1059-1066.

334. Ryall NH, Eyres SB, Neumann VC, Bhakta BB, Tennant A. The SIGAM mobility grades: a new population-specific measure for lower limb amputees. Disabil Rehabil 2003;25:833-844.

335. Leong KP, Kong KO, Thong BY, Koh ET, Lian TY, Teh CL, Cheng YK, Chng HH, Badsha H, Law WG, Lau TC, Chew LC, Ho HJ, Pong LY, Hoi LS, Sangeetha N, Chan SP, Howe HS. Development and preliminary validation of a systemic lupus erythematosus-specific quality-of-life instrument (SLEQOL). Rheumatology (Oxford) 2005;44:1267-1276.

336. Ludlow LH, Haley SGBM. A Hierarchical Model of Functional Performance in Rehabilitation Medicine. Eval Health Prof 1992;15:59-75.

337. Velozo CA, Peterson EW. Developing meaningful Fear of Falling Measures for community dwelling elderly. Am J Phys Med Rehabil 2001;80:662-673.

338. Gehlert S, Chang CH, Bock RD, Hartlage SA. The WOMQOL instrument measured quality of life in women of reproductive age with no known pathology. J Clin Epidemiol 2006;59:525-533.

339. Buffart LM, Roebroeck ME, Pesch-Batenburg JMF, Janssen WGM, Stam HJ. Assessment of arm/hand functioning in children with a congenital transverse or longitudinal reduction deficiency of the upper limb. Disabil Rehabil 1;28:85-95.

340. Ahlstrom S, Bernspang B. Occupational performance of persons who have suffered a stroke: a follow-up study. Scand J Occup Ther 6;10:88-94.

341. Bonnier B, Eliasson AC, Krumlinde-Sundholm L. Effects of constraint-induced movement therapy in adolescents with hemiplegic cerebral palsy: a day camp model. Scand J Occup Ther 2006;13:13-22.

342. Kottorp A, Hallgren M, Bernspang B, Fisher AG. Client-centred occupational therapy for persons with mental retardation: implementation of an intervention programme in activities of daily living tasks. Scand J Occup Ther 2003;10:51-60.

343. Merritt BK, Fisher AG. Gender differences in the performance of activities of daily living. Arch Phys Med Rehabil 2003;84:1872-1877.

344. Sorensen LV, Axelsen U, Avlund K. Social participation and functional ability from age 75 to age 80. Scand J Occup Ther 2002;9:71-78.

345. Bedell GM, Dumas HM. Social participation of children and youth with acquired brain injuries discharged from inpatient rehabilitation: a follow-up study. Brain Inj 2004;18:65-82.

346. McColl MA, Charlifue S, Glass C, Savic G, Meehan M. International differences in ageing and spinal cord injury. Spinal Cord 2002;40:128-136.

347. Sliwa JA, Heinemann A, Semik P. Inpatient rehabilitation following burn injury: patient demographics and functional outcomes. Arch Phys Med Rehabil 2005;86:1920-1923.

348. Hammond FM, Grattan KD, Sasser H, Corrigan JD, Bushnik T, Zafonte RD. Long-term recovery course after traumatic brain injury: a comparison of the Functional Independence Measure and Disability Rating Scale. J Head Trauma Rehabil 8;16:318-329.

349. Johnston MV, Wood KD. Family help available: instrumental support and rehabilitation outcomes. J Rehabil Outcomes Meas 2000;4:16-18.

350. Forchheimer MB, Kalpakjian CZ, Tate DG. Gender differences in community integration after spinal cord injury. Top Spinal Cord Inj Rehabil 2004;10:163-174.

351. Heinemann AW, Linacre JM, Wright BD, Hamilton BB, Granger C. Prediction of rehabilitation outcomes with disability measures. Arch Phys Med Rehabil 1994;75:133-143.

352. Meythaler JM, DeVivo MJ, Braswell WC. Rehabilitation outcomes of patients who have developed Guillain-Barre syndrome. Am J Phys Med Rehabil 1997;76:411-419.

353. Werner RA, Kessler S. Effectiveness of an intensive outpatient rehabilitation program for postacute stroke patients. Am J Phys Med Rehabil 1996;75:114-120.

354. Hicken BL, Putzke JD, Novack T, Sherer M, Richards JS. Life satisfaction following spinal cord and traumatic brain injury: a comparative study. J Rehabil Res Dev 2002;39:359-366.

355. New PW. Functional outcomes and disability after nontraumatic spinal cord injury rehabilitation: Results from a retrospective study. Arch Phys Med Rehabil 2005;86:250-261.

356. Malec JF. Impact of comprehensive day treatment on societal participation for persons with acquired brain injury. Arch Phys Med Rehabil 2001;82:885-895.

357. Granger CV, Ottenbacher KJ, Baker JG, Sehgal A. Reliablity of a brief outpatient functional otucome assessment measure. Am J Phys Med Rehabil 1995;74:469-475.

358. Page SJ, Garner C. Psychometric evaluation of the Motor Readiness Questionnaire for Stroke. Clin Rehabil 2005;19:531-537.

359. Jette AM, Keysor J, Coster W, Ni P, Haley S. Beyond function: predicting participation in a rehabilitation cohort. Arch Phys Med Rehabil 2005;86:2087-2094.

360. Ostensjo S, Bjorbaekmo W, Carlberg EB, Vollestad NK. Assessment of everyday functioning in young children with disabilities: an ICF-based analysis of concepts and content of the Pediatric Evaluation of Disability Inventory (PEDI). Disabil Rehabil 2006;28:489-504.

361. Hart T, Millis S, Novack T, Englander J, Fidler-Sheppard R, Bell KR. The relationship between neuropsychologic function and level of caregiver supervision at 1 year after traumatic brain injury. Arch Phys Med Rehabil 2003;84:221-230.

362. Kolobe THA, Bulanda M, Susman L. Predicting motor outcome at preschool age for infants tested at 7, 30, 60, and 90 days after term age using the Test of Infant Motor Performance. Phys Ther 2004;84:1144-1156.

363. Rigby P, Gaik S. Stability of playfulness across environmental settings: a pilot study. Phys Occup Ther Pediatr 2007;27:27-43.

364. Hesse S, Werner C, Pohl M, Rueckriem S, Mehrholz J, Lingnau ML. Computerized arm training improves the motor control of the severely affected arm after stroke: a single-blinded randomized trial in two centers. Stroke 2005;36:1960-1966.

365. Linacre JM. Optimizing rating scale category effectiveness. J Appl Meas 2002;3:85-106.

366. White CL, Mayo N, Hanley JA, Wood-Dauphinee S. Evolution of the caregiving experience in the initial 2 years following stroke. Res Nurs Health 2003;26:177-189.

367. White CL, Lauzon S, Yaffe MJ, Wood-Dauphinee S. Toward a model of quality of life for family caregivers of stroke survivors. Qual Life Res 2004;13:625-638.

368. Mayo NE, Wood-Dauphinee S, Cote R, Durcan L, Carlton J. Activity, participation, and quality of life six months post-stroke. Arch Phys Med Rehab 2002;83:1035-1042.

369. Mayo NE, Wood-Dauphinee S, Cote R, Gayton D, Carlton J, Buttery J, Tamblyn R. There's no place like home: an evaluation of early supported discharge for stroke. Stroke 2000;31:1016-1023.

370. Teng J, Mayo NE, Latimer E, Hanley J, Wood-Dauphinee S, Cote R, Scott S. Costs and Caregiver Consequences of Early Supported Discharge for Stroke Patients. Stroke 2003;34:528-536.

371. Mayo NE, Nadeau L, Ahmed S, White CL, Grad R, Huang A, Yaffe M, Wood-Dauphinee S. Bridging the Gap: The Effectiveness of Teaming a Stroke Co-ordinator with Patient's Personal Physician on the Outcome of Stroke. Age Ageing. Accepted for publication 2007.

372. Jette AM. Toward a common language for function, disability, and health. Phys Ther 2006;86:726-734.

373. Berk, RA. Standard Setting: the next generation (Where few pshychometricians have gone before!). Appl Meas Educ 1996;9:215-235.

374. Tabachnick BG, Fidell LS. Using multivariate statistics. Boston: Allyn and Bacon, 2001.

375. Ferguson E, Cox T. Exploratory factor analysis: a users' guide. International Journal of Selection and Assessment 1993;1:84-94.

376. Franklin SB, Gibson DJ, Robertson PA, Pohlmann JT, Fralish JS. Parallel Analysis - A Method for Determining Significant Principal Components. J Veg Sci 1995;6:99-106.

377. Rojas Tejada AJ, Gonzalez GA, Padilla Garcia JL, Perez MC. Two strategies for fitting real data to Rasch polytomous models. J Appl Meas 2002;3:129-145.

378. Wise SL., Kingsbury GG. Practical Issues in Developing and Maintaining a Computerized Adaptive Testing Program. Psicologica 2000;21:135-155.

379. Smith EV, Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002;3:205-231.

380. Steinberg L, Thissen D, Wainer H. Computerized Adaptive Testing: A primer. Hillsdale, NJ: Lawrence Erlbaum Associates, 2000.

381. Lindmark B, Hamrin E. Evaluation of functional capacity after stroke as a basis for active intervention. Presentation of a modified chart for motor capacity assessment and its reliability. Scand J Rehabil Med 1988;20:103-109.

382. Johnstone M. Restoration of motor function in the stroke patient. Edinburgh: Chruchill Livingstone, 1983:14-17.

383. Mudie MH, Matyas TA. Upper extremity retraining following stroke: Effects of bilateral practice. J Neurol Rehabil 1996;10:167-184.

384. Nelson DL, Konosky K, Fleharty K, Webb R, Newer K, Hazboun VP, Fontane C, Licht BC. The effects of an occupationally embedded exercise on bilaterally assisted supination in persons with hemiplegia. Am J Occup Ther 1996;50:639-646.

385. Mayo NE. Epidemiology and Recovery of Stroke. Physical Medicine and Rehabilitation: State of the Art Reviews 1998;12:355-366.

386. Wright BD, Mok M. Rasch models overview. J Appl Meas 2000;1:83-106.

387. Cote R, Hachinski VC, Shurvell BL, Norris JW, Wolfson C. The Canadian Neurological Scale: a preliminary study in acute stroke. Stroke 1986;17:731-737.

388. Linacre JM. Sample Size and Item Calibration Stability. Rasch Measurement Transactions 1994;7:328.

389. Wright BD, Tennant A. Sample Size Again. Rasch Measurement Transactions 1996;9:468.

390. Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to their Development and Use. Oxford: Oxford University Press, 1995:231.

391. Messick S. Validity. In Linn R.L., editor. Educational Measurement. Phoenix: Oryx Press, 1993:13-103.

392. Wright BD, Stone MH. Best Test Design: Rasch Measurement. Chicago: MESA Press, 1979.

393. Schumacker RE. Rasch measurement using dichotomous scoring. J Appl Meas 2004;5:328-349.

394. Hattie J. Methodology Review: Assessing Unidimensionality of Tests and Items. Appl Psychol Meas 1985;9:139-164.

395. Smith EV, Jr. Metric development and score reporting in Rasch measurement. J Appl Meas 2000;1:303-326.

396. Muthen BO, Muthen L. Mplus User's guide. Los Angeles, CA: 1988.

397. Yu C.Y. Evaluating cutoff criteria of model fit indices for latent variable models with binary outcomes. 2002. University of California, Los Angeles.

398. Penfield RD. Unique properties of Rasch model item information functions. J Appl Meas 2005;6:355-365.

399. Stavem K, Lossius M, Ronning OM. Reliability and validity of the Canadian Neurological Scale in retrospective assessment of initial stroke severity. Cerebrovasc Dis 2003;16:286-291.

400. Waugh RF, Chapman ES. An analysis of dimensionality using factor analysis (true-score theory) and Rasch measurement: what is the difference? Which method is better? J Appl Meas 2005;6:80-99.

401. Reckase MD. Unifactor Latent Trait Models Applied to Multifactor Tests. Journal of Educational Statistics 1979;4:207-230.

402. Andrich D. Rasch Models for Ordered Response Categories. Encyclopedia of Statistics in Behavioral Science 2005;4:1698-1707.

403. Wilson M. On Choosing a Model for Measuring. In Smith EVJ, Smith RM, editors. Introduction to Rasch Measurement. Maple Grove: JAM Press, 2004:123-142.

404. Haley SM, Andres PL, Coster WJ, Kosinski M, Ni P, Jette AM. Short-form activity measure for post-acute care. Arch Phys Med Rehabil 2004;85:649-660.

405. Wright BD. Local Dependency, Correlations and Principal Components. Rasch Measurement Transactions 1996;10:509-511.

406. Wright BD, Masters GN. Rating scale analysis. Chicago: MESA Press, 1982.

407. Andrich D, de Jong JHAL, Sheridan BE. Diagnostic Opportunities with the Rasch model for ordered response categories. In Rost J, Langeheine R, editors. Applications of Latent Trait and Latent Class Models in the Social Sciences. New York: Waxmann, 1997:59-70.

408. Mallinson T, Stelmack J, Velozo C. A comparison of the separation ratio and coefficient alpha in the creation of minimum item sets. Med Care 2004;42:I17-I24.

409. Masters GN. Item Discrimination: When more is worse. Journal of Educational Measurement 1988;25:15-29.

410. Smith RM. Detecting item bias with the Rasch model. J Appl Meas 2004;5:430-449.

411. Andrich D. A general form of Rasch's extended logistic model for partial credit scoring. Applied Measurement in Education 1988;1:363-378.

412. Andrich D. An index of person separation in latent trait theory, the traditional KR.20 index, and the Guttman scale response pattern. Education Research and Perspectives 1982;9:95-104.

413. Rigdon EE, Ferguson CE. The Performance of the Polychoric Correlation-Coefficient and Selected Fitting Functions in Confirmatory Factor-Analysis with Ordinal Data
1. Journal of Marketing Research 1991;28:491-497.

414. Coenders G, Saris WE. Categorization and measurement quality. The choice between Pearson and Polychoric correlations. In Saris WE Munnich A, editor. *The Multitrait-Multimethod Approach to Evaluate Measurement Instruments.* Budapest: Eötvös University Press, 1995:125-144.

APPENDICES

APPENDIX 1: CONSENT FORM: The Effectiveness of Rehabilitation Therapy in Stroke

APPENDIX 2: CERTIFICAT OF ETHICAL APPROVAL

APPENDIX 3: AGREEMENT OF REPRODUCTION – MANUSCRIPT 1

APPENDIX 4: LIST OF STUDY MEASURES

APPENDIX 5: Model, item and person quality criteria for Rasch measures

APPENDIX 6: ADDITIONAL STATISTICAL CONSIDERATIONS

APPENDIX 1

PATIENT CONSENT FORM

Neurology Department
Royal Victoria Hospital
McGill University Health Centre

Title of the Study: **The Effectiveness of Rehabilitation Therapy in Stroke**

Clinical Contact Person: **Dr. Allen Huang (514-842-1231 ext. 4678 or 5704)**

Introduction: Researchers at this hospital and at McGill University are conducting a study to evaluate two rehabilitation programs that are specially designed for persons following a stroke. One program aims to improve arm and hand function, while the other aims to improve walking ability. The therapy provided in each of these two programs is a special type of therapy that focuses on the performance of functional tasks. This therapy is not provided in a standard fashion by rehabilitation professionals as its effectiveness has not yet been demonstrated in research. This is the goal of the study we are inviting you to enter. We realize that you may be involved in other studies. Your participation in this study will not affect your involvement in the other studies.

Time of Entry into the Study: We are asking if you would like to participate in this study. If you agree, then you would enter the study once you have completed formal rehabilitation therapy. For example, the situation described below that applies to you indicates when you would begin to participate:

1) If you have been discharged home from an acute-care hospital and you *are not receiving* physical or occupational therapy, then you would begin participation in this study three to four weeks after arriving home.

2) If you have been discharged home and you *are receiving* physical or occupational therapy as an out-patient, then you would begin participation in this study on completion of your therapy.

3) If you have been discharged to an in-patient rehabilitation centre where you *are receiving* physical or occupational therapy, then you would begin participation in this study on completion of your therapy.

Schedule of Evaluations: On entry into the study, you will undergo a baseline evaluation. The evaluations will be performed by a trained health professional who will assess your balance, how well you move your arms and legs, and how well you can do activities like walking, and climbing stairs. We will also ask you questions about how you feel about your health, and what you are able to do at home. Following this baseline evaluation, you will participate in one of the rehabilitation programs for six weeks. You will be re-evaluated on completion of this program, and then again six months later. The evaluations will be performed at the Richardson Hospital Centre.

In addition to these evaluations, we need to obtain some basic information about your medical history and your stroke from your medical chart.

Description of the Rehabilitation Programs: After your baseline evaluation, you will be assigned to one of the two rehabilitation programs by a random procedure (like flipping a coin). In other words, you will have a 50% chance of being assigned to one program or the other. If you are assigned to program A, you will perform functional, challenging, and creative tasks that involve coordination and strength of the arms and hands. You will also learn to perform activities on a computer. If you are assigned to program B, you will perform mobility tasks such as standing up from chairs of different heights, walking forwards, backwards, over obstacles, and up and down ramps and stairs. You will also participate in endurance training. In both programs, the exercises will be tailored to the level of ability of the individual.

As a participant in the rehabilitation program you will attend three sessions of exercise a week at the Richardson Hospital Centre. Each exercise session will last approximately one hour and a half. The therapist who designs your rehabilitation program may need to perform a few additional tests during your initial visits.

Transportation and Parking: The Richardson Hospital Centre is located at 5425 Bessborough Avenue in Montreal near the corner of Côte-St-Luc Road and Cavendish Road. Free visitors parking is available at the centre as well as on the quiet surrounding streets. While we cannot directly pay for transportation, we will make arrangements for car pooling, adapted transport, or taxi services as needed.

Participation and Confidentiality: Participation is voluntary. You may refuse to participate or withdraw from the study at any time without this having an effect on your health care. All of the information that we obtain from you will be kept strictly confidential. The data will be kept in a locked filing cabinet in the investigator's office. You will be assigned a study number and this will be the only identifying mark that will appear on your results. The results of the study will be published in scientific journals but your data will appear as numbers in statistical summaries.

Risks: We do not envision that the therapy provided through this study will cause you any harm. There may be a potential risk of falling for participants in the program that focuses on walking-related tasks. For this reason, you will be provided with physical assistance to walk when necessary to ensure your physical safety. The exercises that you will undertake will be performed at your own pace. All activities will be supervised so that if you do not feel well, or if you are anxious about your health, the appropriate action will be taken.

Benefits: The study offers you the opportunity to receive further rehabilitation therapy at a time when such services are no longer being provided by the public health care system. The results of this study will help us to identify which rehabilitation programs are most effective in improving physical function in persons who have been discharged home and who are no longer receiving therapy after stroke.

Contact Numbers: If you have any questions about the research, please contact Dr. Nancy Mayo, the principal investigator, or Nancy Salbach, the study coordinator, at (514) 842-1231 ext. 6906. If you have any questions concerning your rights as a participant in this study, you may contact the Ombudsman at (514) 842-1231 ext. 5655.

In signing this consent form, you recognize that the study has been explained to you and that you understand the study. You also agree that you have had the opportunity to ask questions, and that you are satisfied with the responses.

Declaration of the Participant: I understand what is expected of me and I freely and voluntarily agree to participate in this study entitled "The Effectiveness of Rehabilitation Therapy in Stroke".

A copy of this consent form has been given to the participant named below.

Name of Participant	*Signature of Participant*	*Date*
Name of Researcher	*Signature of Researcher*	*Date*

APPENDIX 2 - CERTIFICAT OF ETHICAL APPROVAL - Institutional Review Board of McGill University.

McGill

Faculty of Medicine
3655 Drummond Street
Montreal, QC H3G 1Y6
Fax: (514) 398-3595

Faculté de médecine
3655, rue Drummond
Montréal, QC, H3G 1Y6
Télécopieur: (514) 398-3595

CERTIFICATION OF ETHICAL ACCEPTABILITY FOR RESEARCH INVOLVING HUMAN SUBJECTS

The Faculty of Medicine Institutional Review Board consisting of:

LAWRENCE HUTCHISON, MD

SHARI BAUM, PHD

HAROLD FRANK, MD

NEIL MACDONALD, MD

ROBERTA PALMOUR, PHD

HARVEY SIGMAN, MD

MICHAEL THIRLWELL, MD

SALLY TINGLEY, BA

has examined the research project **A06-M10-99** entitled **"The Effectiveness of Rehabilitation Therapy in Stroke"**

as proposed by: Dr. Nancy E. Mayo (Applicant) to ______________ (Granting Agency, if any)

and consider the experimental procedures to be acceptable on ethical grounds for research involving human subjects.

June 14, 1999 — Date

Chair, IRB

Dean of Faculty

Institutional Review Board Assurance Number: M-1458

APPENDIX 3: Agreement of Reproduction – MANUSCRIPT 1

Notes for Contributors

1. The Journal's policy is to acquire copyright in all contributions. There are two reasons for this: (a) ownership of copyright by one central organisation tends to make it easier to maintain effective international protection against unauthorised use; (b) it also allows for requests from third parties to reprint or reproduce a contribution, or part of it, to be handled in accordance with a general policy which is sensitive both to any relevant changes in international copyright legislation and to the general desirability of encouraging the dissemination of knowledge.

2. Hodder Arnold co-operates in various licensing schemes which allow organisations to copy material within agreed restrains (e.g. the CLA in the UK and the CCC in the USA).

3. All contributors retain the rights to reproduce their paper for their own purposes provided no sale is involved. The author may post preprint electronic files on their own website for personal or professional use, on the author's institution's internal network or on a secure external site at the author's institution provided the article is not for resale. (The final typeset version of the paper may not be posted on any external website.)

4. All contributors retain the rights to forward an electronic preprint of their paper to a colleague for the colleague's personal or professional use.

5. All contributors retain the rights to reprint their paper in any volume of which they are editor or author. Permission will automatically be given to the publisher of such a volume, subject to the normal acknowledgement.

6. It is understood that in some cases copyrights will be held by the contributor's employer. If so, Hodder Arnold requires non-exclusive permission to deal with requests from third parties, on the understanding that any requests it receives will be handled in accordance with paragraph 3.

APPENDIX 4: LIST OF STUDY MEASURES

Tests and indices of impairment

- Box and Block Test (BBT)
- Canadian Neurological Scale (CNS)
- Chedoke McMaster Stroke Assessment Impairment Inventory
- Grip Strength
- Nine Hole Peg Test (NHPT)
- Rankin Index
- Upper Extremity Subscale of The STroke REhabilitation Assessment of Movement (STREAM)

Tests and indices of activity limitation (capacity)

- Frenchay Arm Test
- Test Évaluant la Performance des Membres Supérieurs des Personnes Agées (TEMPA)

Indices of the Amount of Use of the Upper Extremity

- Barthel Index
- Older Americans Resources and Services Scale-Instrumental Activities of Daily Living (OARS-IADL)
- Preference Based Stroke Index (PBSI)
- Reintegration to Normal Living Index (RNL)

Indices of Health-Related Quality of Life

- EUROQOL EQ-5D
- Health Utility Index
- Medical Outcomes Study 36-Item Short form Questionnaire (SF-36)
- Stroke Impact Scale (SIS)

The Box and Block Test

Instructions/Data Sheet

The evaluator is sitting in front of the subject. The subject is permitted to have a trial period of 15 seconds. At the signal, the subject is asked to take the blocks, one by one, from the compartment on the side of the hand being evaluated, take them to the other side of the box and release them. Start the test with the dominant hand. Count the number of blocks transferred in 60 seconds.

If the subject picks up two blocks at a time, they are counted as one. If the block is dropped on the table or floor after it is carried across the box, it is still counted but if it is tossed across without the fingertips crossing the partition, it will not be counted.

Number of blocks in 60 seconds - right hand __________

Number of blocks in 60 seconds - left hand __________

Le Test «Box and Block»

Instructions / Formulaire des données

L'évaluateur est assis en face du sujet. Le sujet a droit à une période d'essai de 15 secondes. Au signal, le sujet doit prendre les blocs, un par un, d'un compartiment situé du côté de la main évaluée, les transporter et les relâcher dans l'autre compartiment. Commencez le test avec la main dominante. Comptez le nombre de blocs transféré dans un délai de 60 secondes.

Si le sujet prends deux blocs à la fois, ils ne compterons que pour un. Si le sujet échappe un bloc sur la table ou par terre après l'avoir traverser le l'autre côté de la boîte, il sera compté mais si le bloc est lancez sans avoir les bouts des doigts ait traversé la séparation du milieu, il ne sera pas compté.

Nombre de blocs en 60 secondes – main droite __________

Nombre de blocs en 60 secondes – main gauche __________

The Canadian Neurological Scale

Section A - Mentation

1) Level of Consciousness

Alert: Normal consciousness.

Drowsy: Patient, when stimulated verbally, remains awake and alert for a short period of time but tends to doze even when examined.

2) Orientation

Oriented: Patient is oriented to both place (i.e., city *or* hospital) *and* to time (i.e., patient must be at least correct within 2 weeks). If early in the month (i.e., first 3 days), previous month is acceptable. Speech can be dysarthric (mispronounced or slurred) but intelligible.

Disoriented or non-applicable: If, for any reason, patient cannot answer the preceding questions on orientation (i.e., does not know answer, gives wrong answer, answers only partially, cannot express himself either by lack of words or unintelligible speech or finally ignores questions).

3) Speech (Language and Pronunciation)

a) Receptive Language

Patient is asked: (i) Close your eyes.
(ii) Open your mouth.
(iii) Point to the ceiling.

- Repeat twice if necessary.
- If patient obeys 3 commands, continue to b) Expressive Language.
- If patient obeys only 2 or less commands, score *receptive defect* in Speech Scale, and then proceed directly to motor function testing.

b) Expressive Language

- Objects needed: pencil, key, watch.
- In this section, pay special attention not only to answer but also to word pronunciation (i.e., dysarthria or slurred speech).

(i) Ask patient to name each object. Make sure patients see objects.
- If patient names only 2 or less of the objects, patient is scored *expressive defect* in Speech Scale.
- If patient names correctly 3 objects, proceed to (ii) below.

(ii) Ask the patient the following questions: What do you do with a pencil?
What do you do with a key?
What do you do with a watch?
- If patient answers correctly 3 questions, he/she is scored *normal speech*.
- If patient answers only 2 or less questions, he/she is scored *expressive defect* in Speech Scale.

n.b. The above scoring system relates to language only. Problems with pronunciation of words (i.e., dysarthria or slurred speech) is graded directly on Speech Scale below.

- Patient should always be scored according to worst speech deficit (i.e., language score or mispronunciation).
- Do not mimic commands in Section a) on Receptive Language.

Speech Scale

Normal Speech:	Answers all commands and questions in speech section; patient can have slurred speech (dysarthria) but still intelligible.
Expressive Defect:	Patient obeys command in receptive language section but makes one or more errors in section on expressive language and/or mispronunciation of words (slurred speech), with speech totally or partially non-intelligible (severe dysarthria).
Receptive Defect:	Patient obeys only 2 or less commands in section on receptive language.

Section A1 – Motor, No Receptive Deficit

This section to be used if patient does not have comprehension problems (i.e., normal speech or expressive defect only).

Motor Function

When evaluating strength and range of motion in limbs, always submit both limbs to same testing (i.e., apply same resistance at same position bilaterally).

4) Face

Test: Ask patient to show teeth or gums.

No weakness:	Symmetrical grin, no asymmetry in smile.
Weakness:	Facial asymmetry. One corner of mouth lower than other, either at rest or while showing teeth.

5) Upper Limb (Proximal)

- Patient should be tested in sitting position if possible. Test: Abduction arms (to 90°).
- If patient lying in bed: Test: Elevate arms to approximately 45° to 90°.
- Strength in both arms tested simultaneously.
- Resistance applied at midpoint between shoulder and elbow at all times.

6) Upper Limb (Distal)

- Patient tested in sitting or lying position, arms elevated.
  Test: Patient asked to make fists and to extend wrists.
- Compare range of movement in both wrists simultaneously.

- If full range of extension in both wrists, proceed to test strength by applying resistance separately to both fists while stabilizing patient's arm firmly.

7) **Lower Limb (Proximal)**
Test: Hip flexion – Ask patient to flex thighs toward trunk with knees flexed at 90°. Movement in both thighs tested separately.

8) **Lower Limb (Distal)**
Test: Dorsiflexion foot – Ask patient to point toes and foot upward. Compare both feet simultaneously (i.e., complete or partial movement).

In both **7)** and **8)**: Patient lying in bed for testing should always be scored according to worst deficit.

Apply resistance alternately to each thigh and foot after the full movement has been leted to test strength.

For sections **5)** through **8)**:

No weakness:	No detectable weakness.
Mild weakness:	Normal range of motion against gravity, but succumbs to resistance by observer either partially or totally.
Significant weakness:	Cannot completely overcome gravity in range of motion (i.e., partial movement).
Total weakness:	Absence of motion in movement tested or only contraction of muscles without actual movement of limb.

Section A2 – Motor, Receptive Deficit
This section to be used for patients *with* comprehension problems (i.e., receptive defect in Speech Scale).

Motor function in this section can be monitored in *one* of two ways:
(i) The ability of the patient to maintain a *fixed posture* in upper or lower limbs for a few seconds (3 – 5 seconds). The observer will alternately place the limbs in the desired position.

Upper Limbs:	Place arms outstretched at 90° in front of patient.
Lower Limbs:	Flexion of thighs with knees flexed at 90°.
Facial Power:	Have patient mimic your own grin. If patient does not cooperate then one proceeds to:

(ii) Comparison of motor response to a noxious stimuli (i.e., pressure on nailbed of fingers or toes alternately with a pencil). Facial response (grimacing) to pain is tested by applying pressure on sternum.

4) Face (grimacing)
Symmetrical
Asymmetrical (note side)

5) Upper Limbs

Equal motor response:	Patient *can* maintain the fixed posture equally in both upper limbs for a few seconds or withdraws equally on both sides to pain.
Unequal motor response:	Patient *cannot* maintain equally on both sides the fixed posture, weakness is noted on one side or there is an unequal withdrawal to pain. Note side where withdrawal not as brisk.

6) Lower Limbs

Equal motor response:	Patient *can* maintain the fixed posture equally in both lower limbs for a few seconds or withdraws equally on both sides to pain.
Unequal motor response:	Patient *cannot* maintain equally on both sides the fixed posture, weakness is noted on one side or there is an unequal withdrawal to pain. Note side where withdrawal not as brisk.

Chedoke-McMaster Stroke Assessment
SCORE FORM Page 1 of 4
IMPAIRMENT INVENTORY: SHOULDER PAIN AND POSTURAL CONTROL

POSTURAL CONTROL: Start at Stage 4. Starting position is indicated beside the item or underlined. No support is permitted.
Place an X in the box of each task that is accomplished. Score the highest Stage in which the client achieves at least two Xs.

SHOULDER PAIN

1	☐	constant, severe arm and shoulder pain with pain pathology in more than just the shoulder
2	☐	intermittent, severe arm and shoulder pain with pain pathology in more than just the shoulder
3	☐	constant shoulder pain with pain pathology in just the shoulder
4	☐	intermittent shoulder pain with pain pathology in just the shoulder
5	☐	shoulder pain is noted during testing, but the functional activities that the client normally performs are not affected by the pain
6	☐	no shoulder pain, but at least one prognostic indicator is present • Arm Stage 1 or 2 • Scapula malaligned • Loss of range of shoulder movt - flexion/abduction < 90° or external rotation < 60°
7	☐	shoulder pain and prognostic indicators are absent

☐ **STAGE OF SHOULDER PAIN**

POSTURAL CONTROL

1		☐	not yet Stage 2
2	Supine	☐	facilitated log roll to side lying
	Side lying	☐	resistance to trunk rotation
	Sit	☐	static righting with facilitation
3	Supine	☐	log roll to side lying
	Sit	☐	move forward and backward
	Stand	☐	remain upright 5 sec
4	Supine	☐	segmental rolling to side lying
	Sit	☐	static righting
	Sit	☐	stand
5	Sit	☐	dynamic righting side to side, feet on floor
	Sit	☐	stand with equal weight bearing
	Stand	☐	step forward onto weak foot, transfer weight
6	Sit	☐	dynamic righting backward and sideways with displacement, feet off floor
	Stand	☐	on weak leg, 5 seconds ☐ sec
	Stand	☐	sideways braiding 2 m
7	Stand	☐	on weak leg: abduction of strong leg
	Stand	☐	tandem walking 2 m in 5 sec
	Stand	☐	walk on toes 2 m

☐ **STAGE OF POSTURAL CONTROL**

Chedoke-McMaster Stroke Assessment
SCORE FORM Page 2 of 4
IMPAIRMENT INVENTORY: STAGE OF RECOVERY OF ARM AND HAND

ARM and HAND: Start at Stage 3. Starting position: sitting with forearm in lap in a neutral position, wrist at 0° and fingers slightly flexed. Changes from this position are indicated by underlining. Place an X in the box of each task accomplished. Score the highest Stage in which the client achieves at least two Xs.

Stage	ARM	HAND
1	☐ not yet Stage 2	☐ not yet Stage 2
2	☐ resistance to passive shoulder abduction or elbow extension ☐ facilitated elbow extension ☐ facilitated elbow flexion	☐ positive Hoffman ☐ resistance to passive wrist or finger extension ☐ facilitated finger flexion
3	☐ touch opposite knee ☐ touch chin ☐ shoulder shrugging > ½ range	☐ wrist extension > ½ range ☐ finger/wrist flexion > ½ range ☐ supination, thumb in extension: thumb to index finger
4	☐ extension synergy, then flexion synergy ☐ shoulder flexion to 90° ☐ elbow at side, 90° flexion: supination, then pronation	☐ finger extension, then flexion ☐ thumb extension > ½ range, then lateral prehension ☐ finger flexion with lateral prehension
5	☐ flexion synergy, then extension synergy ☐ shoulder abduction to 90° with pronation ☐ shoulder flexion to 90°: pronation then supination	☐ finger flexion, then extension ☐ pronation: finger abduction ☐ hand unsupported: opposition of thumb to little finger
6	☐ hand from knee to forehead 5 x in 5 sec. ☐ shoulder flexion to 90°: trace a figure 8 ☐ arm resting at side of body: raise arm overhead with full supination	☐ pronation: tap index finger 10 x in 5 sec ☐ pistol grip: pull trigger, then return ☐ pronation: wrist and finger extension with finger abduction
7	☐ clap hands overhead, then behind back 3 x in 5 sec ☐ shoulder flexion to 90°: scissor in front 3 x in 5 sec ☐ elbow at side, 90° flexion: resisted shoulder external rotation	☐ thumb to finger tips, then reverse 3 x in 12 sec ☐ bounce a ball 4 times in succession, then catch ☐ pour 250 ml. from 1 litre pitcher, then reverse
	☐ STAGE OF ARM	☐ STAGE OF HAND

Chedoke-McMaster Stroke Assessment
SCORE FORM Page 3 of 4
IMPAIRMENT INVENTORY: STAGE OF RECOVERY OF LEG AND FOOT

LEG: Start at Stage 4 with the client in crook lying. FOOT: Start at Stage 3 with the client in supine. Test position is beside the it or underlined. If not indicated, the position has not changed. Place an X in the box of each task accomplished. Score the highest in which the client achieves at least two Xs. For "standing" test items, light support may be provided but weight bearing through hand is not allowed. Shoes and socks off.

Stage	Position	LEG	Stage	Position	FOOT
1		☐ not yet Stage 2	1		☐ not yet Stage 2
2	Crook lying	☐ resistance to passive hip or knee flexion	2	Crook lying	☐ resistance to passive dorsiflexion
		☐ facilitated hip flexion			☐ facilitated dorsiflexion or toe xtension
		☐ facilitated extension			☐ facilitated plantarflexion
3		☐ abduction: adduction to neutral	3	Supine	☐ plantarflexion > ½ range
		☐ hip flexion to 90°		Sit	☐ some dorsiflexion
		☐ full extension			☐ extension of toes
4	Sit	☐ hip flexion to 90° then extension synergy	4		☐ some eversion
		☐ bridging hip with equal weightbearing			☐ inversion
		☐ knee flexion beyond 100°			☐ legs crossed: dorsiflexion, then plantarfk
5	Crook lying	☐ extension synergy, then flexion synergy	5		☐ legs crossed: toe extension with ankle plantarflexion
	Sit	☐ raise thigh off bed			☐ sitting with knee extended: ankle plantarflexion, then dorsiflexion
	Stand	☐ hip extension with knee flexion		Stand	☐ heel on floor: eversion
6	Sit	☐ lift foot off floor 5 x in 5 sec.	6		☐ heel on floor: tap foot 5 x in 5 sec
	Stand	☐ full range internal rotation			☐ foot off floor: foot circumduction
		☐ trace a pattern: forward, side, back, return			☐ knee straight, heel off floor: eversion
7	Stand	☐ unsupported: rapid high stepping 10 x in 5 sec	7		☐ heel touching forward, then toe touching behind, repeat 5 x in 10 sec
		☐ unsupported: trace a pattern quickly; forward, side, back, reverse			☐ foot off floor: circumduction quickly, rev
		☐ on weak leg with support: hop on weak leg			☐ up on toes, then back on heels 5 x
		☐ STAGE OF LEG			☐ STAGE OF FOOT

Chedoke-McMaster Stroke Assessment
SCORE FORM Page 4 of 4
DISABILITY INVENTORY

	SCORING LEVELS	
NO HELPER	Independence	
	7 Complete Independence	(Timely, Safely)
	6 Modified Independence	(Device)
HELPER	Modified Dependence	
	5 Supervision	
	4 Minimal Assist	(Client = 75%)
	3 Moderate Assist	(Client = 50%)
	Complete Dependence	
	2 Maximal Assist	(Client = 25%)
	1 Total Assist	(Client = 0%)

		Score
1.	Supine to side lying on strong side	☐
2.	Supine to side lying on weak side	☐
3.	Side lying to long sitting through strong side	☐
4.	Side lying to sitting on side of the bed through strong side	☐
5.	Side lying to sitting on side of the bed through the weak side	☐
6.	Remain standing	☐
7.	Transfer to and from bed towards strong side	☐
8.	Transfer to and from bed towards weak side	☐
9.	Transfer up and down from floor and chair	☐
10.	Transfer up and down from floor and standing	☐
11.	Walk indoors - 25 meters	☐
12.	Walk outdoors, over rough ground, ramps, and curbs - 150 meters	☐
13.	Walk outdoors several blocks - 900 meters	☐
14.	Walk up and down stairs	☐
15.	Age appropriate walking distance for 2 minutes (2 Point Bonus)	☐

Distance ☐ meters **Total Score** ☐

Walking aids:

walker ☐
4 point cane ☐
1 point cane ☐
brace ☐

To score Bonus:
for age less than 70 years distance must be > 95 meters or greater
for age 70 years or greater distance must be > 85 meters or greater

Grip Strength

Instructions / Data Sheets

Subjects are seated on a standard height chair without armrests with their elbow at 90 degrees. Three grip strength measures of each hand are taken using the Jamar dynamometer. The highest score will be retained.

Right hand:	1) ________	Left hand:	1) ________
	2) ________		2) ________
	3) ________		3) ________

Force de Préhension

Instructions / Formulaire des données

Le sujet est assis sur une chaise de hauteur standard, sans appui-bras. Le coude est placé a 90 degrés. Trois mesure de chaque mains sont prises avec un dynamomètre Jamar. La mesure la plus haute sera retenue.

Main droite:	1) ________	Main gauche:	1) ________
	2) ________		2) ________
	3) ________		3) ________

Nine-Hole Peg Test

The board is placed in front of the subject. The dowels are placed beside the board on the side of the hand being evaluated. At the signal, the subject takes the 9 dowels, one by one, and places them in the holes. Once they are all in place, the subject takes them out, one by one, and puts them back on the table beside the board. The task is timed.

Time __________ seconds

Le Test «Nine-Hole Peg»

La planchette de bois est placée en face du sujet; les chevilles de bois sont dans un contenant adjacent à la planchette de bois (du côté de la main évaluée). Au signal, le sujet doit prendre les 9 cheville (une à la fois), les placer dans les trous de la planchette puis les retirer (une à la fois) et les remettre dans le contenant. Chronométrez la tâche.

Temps __________ secondes

The Rankin Index

Grade Description

0	No symptoms at all
1	No significant disability; despite symptoms; able to carry out all usual duties and activities
2	Slight disability; unable to carry out all previous activities but able to look after won affairs without assistance
3	Moderate disability; requiring some help, but able to walk without assistance
4	Moderately severe disability; unable to walk without assistance and unable to attend to own bodily needs without assistance
5	Severe disability; bedridden, incontinent and requiring constant nursing care and attention

STroke Rehabilitation Assessment of Movement

Data Sheet

Name: ______________________ Date: ____________

Hospital#: ______________________

Score	**SUPINE**
/2	1. **Protracts scapula in supine**
	"Lift your shoulder blade so that your hand moves towards the ceiling." Note: Therapist stabilizes arm with shoulder 90° flexed and elbow extended.
/2	2. **Extends elbow in supine** (starting with elbow fully flexed)
	"Lift your hands toward the ceiling, straightening your elbow as much as you can." Note: Therapist stabilizes arm with shoulder 90° flexed, strong associated shoulder extension and/or abduction = marked deviation (score 1a or 1c)
/2	3. **Flexes hip and knee in supine** (attains half crook lying)
	"Bend your hip and knee so that your foot rests flat on the bed."
/3	4. **Rolls onto side** (starting from supine)
	"Roll onto your side." Note: May roll onto either side; pulling with arms to turn over = aid (score 2).
/3	5. **Raises hips off bed in crook lying position** (bridging)
	"Lift your hips as high as you can." Note: Therapist must stabilize foot, but if knee pushes strongly enough into extension with bridging = marked deviation (score 1a or 1c); if requires aid (external or from therapist) to maintain knees in midline = aid (score 2).
/3	6. **Moves from lying supine to sitting** (with feet on the floor)
	"Sit up and place your feet on the floor." Note: may sit up to either side using any functional and safe method; longer than 20 seconds = marked deviation (score 1a or 1c); pulling up using bed rail or edge of plinth = aid (score 2).
	SITTING (feet supported; hands resting on pillow on lap for items 7-14)
/2	7. **Shrugs shoulders** (scapular elevation)
	"Shrug your shoulders as high as you can." Note: Both shoulders are shrugged simultaneously.
/2	8. **Raises hand to touch top of head**
	"Raise your hand to touch the top of your head."

/2	**9. Places hand on sacrum**
	"Reach behind your back and as far across toward the other side as you can."
/2	**10. Raises arm overhead to fullest elevation**
	"Reach your hand as high as you can towards the ceiling."
/2	**11. Supinates and pronates forearm** (elbow flexed at 90^{o}) "Keeping your elbow bent and close to your side, turn your forearm over so that your palm faces up, then turn your forearm over so that your palm faces down."
	Note: Movement in one direction only = partial movement (score 1a or 1b).
/2	**12. Closes hand from fully opened position** "Make a fist, keeping your thumb on the outside."
	Note: Must extend wrist slightly (wrist cocked) to obtain full marks.
/2	**13. Opens hand from fully closed position**
	"Now open your hand all the way."
/2	**14. Opposes thumb to index finger** (tip to tip)
	"Make a circle with your thumb and index finger."
/2	**15. Flexes hip in sitting**
	"Lift your knee as high as you can."
/2	**16. Extends knee in sitting**
	"Straighten your knee by lifting your foot up."
/2	**17. Flexes knee in sitting**
	"Slide your foot back as far as you can."Note: Start with affected foot forward (heel in line with toes of other foot).
/2	**18. Dorsiflexes ankle in sitting**
	"Keep your heel on the ground and lift your toes off the floor as far as you can."
/2	**19. Plantarflexes ankle in sitting**
	"Keep your toes on the ground and lift your heel off the floor as far as you can."
/2	**20. Extends knee and dorsiflexes ankle in sitting** "Straighten your knee as you bring your toes towards you."
	Note: Extension of the knee without dorsiflexion of ankle = partial movement (score 1a or 1b).
/3	**21. Rises to standing from sitting** "Stand up; try to take equal weight on both legs." Note: pushing up with hand(s) to stand = aid (score 2); asymmetry such as trunk lean, trendelenburg, hip retraction, or excessive flexion or extension of the affected knee = marked deviation (score 1a or 1c).

/3	*STANDING* **22. Maintains standing for 20 counts** *"Stand on the spot while I count to 20."*
/2	**STANDING** (holding onto a stable support to assist balance for items 23-25) **23. Abducts affected hip with knee extended** *"Keep your knee straight and your hips level, and raise your leg to the side."*
/2	**24. Flexes affected knee with hip extended** *"Keep your hip straight, bend your knee back and bring your heel towards your bottom."*
/2	**25. Dorsiflexes affected ankle with knee extended** *"Keep your heel on the ground and lift your toes off the floor as far as you can."*
/3	***Standing and Walking Activities*** **26. Places affected foot onto first step** (or stool 18 cm high) "Lift your foot and place it onto the first step (or stool) in front of you." Note: Returning the foot to the ground is not scored; use of handrail = aid (score 2).
/3	**27. Takes 3 steps backwards** (one and a half gait cycles) *"Take 3 average sized steps backwards, placing one foot behind the other."*
/3	**28. Takes 3 steps sideways to affected side** *"Take 3 average sized steps sideways towards your weak side."*
/3	**29. Walks 10 meters indoors** (on smooth, obstacle free surface) "Walk in a straight line over to … (a specified point 10 meters away)." Note: orthotic = aid (score 2); longer than 20 seconds = marked deviation (score 1c).
/3	**30. Walks down 3 stairs alternating feet** "Walk down 3 stairs; place only one foot at a time on each step if you can."Note: handrail = aid (score 2); non-alternating feet = marked deviation (score 1a or 1c).

Frenchay Arm Test

This test consists of five pass/fail tasks. One point is given for each task completed successfully. The subject begins each task by sitting at a table with their hands on their lap and is asked to use the affected arm/hand to:

1. Stabilize a ruler, while drawing a line with a pencil held in the other hand. To pass, the ruler must be held firmly.
2. Grasp a cylinder (12 mm diameter, 5 cm long), set it on its side approximately 15 cm from the table edge, lift it about to 30 cm and replace it without dropping it.
3. Pick up a glass of water (half-full) positioned about 15 to 30 cm from the edge of a table, drink some water and replace it without spilling any water.
4. Remove and replace a spring clothes peg from a 10 mm diameter dowel (15 cm long set in a 10 cm base, 15 to 30 cm from table edge) without dropping the peg or knock dowel over.
5. Comb hair (or imitate) across top, down the back and down each side of their head.

Total score __________

Test D'Evaluation Des Membres Supérieurs Chez Les Personnes Agées (TEMPA)

Examiner-Subject Positioning

The person being evaluated sits on a chair or armchair of standard height (44 cm ± 2.5 cm) or in her own wheelchair facing a table of regular height (76 cm ± 2.5), representing a normal, everyday situation. The therapist, with the score sheet and stopwatch in hand, sits beside the table at an angle of 90° to the subject. The equipment to be used is within reach.

Before each task, the subject puts her hands on the edge of the table waiting until the therapist gives the signal to begin. **The stopwatch is started as soon as the subject's hands leave the table (*not* on the word "Go").**

Equipment: Test material, table, chair, stopwatch, Jamar dynamometer.

General Instructions: ***For the therapist...***

Each task is preceded by specific instructions and a demonstration. Encourage the subject to perform a trial to ensure that the task is fully understood. For unilateral tasks, do the trial using the better side.

When the subject suffers from visual impairments, ensure that glasses are worn.

Unilateral tasks *begin* on the *dominant or more functional side* when there is a unilateral impairment.

Complete the upper part of the score sheet (name, age, main diagnosis, presence of visual impairments, dominance, wearing of corrective lenses, perceptual or cognitive impairment). Evaluate and record the overall passive range of motions of the upper extremities. All the tasks must be performed in the order in which they appear on the score sheet. Record scores at the end of each task, and note any specific scoring details in the "comments" column.

For the person being evaluated...

The therapist will say the following:

"I want to evaluate how you use your arms in some daily activities. I am going to ask you to do a few tasks and, before each task, I will demonstrate it to you. Before doing each task, you will have a chance to practice. I am going to time how long it takes you to do each task and, at the same time, I will be watching to see how you do it. I can help you if you need assistance. After each task, I will take a few seconds to write my observations on my sheet."

"Some of the activities must be done twice, once with the right hand and once with the left. The other tasks will be done using both hands together if you can do this. Finally, I will measure the strength of your hand and your muscular endurance with this instrument."

"Do you have any questions?"

Description of the Tasks

Numbers on the shelves indicate the exact position for placing the test material for each task. **When in doubt, record the lower score being considered for a task.**

1. **Pick up and move a jar** (unilateral task)

   Instructions

   "With your right (or left) hand, pick up this jar and put it down here, right in the middle of the shelf."
   "You may practice if you wish."
   "Are you ready?"
   "Keep your hands on the edge of the table until I say go."
   "Go."

   "Do the same thing with the other hand."

   Les consignes

   "Prenez avec votre main droite (ou gauche) le pot placé ici et venez le porter à cet endroit, juste au centre du plateau."
   "Vous pouvez pratiquer."
   "Êtes-vous prêt?"
   "Gardez vos mains sur le bord de la table jusqu'à mon signal."
   "Allez-y" ou "Quand vous êtes prêt, vous pouvez y aller".

   "Refaites la même chose avec l'autre main."

2. **Open a jar and remove a spoonful of coffee** (bilateral task)

   Instructions

   "Using whichever hand you prefer, pick up the jar of coffee that is here and open it. Then take the spoon that is in the cup and remove a spoonful of coffee. Put the spoonful of coffee into the cup, close the jar and put it back in its original place."
   "You may practice if you wish."
   "Are you ready?"
   "Keep your hands on the edge of the table until I say go."
   "Go."

   Les consignes

   "Prenez le pot de café qui est ici, avec la main que vous désirez. Vous devez l'ouvrir puis saisir la cuillère placée dans la tasse pour prendre une cuillerée de café. Vous devez ensuite verser cette cuillerée de café dans la tasse puis refermer le pot et le remettre à sa place."
   "Vous pouvez pratiquer."
   "Êtes-vous prêt?"

"Gardez vos mains sur le bord de la table jusqu'à mon signal."
"Allez-y" ou "Quand vous êtes prêt, vous pouvez y aller".

Note: After the task is completed, the therapist puts the coffee back in the jar.

Scoring

Both upper limbs are scored at the same time. One of limbs acts mainly and almost exclusively as a stabilizer. Record the better performance. The loss of function of the upper limb will be determined in the unilateral tasks.

3. **Pick up a pitcher and pour water into a glass** (unilateral task)

Note: Fill pitcher with 400 mL water. Place pitcher with the handle towards the right for right hand performance and towards the left for left hand performance. The glass is located on the side of the performing hand.

Instructions

"With this hand, pick up this pitcher that has been filled with water and fill this glass about three-quarters full. Then put the pitcher on the table, pick up the glass and touch your chin with it. Then put the glass back on the table."
"You may practice if you wish."
"Are you ready?"
"Go."

"Do the same thing with the other hand."

Les consignes

"Prenez ce pichet rempli d'eau avec cette main et remplissez ce verre aux trois quarts environ. Ensuite, vous déposez le pichet sur la table, prenez le verre et vous le portez à votre menton. Puis vous remettez le verre sur la table."
"Vous pouvez pratiquer."
"Êtes-vous prêt?"
"Gardez vos mains sur le bord de la table jusqu'à mon signal."
"Allez-y" ou "Quand vous êtes prêt, vous pouvez y aller".

"Refaites la même chose avec l'autre main."

4. **Unlock a lock and open a pill container** (bilateral task)

Note: The key notches are oriented to the same side as the dominant hand, this facilitates handling the key and opening the lock. The therapist requests that the key be grasped with a lateral pinch, thumb on top.

Instructions

"Pick up this key like this and open the cabinet like this, leaving the key in the lock. Pick up the pill container, open it, take out two pills and put them on the table. Then close the container and put it on the table."
"You may practice if you wish."
"Are you ready?"
"Go."

Les consignes

"Prenez cet clé comme ceci et ouvrez l'armoire de cette façon, laissez la clé dans la serrure et prenez le pot de pilules. Ouvrez-le et retirez deux comprimés que vous déposez sur la table où vous voulez. Puis, vous refermez le contenant et le placez aussi sur la table."
"Vous pouvez pratiquer."
"Êtes-vous prêt?"
"Gardez vos mains sur le bord de la table jusqu'à mon signal."
"Allez-y" ou "Quand vous êtes prêt, vous pouvez y aller".

Scoring

The cabinet on the upper shelf is removable. If the subject cannot reach it for any reason, the therapist can pull the cabinet out of its space and move it to the level of the lower shelf. This allows other aspects of the task to be evaluated. However, the subject will obtain a -2 on the functional rating.

5. **Write on an envelope and stick a stamp on it** (bilateral task)

Note: The subject may move the envelope to be more comfortable while writing. If the subject is illiterate, substitute their name for 'Bell Canada'.

Instructions

"Pick up this pen and write the words 'Bell Canada' in the middle of this envelope. Then take this stamp and stick it in the right-hand corner."
"You may practice on a sheet of paper or begin right away on the envelope."
"Are you ready?"
"Go."

Les consignes

"Prenez ce crayon et écrivez au centre de cette enveloppe les mots "Bell Canada". Par la suite, vous prenez ce timbre et vous le collez dans le coin habituel."
"Vous pouvez pratiquer sur une feuille ou y aller directement sur l'enveloppe."
"Êtes-vous prêt?"
"Gardez vos mains sur le bord de la table jusqu'à mon signal."
"Allez-y" ou "Quand vous êtes prêt, vous pouvez y aller".

Scoring

If you must stabilize the envelope while the subject is writing, a functional rating of -2 is obtained. If the subject is aphasic and cannot write but can do the rest of the task, a functional rating of -2 is also obtained.

6. **Tie a scarf around one's neck** (bilateral task)

Note: A dark blue scarf is folded in eight and centered on the lower shelf with the fringe on the left. The subject picks it up, unfolds it, wraps it around the neck and ties a simple knot. The neck must not bend during the task. The objective is that the task be performed using both upper limbs as symmetrically as possible.

Instructions

"Pick up the scarf, unfold it, put it around your neck and then tie it using a simple knot. Try to keep your head as soon as possible while you are doing it."
"You may practice if you wish."
"Are you ready?"
"Go."

Les consignes

"Prenez ce foulard, dépliez-le et mettez-le autour de votre cou, et attachez-le en faisant un simple noeud. Essayez de faire le tout en gardant la tête la plus droite possible."
"Vous pouvez pratiquer."
"Êtes-vous prêt?"
"Gardez vos mains sur le bord de la table jusqu'à mon signal."
"Allez-y" ou "Quand vous êtes prêt, vous pouvez y aller".

Scoring

- The functional rating is evaluated for the overall task. On the other hand, the Task Analysis dimensions are scored individually because of the symmetrical nature of the task.
- Refer to the section 'Scoring-specific cases' as needed.
- It is acceptable for the subject to bend the neck in a normal way.

7. **Shuffle and deal playing cards** (bilateral task)

Note: The subject picks up the cards one by one sliding them to the edge of the table (as is normally done when playing cards).

Instructions

"Pick up this deck of cards and remove the rubber band. Shuffle three times, like this, and deal five cards in front of you. Then put down the rest of the deck and pick up the five cards one by one."
"You may practice if you wish."
"Are you ready."
"Go."

Les consignes

"Prenez ce paquet de cartes et enlevez-en l'élastique. Coupez le paquet trois fois comme ceci et passez cinq cartes en avant de vous. Déposez ensuite le reste du paquet et ramassez les cinq cartes une par une."
"Vous pouvez pratiquer."
"Êtes-vous prêt?"
"Gardez vos mains sur le bord de la table jusqu'à mon signal."
"Allez-y" ou "Quand vous êtes prêt, vous pouvez y aller".

8. **Handle coins** (unilateral tasks)

Note: The coins are stacked from the smallest to the largest (the largest on top) on the lower shelf in the identified location. As explained in task no. 4, the cabinet in the upper shelf is removable. Therefore, if the subject cannot reach the slot to insert the coins, the therapist may remove the cabinet.

Instructions

"Using this hand, pick up these coins one by one beginning with the dollar and put them in this slot here."
"You may practice if you wish."
"Are you ready?"
"Go."

Les consignes

"Prenez une par une ces pièces de monnaie avec cette main en commençant avec le dollar et mettez-les dans la fente ici."
"Vous pouvez pratiquer."
"Êtes-vous prêt?"
"Gardez vos mains sur le bord de la table jusqu'à mon signal."
"Allez-y" ou "Quand vous êtes prêt, vous pouvez y aller".

"Refaites la même chose avec l'autre main."

9. **Pick up and move small objects** (unilateral tasks)

Instructions

"Pick up these objects one at a time with this hand and put them in the dish."
"You may practice if you wish."
"Are you ready?"
"Go."

"Do the same thing with the other hand."

Les consignes

"Prenez ces objets un par un avec cette main et mettez-les dans le pot."
"Vous pouvez pratiquer."
"Êtes-vous prêt?"
"Gardez vos mains sur le bord de la table jusqu'à mon signal."
"Allez-y" ou "Quand vous êtes prêt, vous pouvez y aller".

"Refaites la même chose avec l'autre main."

Materials

A glass dish, a safety pin, a flat toothpick, a black button, a bolt, a nail.

Scoring

All the objects must have been picked up, carried and put in the dish to obtain a score of 0 or -1 on the functional rating.

Grip Strength

Subject's Position

The subject sits on a chair without armrests and holds the Jamar dynamometer in hand; the evaluator supports the dynamometer slightly from beneath.

Upper Limb Position

Shoulder in adduction, neutral rotation, elbow bent at 90° and forearm in a neutral position. The wrist is slightly extended (0° to 30°). The handle of the dynamometer is on the *second position.*

Procedure

The subject must squeeze the handle of the Jamar dynamometer as hard as possible. The therapist gives as much verbal encouragement as desired. Three measurements are taken on each side, alternating right and left and starting with the dominant side, with a rest of about 20 seconds between two measurements on the same side.

Instructions

"I am going to measure the strength of your hands. I am going to measure each hand three times."
"I want you to hold the handle like this and squeeze as hard as you can."
"Are you ready? Squeeze as hard as you can."
"Harder!... Harder!... Relax."

Les consignes

"Je vais maintenant mesurer la force de vos mains. Je vais prendre trois mesures de chaque main."
"Prenez cet appareil dans votre main comme ceci et mettez votre bras dans cette position. Je vais soutenir légèrement l'appareil."
"À mon signal, vous serrez la poignée le plus fort que vous pouvez."
"Plus fort!... Plus fort!... Arrêtez."

The same is said for the other hand. Repeat the same procedure two more times.

Scoring

The results for each hand are recorded on the score sheet (in kg) and the average of the three measurements is calculated.

Relative Isometric Muscular Endurance

Relative isometric endurance may be defined as the subject's ability to maintain a force equivalent to 50% of maximum strength over a certain length of time.

Materials

A Jamar dynamometer, a stopwatch.

Subject's Position

The subject sits on a chair without armrests and holds the Jamar dynamometer in hand, with the evaluator supporting the dynamometer slightly.

Upper Limb Position

Shoulder is in adduction, neutral rotation, elbow bent at 90° and forearm in a neutral position. The wrist is slightly extended (0° to 30°).

Procedure

The average value obtained when measuring the prehension strength of each hand with the Jamar dynamometer is divided by two. The subject is asked to reach the resulting value and maintain it for as long as possible; the procedure is repeated with the other hand.

Instructions

"I want to measure the muscular endurance in your hands. Hold this instrument. I am going to ask you to squeeze the handle gently until the needle on the dial reaches a value that corresponds to half your strength. I want you to maintain this position as long as possible. I am going to time you."

Les consignes

"Je désire mesurer l'endurance des muscles de vos mains. Prenez cet appareil. Je vais vous demander de serrer doucement la poignée pour amener l'aiguille du cadran à une valeur qui correspond à la moitié de votre force. Je désire que vous mainteniez cette position le plus longtemps possible. Je vais chronométrer ce temps."

Scoring

- The length of time the force the maintained is timed with the stopwatch.
- Start with the dominant or better hand.
- The therapist must give the subject regular instructions (e.g. a little harder, not so hard) about the force he/she must maintain (Note: the subject cannot see the dial because of the design of the instrument).
- The therapist stops the stopwatch when, in spit of encouragement, the needle on the dial falls to less than 10% of the desired force or when the subject suddenly stops trying.
- The opposite end of the movable needle can be used as a guide by placing it at the required value.
- The test is repeated with the other hand.

Task Scoring System

During or immediately upon completion of each task, the therapist scores the performance obtained according to three measurement criteria: speed of execution, functional rating and task analysis.

Speed of Execution

Each task is timed with s stopwatch to the nearest tenth of a second, beginning as soon as the subject's hands leave the table and ending the moment the task is completed.

Functional Rating

The functional rating refers to the subject's independence in each of the tasks; it is measured using a four-level scale:

Score	Scale
0	The task is successfully completed, without hesitation or difficulty, as instructed or demonstrated.
-1	The task is executed completely, but with some hesitation or difficulty.
-2	The task is partially executed (more than 25%) or certain steps are executed with major difficulties necessitating repeated efforts. Part of the task may have had to be modified or needed assistance to make it achievable.
-3	The task cannot be performed to more than 25%.

Task Analysis:

1) Active Range of Motion

The subject's ability to reach the material and execute the movements required by the task; the quality of the movement is not taken into consideration.

Score	Scale
0	All the ranges required by the task are present.
-1	Certain parts of the task are difficult or compensated for because of a partial limitation in the active range of motion.
-2	Certain ranges of motions are very limited making it impossible to reach the material or execute part of the task.
-3	The majority of ranges are necessary to perform the task are very restricted, substantially compromising the accomplishment of the task.

2) Strength

The subject's ability to use the "task heavy" material (offering resistance other than gravity) without compensatory movements.

Score	Scale
0	The strength of the upper extremity(ies) is sufficient to do the task as described (at least against gravity and the resistance of the object).
-1	The strength of the upper extremity(ies) is diminished, contributing to the emergence of some compensatory movements.
-2	The strength of the upper extremity(ies) is greatly diminished and the loss cannot be easily compensated for by substitution movements.
-3	The strength of the upper extremity(ies) is too diminished to lift the material making the task impossible to execute or impossible to evaluate because it proved impossible to handle the material.

3) Precision of Gross Motor Movements

The subject's ability to execute precise, rapid and appropriate movements with the whole of the upper limb, excluding the hand.

Score	Scale
0	The unilateral or bilateral gross movements are precise, rapid and appropriate to the task.
-1	The gross motor movements are slightly uncoordinated or done slowly.
-2	The gross motor movements are imprecise, poorly directed or very slow.
-3	No precision in the gross motor movements in terms of achieving the task objectives.
	Cannot be evaluated

4) Prehension

The subject's ability to take different objects, small or large, with the hand or fingers using grasping and pinching actions appropriately.

Score	Scale
0	All the prehension patterns (grasps and pinches) required to perform the task are executed without apparent difficulty.
-1	All the objects can be grasped in spite of some prehension difficulties.
-2	Certain prehensions are impossible or very difficult and require several attempts.
-3	No prehension is possible in the activity.
	Cannot be evaluated.

5) Precision of Fine Motor Movements

The subject's ability to use both hands and fingers to execute precise, rapid movements that are well directed towards accomplishing the task.

Score	Scale
0	The movements of the hands and fingers are very precise and goal-directed. Manipulation of the objects is done normally.
-1	The precision of the fine motor movements is diminished or the movements are slow. Manipulation may be difficult but possible. There may be some slight trembling.
-2	Distal movements significantly lack precision. Objects are often dropped. There may be some substantial trembling.
-3	The fine motor movements are very imprecise or are impossible to accomplish. It becomes impossible to perform the fine motor movements desired, which are necessary to do the activity.
	Cannot be evaluated.

Calculation of Total Scores

When the nine tasks are completed and scored, the therapist adds vertically the scores on the functional rating as follows: all the scores obtained on the right unilateral tasks (0 to -12), left unilateral tasks (0 to -12) and bilateral tasks (0 to -15) and then the total of the scores on all the tasks (unilateral right + unilateral left + bilateral) (0 to -39). In a similar manner, the same additions are done for the five dimensions in the Task Analysis section.

See the sample score sheet with scores tallied at the end of this section.

Scoring – Specific Cases

1. If the task **is completed,** with or without difficulty, as instructed or demonstrated, the speed of execution is recorded and each of the other measurement criteria is scored according to performance.

2. If a **part of the task cannot be done** because of some limitation, whatever its nature, the therapist may give physical help or reduce the degree of difficulty of the task in such a way as to be able to observe a maximum number of dimensions. The subject can thus proceed with the task in spite of difficulties. However, the speed of execution cannot be counted in this instance. The subject will obtain a score of -2 on the functional rating and the task analysis dimensions will be scored according to the pre-established scale for each task.

Examples of help given or modifications

- steadying the material
- reducing the weight of an object (e.g. taking some water out of the pitcher)
- reducing the height (e.g. putting the material on the lower shelf instead of the upper shelf)

3. If there is **unequal function** of the upper extremities during bilateral tasks (unilateral paresis), the bilateral tasks are scored according to the functional global performance of the upper extremities. This type of task normally requires that the less functional or non dominant upper limb stabilize the material (asymmetrical tasks), except in the case of task no. 5 (tying a scarf) which is a symmetrical bilateral task.

   If the subject uses only one upper extremity in a bilateral task, a maximum score of -1 on the functional rating is obtained because the task presents some difficulty, however slight. Regarding the scoring of the dimensions in the Task Analysis section, you score what you see, namely, the performance of the active upper limb (the better side). In the left or right corner of the appropriate box, you can put a check mark to indicate that a single upper limb performed the task. The unilateral tasks should make it possible to quantify the dimensions for each upper limb individually, thus showing the non-performance of one of the upper limbs, if such is the case.

4. You may give verbal assistance throughout the task.

5. If the subject obtains -3 on the functional rating, it is often impossible to measure the dimensions in the Task Analysis section. In such cases, you score -3 on these dimensions.

6. If the subject is apraxic and perform movements irrelevant to the task, a maximum of -1 on the functional rating will be obtained.

7. Write all pertinent comments in the section provided for this purpose.

<u>TEMPA</u>: **UPPER EXTREMITY PERFORMANCE EVALUATION TEST FOR THE ELDERLY**

NAME : AGE : DIAGNOSIS : DATE :

DOMINANCE : VISUAL DISORDERS : NO ☐ YES ☐ EYE GLASSES : NO ☐ YES ☐

PERCEPTUAL OR COGNIITIVE IMPAIRMENT : NO ☐ YES ☐ NOT EVALUATED ☐

OVERALL PASSIVE RANGE OF MOTION :

TASKS	SPEED OF EXECU		FUNCTIONAL RA		TASK ANALYSIS										COMMENTS
	L	R	L	R	Active range of motion		Strength		Precision of gro motor movemen		Prehensions		Precision of fine motor movement s		
					L	R	L	R	L	R	L	R	L	R	
1. PICK UP AND MOVE A JAR															
2. OPEN A JAR AND REMOVE A SPOONFUL OF COFFEE															
3. PICK UP A PITCHER AND POUR WATER INTO A GLASS															
4. UNLOCK A LOCK AND OPEN A PILL CONTAINER															
5. WRITE ON AN ENVELOPE AND STICK A STAMP ON IT															
6. TIE A SCARF AROUND ONE'S NECK															
7. SHUFFLE AND DEAL PLAYING CARDS															
8. HANDLE COINS															
9. PICK UP AND MOVE SMALL OBJECTS															
UNILATERAL TASKS TOTAL SCORE															
BILATERAL TASKS TOTAL SCORE															
COMBINED TOTAL SCORE															

Barthel Index

1. Feeding

10 = Independent. The patient can feed himself a meal from a tray or table when someone places the food within his reach. He is able to put on an assistive device if required, cut up his food, use salt and pepper, spread butter, etc. He must be able to accomplish this in a reasonable time.

5 = Needs some assistance (with the tasks listed above).

0 = The patient can not meet the criteria as defined above.

2. Personal Hygiene

5 = Independent. The patient can wash his hands and face, comb hair and brush teeth. Male patients must be able to shave with any kind of razor but must be able to put in the blade, plug in the razor and get it from the drawer or cabinet by themselves. Female patients must be able to apply their own make up, if used, but do not need to be able to braid or style their hair.

0 = Needs some assistance.

0 = The patient can not meet the criteria as defined above.

3. Bathing

5 = Independent. The patient must be able to use a bathtub, a shower or take a complete sponge bath. He must be able to perform all the steps involved in any one of these tasks without another person being present.

0 = Needs some assistance.

0 = The patient can not meet the criteria as defined above.

4. Dressing and Undressing

10 = Independent. Patient is able to put on, remove and fasten all clothing and tie shoe laces (unless it is necessary to use adaptive aids for this). This includes putting on, removing and fastening corsets or braces when they are prescribed.

5 = Some help is necessary. The patient needs help in putting on, removing or fastening any clothing. He must do at least half the work himself and accomplish this within reasonable time. Women need not be scored on the use of a brassiere or girdle unless these are prescribed garments.

0 = The patient can not meet the criteria as defined above.

5. Getting On and Off the Toilet

10 = Independent. The patient is able to get on and off the toilet, fasten and unfasten clothes, prevent soiling of clothes and use toilet paper without help. A wall bar or any other stable object may be used for support if needed. If a bedpan is used, he must be able to place it on the chair, empty it and clean it.

5 = Needs some assistance. The patient requires help because of imbalance, in handling clothes or in using toilet paper.

0 = The patient can not meet the criteria as defined above.

6. Continence of Bowels

10 = Independent. The patient is able to control his bowel and have no accidents. He can use a suppository or take an enema when necessary.

5 = Needs some assistance. The patient needs help in using a suppository, taking an enema or has occasional accidents.

0 = The patient can not meet the criteria as defined above.

7. Bladder Control

10 = Independent. The patient is able to control his bladder day and night. Spinal cord injury patients who wear an external device and a leg bag must be able to put them on independently, clean and empty the bag and stay dry day and night.

5 = Needs some assistance. The patient has occasional accidents, can not wait for the bed pan, get to the toilet in time or needs help with an external device.

0 = The patient can not meet the criteria as defined above.

8. Chair / Bed Transfers

15 = Independent. The patient can transfer from a chair to a bed and back again safely. For those in a wheelchair, the patient can safely approach the bed in his wheelchair, lock breaks, lift footrests, move safely onto the bed, lie down, come to a sitting position on the side of the bed, change the position of the wheelchair, if necessary, to transfer back into it safely.

10 = Needs some assistance. The patient needs to be reminded or supervised for safety of one or more steps of this activity.

0 = The patient can not meet the criteria as defined above.

9. Walking on a Level Surface

15 = Independent. The patient can walk at least 50 yards without help or supervision. He may wear braces or prostheses and use crutches, canes or a walkerette but not a rolling walker. He must be able to lock and unlock braces if used, assume the standing position and sit down, get the necessary mechanical aids into position for use and dispose of them when he sits (putting on and taking off braces is scored under #4. Dressing and Undressing).

10 = Needs some assistance. The patient needs supervision in any of the above tasks but can walk at least 50 yards with minimal help.

0 = The patient can not meet the criteria as defined above.

10. Ascending and Descending Stairs

10 = Independent. The patient is able to go up and down a flight of stairs safely without help or supervision. He may, and should, use handrails, canes or crutches when needed. He must be able to carry the cane or crutches as he ascends and descends the stairs.

5 = Needs some assistance. The patient needs help with or supervision of any one of the above tasks.

0 = The patient can not meet the criteria as defined above.

11. Propelling a Wheelchair

Do not score this item if the patient gets a score for #9.

10 = Independent. The patient can not ambulate but can propel a wheelchair independently. He must be able to go around corners, turn around, maneuver the chair to a table, bed, toilet, etc. He must be able to propel the chair at least 50 yards.

0 = Needs some assistance.

0 = The patient can not meet the criteria as defined above.

Older Americans Resources Scale (OARS-IADL)

I'd like to ask you about some of the activities of daily living, things that we all need to do as part of our daily lives. I would like to know if you can do these activities without any help at all, or if you need some help to do them, or if you can't do them at all.

(Be sure to read all answer choices if applicable in questions 1 through 15 to respondent.)

1. Can you use the telephone...
   (2) without help, including looking up numbers and dialing
   (1) with some help (can answer the phone or dial the operator in an emergency, but needs a special phone or help in getting the number or dialing)
   (0) or are you completely unable to use the telephone
   --- not answered

2. Can you get to places out of walking distance...
   (2) without help (can travel on buses, taxis or drive your own car)
   (1) with some help (need someone to help you or go with you when travelling
   (0) or are you unable to travel unless emergency arrangements are made for a specialized vehicle like an ambulance
   --- not answered

3. Can you go shopping for groceries or clothes (assuming has transportation)
   (2) without help (taking care of all shopping needs yourself assuming that you had transportation)
   (1) with some help (need someone to go with you on all shopping trips)
   (0) or are you completely unable to do any shopping
   --- not answered

4. Can you prepare your own meals…
   (2) without help (plan and cook meals yourself)
   (1) with some help (can prepare some things but unable to cook full meals yourself)
   (0) or are you completely unable to prepare any meals
   --- not answered

5. Can you do your own housework…
   (2) without help (can scrub floors etc.)
   (1) with some help (can do light housework but need help with heavy work)
   (0) or are you completely unable to do any housework
   --- not answered

6. Can you take your own medicine…
   (2) without help (in the right doses at the right time)
   (1) with some help (able to take medicine if someone prepares it for you and/or reminds you to take it)
   (0) or are you completely unable to take your own medicine
   --- not answered

7. Can you handle your own money…
   (2) without help (write checks, pay bills, etc.)
   (1) with some help (manage day to day buying but need some help with managing cheque book and paying bills)
   (0) or are you completely unable to handle your own money
   --- not answered

Preference Based Stroke Index (PBSI)

For each of the following questions, please indicate the choice that best describes your capacity to perform different activities considering your own health state today. Make sure to answer all questions and to only tick one box per question.

Mobility:

Walking

How would you best describe your ability to walk with or without a walking aid.

- ☐ I am able to walk in the community, as I need to
- ☐ I am able to walk inside the house, but I have difficulty walking alone outside
- ☐ I am able to walk only a few steps or I use a wheelchair

Stairs

How would you best describe your ability to go up and down stairs.

- ☐ I can go up and down several flights of stairs
- ☐ I can go up and down only a few steps
- ☐ I can't go up and down stairs

Physical activities

How would you best describe your ability to perform physically demanding activities.

- ☐ I can do sports or physically demanding activities as usual
- ☐ I cannot do sports anymore but I can still manage some physically demanding activities (like, carrying heavy bags, gardening, etc)
- ☐ I can no longer do sports or any physically demanding activities

Usual activities:

Recreational activities

How would you best describe your participation in recreational activities (like painting, knitting, playing cards, etc)

- ☐ I can participate in all recreational activities I wish to
- ☐ I can participate only in some recreational activities that I wish to
- ☐ I cannot participate in the recreational activities I wish to

Work/activity

How would you best describe your ability to accomplish work or any other activities

- ☐ I can work or perform activities as I used to
- ☐ I do not always perform my work or activities as I used to
- ☐ I can no longer work or perform activities as I used to

Driving

How would you best describe your ability to drive a car.

- ☐ I can drive a car anywhere, as before
- ☐ I can drive a car in my neighborhood, avoiding traffic or highways
- ☐ I am unable to drive since my stroke

- ☐ I have never driven a car or had stopped driving long before I had my stroke

Memory

How would you best describe your ability to remember things

- ☐ I am able to remember most things
- ☐ I am somewhat forgetful
- ☐ I am very forgetful

Speech

How would you best describe your ability to speak

- ☐ I can be completely understood when speaking with strangers
- ☐ I can be completely understood when speaking with those who know me well but only partially understood by strangers
- ☐ I can hardly be understood by anyone

Coping

How would you best describe your ability to deal with life problems

- ☐ I can cope with life problems as they come
- ☐ I am sometimes overwhelmed by life problems
- ☐ I often feel helpless when dealing with life problems

Self-esteem

How would you best describe your appreciation of yourself.

- ☐ I am satisfied with myself most of the times
- ☐ I sometimes feel I have good qualities but do not consider myself equal to others
- ☐ I often feel I am a failure, with much less worth than others

The Reintegration to Normal Living Index

Questionnaire

Name: ______________________________ Date: _____________

Hospital: ____________________________

The next set of questions will be a set of statements and you will have three choices. The choices are: yes, partially or no

	Y	P	N
1. Do you move around your living quarters as you feel is necessary?	0	1	2
2. Do you move around your community as you feel is necessary? (shopping, banking, etc.)	0	1	2
3. Are you able to take trips out of town as you feel necessary?	0	1	2
4. Are you comfortable with how you feel your self–care needs are met?	0	1	2
5. Do you spend most of your days occupied in activity that is necessary or important to you?	0	1	2
6. Are you able to participate in recreational activities as you want to? (hobbies, sports, cards, etc.)	0	1	2
7. Are you participating in social activities with family, friends, and/or business acquaintances as is necessary or desirable to you?	0	1	2
8. Are you assuming a role in your family which meets your needs and those of other family members? (family means people with whom you live and/or relatives who you see on a regular basis)	0	1	2
9. In general, are you comfortable with your personal relationships?	0	1	2
10. In general, are you comfortable with yourself when you are in the company of others?	0	1	2
11. Do you feel that you can deal with life events when they happen?	0	1	2

EURO-QOL (EQ) - 5D

Questionnaire

Name: ______________________ Date: ______________

Hospital: ______________________

Please indicate which statement best describes your own health state today. Do not tick more than one box in each group.

Mobility

I have no problems in walking about	☐
I have some problems in walking about	☐
I am confined to bed	☐

Self-Care

I have no problems with self-care	☐
I have some problems washing or dressing myself	☐
I am unable to wash or dress myself	☐

Usual Activities (*e.g. work, study, housework, family or leisure activities*)

I have no problems with performing my usual activities	☐
I have some problems with performing my usual activities	☐
I am unable to perform my usual activities	☐

Pain / Discomfort

I have no pain or discomfort	☐
I have moderate pain or discomfort	☐
I have extreme pain or discomfort	☐

Anxiety / Depression

I am not anxious or depressed	☐
I am moderately anxious or depressed	☐
I am extremely anxious or depressed	☐

Health Utility Index

The next set of questions asks about *[YOUR / PATIENT'S NAME'S]* day-to-day health. The questions are not about illnesses like colds that affect people for short periods of time. They are concerned with a person's usual abilities. You may feel that some of the questions do not apply to *[YOU / PATIENT'S NAME]* but it is important that we ask the same questions of everyone. *(INTERVIEWER : Press <Enter> to continue)*

Vision

1.	**During the past week/month, [*have/has*] [*you/patient's name*] been able to see well enough to read ordinary newsprint *without* glasses or contact lenses?**	☐ Yes (1) Go to #4 ☐ No (2) ☐ DK ☐ Ref
2.	***[Have/has] [you/patient's name]* been able to see well enough to read ordinary newsprint *with* glasses or contact lenses?**	☐ Yes (1) Go to #4 ☐ No (2) ☐ DK/Didn't wear glasses or contact lenses ☐ Ref
3.	**During the past week/month, *[have/has] [you/patient's name]* been able to see at all?**	☐ Yes (1) ☐ No (2) Go to #6 ☐ DK ☐ Ref
4.	**During the past week/month, *[have/has] [YOU/HE/SHE]* been able to see well enough to recognize a friend on the other side of the street without glasses or contact lenses?**	☐ Yes (1) Go to #6 ☐ No (2) ☐ DK ☐ Ref
5.	***[Have/has] [you/HE/SHE]* been able to see well enough to recognize a friend on the other side of the street with glasses or contact lenses?**	☐ Yes (1) ☐ No (2) ☐ Don't know/didn't wear glasses or contact lenses ☐ Ref

Hearing

6.	**During the past week/month, *[have/has] [you/patient's name]* been able to hear what is said in a group conversation with at least 3 other people *without* a hearing aid?**	☐ Yes (1) Go to #11 ☐ No (2) ☐ DK ☐ Ref

7.	***[Have/has] [you/he/she]*** **been able to hear what is said in a group conversation with at least 3 other people *with* a hearing aid?**	☐ Yes (1) Go to # 9 ☐ No (2) ☐ DK/Didn't wear a hearing aid ☐ Ref
8.	**During the past week/month, *[have/has] [you/he/she]* been able to hear at all?**	☐ Yes (1) ☐ No (2) Go to #11 ☐ DK ☐ Ref
9.	**During the past week/month, *[have/has] [you/he/she]* been able to hear what is said in a conversation with one other person in a quiet room *without* a hearing aid?**	☐ Yes (1) Go to #11 ☐ No (2) ☐ DK ☐ Ref
10	***[Have/has] [you/he/she]*** **been able to hear what is said in a conversation with one other person in a quiet room *with* a hearing aid?**	☐ Yes (1) ☐ No (2) ☐ Don't know/didn't wear a hearing aid ☐ Ref
Speech		
11	**During the past week/month, *[have/has] [you/patient's name]* been able to be understood *completely* when speaking *[YOUR/HIS/HER]* own language with people who don't know you?**	☐ Yes (1) Go to #16 ☐ No (2) ☐ DK ☐ Ref -
12	***[Have/has] [you/he/she]*** **been able to be understood *partially* when speaking with people who don't know you?**	☐ Yes (1) ☐ No (2) ☐ DK ☐ Ref
13	**During the past week/month, *[have/has] [you/he/she]* been able to be understood *completely* when speaking with people who know *[YOU/HIM/HER]* well?**	☐ Yes (1) Go to #16 ☐ No (2) ☐ DK ☐ Ref -
14	***[Have/has] [you/he/she]*** **been able to be understood *partially* when speaking with people who know *[you/him/her]* well?**	☐ Yes (1) Go to #16 ☐ No (2) ☐ DK ☐ Ref
15	**During the past week/month, *[have/has] [you/he/she]* been able to speak at all?**	☐ Yes (1) ☐ No (2) ☐ DK ☐ Ref

Getting Around		
16	During the past week/month, *[have/has] [you/patient's name]* been able to bend, lift, jump and run *without difficulty* and *without help or equipment* of any kind?	☐ Yes (1) Go to #24 ☐ No (2) ☐ DK ☐ Ref
17	*[Have/has] [you/patient's name]* been able to walk around the neighbourhood *without difficulty* and *without help or equipment* of any kind?	☐ Yes (1) Go to #24 ☐ No (2) ☐ DK ☐ Ref
18	*[Have/has] [you/he/she]* been able to walk around the neighbourhood *with difficulty* but *without help or equipment* of any kind?	☐ Yes (1) Go to #24 ☐ No (2) ☐ DK ☐ Ref
19	During the past week/month, *[have/has] [you/he/she]* been able to walk at all?	☐ Yes (1) ☐ No (2) Go to #22 ☐ DK ☐ Ref
20	*[Have/has] [you/he/she]* needed mechanical support, such as braces or a cane or crutches, to be able to walk around the neighbourhood?	☐ Yes (1) ☐ No (2) ☐ DK ☐ Ref
21	[Have/has] [you/he/she] needed the help of another person to walk?	☐ Yes(1) ☐ No (2) ☐ DK (3) ☐ Ref (4)
22	[Have/has[[you/he/she] needed a wheelchair to get around the neighbourhood?	☐ Yes (1) ☐ No (2) ☐ DK ☐ Ref
23	*[Have/has] [you/he/she]* needed the help of another person to get around in the wheelchair?	☐ Yes (1) ☐ No (2) ☐ DK ☐ Ref
Hands and Fingers		
24	During the past week/month, *[have/has] [you/he/she]* had the *full use* of both hands and ten fingers?	☐ Yes (1) Go to #28 ☐ No (2) ☐ DK ☐ Ref

25	***[Have/has] [you/he/she]*** **needed the help of another person because of limitations in the use of your hands or fingers?**	☐ Yes (1) ☐ No (2) Go to #27 ☐ DK ☐ Ref
26	***[Have/has] [you/he/she]*** **needed the help of another person with some tasks, most tasks, or all tasks?** **<u>INTERVIEWER</u> : read categories to respondent**	☐ **some tasks** (1) ☐ **most tasks** (2) ☐ **all tasks** (3) ☐ DK ☐ Ref
27	***[Have/has] [you/he/she]*** **needed special equipment, for example special tools to help with dressing or eating, because of limitations in the use of your hands or fingers?**	☐ Yes (1) ☐ No (2) ☐ DK ☐ Ref
Self-Care		
28	**During the past week/month, *[have/has] [you/he/she]* been able to eat, bathe, dress and use the toilet without difficulty?**	☐ Yes (1) Go to #31 ☐ No (2) ☐ DK ☐ Ref
29	***[Have/has] [you/he/she]*** **needed the help of another person to eat, bathe, dress or use the toilet?**	☐ Yes (1) ☐ No (2) ☐ DK ☐ Ref
30	***[Have/has] [you/he/she]*** **needed special equipment or tools to eat, bathe, dress or use the toilet?**	☐ **Yes** ☐ No ☐ DK ☐ Ref

Feelings		
31	During the past week/month, *[have/has] [you/patient's name]* been feeling happy or unhappy? INTERVIEWER : read categories to respondent	□ Happy □ Unhappy Go to #33 □ DK □ Ref
32	Would you describe *[yourself/patient's name]* as having felt: (a) happy and interested in life, or (b) somewhat happy?	□ a – Go to #34 □ b – Go to #34 □ DK □ Ref
33	Would you describe *[yourself/patient's name]* as having felt: somewhat unhappy (a) very unhappy (b) so unhappy that life is not worthwhile	□ a □ b □ c □ DK □ Ref
34	During the past week/month, did *[you/he/she]* ever feel fretful, angry, irritable, anxious or depressed?	□ Yes □ No Go to #37 □ DK □ Ref
35	Ho often did [you/patient's name] feel fretful, angry, irritable, anxious or depressed: rarely, occasionally, often, or almost always? INTERVIEWER : read categories to respondent	□ rarely □ occasionally □ often □ almost always □ DK □Ref
36	During the past week/month, did *[you/patient's name*] feel *extremely* fretful, angry, irritable, anxious or depressed, to the point of needing professional help?	□ Yes □ No □ DK □ Ref
Memory		
37	How would you describe *[your/his/her]* ability to remember things during the past week/month: able to remember most things (a) somewhat forgetful (b) very forgetful (c) unable to remember anything at all? INTERVIEWER : read categories to respondent	□ a □ b □ c □ d □ e □ DK □ Ref

Thinking		
38	How would you describe *[your/patient's name]* ability to think and solve day-to-day problems during the past week/month? (a) Able to think clearly and solve problems (b) Had a little difficulty (c) Had some difficulty (d) had a great deal of difficulty (e) Unable to think or solve problems <u>INTERVIEWER</u> : read categories to respondent	☐ a ☐ b ☐ c ☐ d ☐ DK ☐ Ref
Pain and Discomfort		
39	*[Have/has] [you/patient's name]* had any trouble with pain or discomfort during the past week/month?	☐ Yes ☐ No Go to #41 ☐ DK ☐ Ref
40	How many of *[your/patient's name]* activities during the past week/month were limited by pain or discomfort: none, a few, some, most, all? <u>INTERVIEWER</u> : read categories to respondent	☐ None ☐ A few ☐ Some ☐ Most ☐ All ☐ DK ☐ Ref
41	Overall, how would you rate *[your/patient's name]* health during the past week/month: (a)excellent (b)very good Good Fair Poor <u>INTERVIEWER</u> : read categories to respondent	☐ a ☐ b ☐ c ☐ d ☐ e ☐ DK ☐ Ref

SF-36 Health Status Survey / CANADA

Questionnaire

Name: ______________________ Date: ____________

Hospital: ______________________

INSTRUCTIONS: This survey asks for your views about your health. This information will help keep track of how you feel and how well you are able to do your usual activities.

Answer every question by marking the answer as indicated. If you are unsure about how to answer a question, please give the best answer you can.

1. In general, would you say your health is: (circle one)

Excellent	1
Very good	2
Good	3
Fair	4
Poor	5

2. Compared to one year ago, how would you rate your health in general now? (circle one)

Much better now than one year ago.	1
Somewhat better now than one year ago . . .	2
About the same as one year ago . . .	3
Somewhat worse now than one year ago . .	4
Much worse now than one year ago . . .	5

3. The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much? (circle one number on each line)

	Yes, Limited A Lot	Yes, Limited A Little	No, Not Limited At All
a. **Vigorous activities,** such as running, lifting heavy objects, participating in strenuous	1	2	3
b. **Moderate activities,** such as moving a table, pushing a vacuum cleaner, bowling,	1	2	3
c. Lifting or carrying groceries	1	2	3
d. Climbing **several** flights of stairs	1	2	3
e. Climbing **one** flight of stairs	1	2	3
f. Bending, kneeling, or stooping	1	2	3
g. Walking **more than a kilometre**	1	2	3
h. Walking **several blocks**	1	2	3
i. Walking **one block**	1	2	3
j. Bathing or dressing yourself	1	2	3

4. During the past 4 weeks have you had any of the following problems with your work or other regular daily activities as a result of your physical health? (circle one number on each line)

	YES	NO
a. Cut down the **amount of time** you spent on work or other	1	2
b. **Accomplished less** than you would like	1	2
c. Were limited in the **kind** of work or other activities	1	2
d. Had **difficulty** performing the work or other activities (for example, it took extra effort)	1	2

5. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

	YES	NO
a. Cut down the **amount of time** you spent on work or other	1	2
b. **Accomplished less** than you would like	1	2
c. Didn't do work or other activities as **carefully** as usual	1	2

6. During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups? (circle one)

Not at all	1
Slightly	2
Moderately	3
Quite a bit	4
Extremely	5

7. How much bodily pain have you had during the past 4 weeks? (circle one)

None	1
Very mild	2
Mild	3
Moderate	4
Severe	5
Very severe	6

8. During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)? (circle one)

Not at all	1
A little bit	2
Moderately	3
Quite a bit	4
Extremely	5

9. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks

(circle one number on each line)

	All of the Time	**Most of the Time**	**A Good Bit of the Time**	**Some of the Time**	**A Little of the Time**	**None of the Time**
a. Did you feel full of pep?	1	2	3	4	5	6
b. Have you been a very nervous	1	2	3	4	5	6
c. Have you felt so down in the dumps that nothing could	1	2	3	4	5	6
d. Have you felt calm and	1	2	3	4	5	6
e. Did you have a lot of energy?	1	2	3	4	5	6
f. Have you felt downhearted	1	2	3	4	5	6
g. Did you feel	1	2	3	4	5	6
h. Have you been a happy	1	2	3	4	5	6
i. Did you feel	1	2	3	4	5	6

10. During the <u>past 4 weeks</u>, how much of the time has your <u>physical health or emotional problems</u> interfered with your social activities (like visiting with friends, relatives, etc.)?

(circle one)

All of the time. 1

Most of the time. 2

Some of the time 3

A little of the time 4

None of the time 5

11. How **TRUE** or **FALSE** is each of the following statements for you?

(circle one number on each line)

	Definitely True	Mostly True	Don't Know	Mostly False	Definitely False
a. I seem to get sick a little easier	1	2	3	4	5
b. I am as healthy as anybody I	1	2	3	4	5
c. I expect my health to get	1	2	3	4	5
d. My health is	1	2	3	4	5

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.
These questions are about the physical problems, which may have occurred as a result of your stroke.

1. In the past week, how would you rate the strength of your....	A lot of strength	Quite a bit of strength	Some strength	A little strength	No strength at all
a. Arm that was most affected by your stroke?	5	4	3	2	1
b. Grip of your hand that was most affected by your stroke?	5	4	3	2	1
c. Leg that was most affected by your stroke?	5	4	3	2	1
d. Foot/ankle that was most affected by your stroke?	5	4	3	2	1

These questions are about your memory and thinking.

2. In the past week, how difficult was it for you to...	Not difficult at all	A little difficult	Somewhat difficult	Very difficult	Extremely difficult
a. Remember things that people just told you?	5	4	3	2	1
b. Remember things that happened the day before?	5	4	3	2	1
c. Remember to do things (e.g. keep scheduled appointments or take medication)?	5	4	3	2	1
d. Remember the day of the week?	5	4	3	2	1
e. Concentrate?	5	4	3	2	1
f. Think quickly?	5	4	3	2	1
g. Solve everyday problems?	5	4	3	2	1

These questions are about how you feel, about changes in your mood and about your ability to control your emotions since your stroke.

3. In the past week, how often did you...	None of the time	A little of the time	Some of the time	Most of the time	All of the time
a. Feel sad?	5	4	3	2	1
b. Feel that there is nobody you are close to?	5	4	3	2	1
c. Feel that you are a burden to others?	5	4	3	2	1
d. Feel that you have nothing to look forward to?	5	4	3	2	1
e. Blame yourself for mistakes that you made?	5	4	3	2	1
f. Enjoy things as much as ever?	5	4	3	2	1
g. Feel quite nervous?	5	4	3	2	1
h. Feel that life is worth living?	5	4	3	2	1
i. Smile and laugh at least once a day?	5	4	3	2	1

The following questions are about your ability to communicate with other people, as well as your ability to understand what you read and what you hear in a conversation.

4. In the past week, how difficult was it to...	Not difficult at all	A little difficult	Somewhat difficult	Very difficult	Extremely difficult
a. Say the name of someone who was in front of you?	5	4	3	2	1
b. Understand what was being said to you in a conversation?	5	4	3	2	1
c. Reply to questions?	5	4	3	2	1
d. Correctly name objects?	5	4	3	2	1
e. Participate in a conversation with a group of people?	5	4	3	2	1
f. Have a conversation on the telephone?	5	4	3	2	1
g. Call another person on the telephone, including selecting the correct phone number and dialing?	5	4	3	2	1

The following questions ask about activities you might do during a typical day.

5. In the past 2 weeks, how difficult was it to...	Not difficult at all	A little difficult	Somewhat difficult	Very difficult	Could not do at all
a. Cut your food with a knife and fork?	5	4	3	2	1
b. Dress the top part of your body?	5	4	3	2	1
c. Bathe yourself?	5	4	3	2	1
d. Clip your toenails?	5	4	3	2	1
e. Get to the toilet on time?	5	4	3	2	1
f. Control your bladder (not have an accident)?	5	4	3	2	1
g. Control your bowels (not have an accident)?	5	4	3	2	1
h. Do light household tasks/chores (e.g. dust, make a bed, take out garbage, do the dishes)?	5	4	3	2	1
i. Go shopping?	5	4	3	2	1
j. Do heavy household chores (e.g. vacuum, laundry or yard work)?	5	4	3	2	1

The following questions are about your ability to be mobile, at home and in the community.

6. In the past 2 weeks, how difficult was it to...	Not difficult at all	A little difficult	Somewhat difficult	Very difficult	Could not do at all
a. Stay sitting without losing your balance?	5	4	3	2	1
b. Stay standing without losing your balance?	5	4	3	2	1
c. Walk without losing your balance?	5	4	3	2	1
d. Move from a bed to a chair?	5	4	3	2	1
e. Walk one block?	5	4	3	2	1
f. Walk fast?	5	4	3	2	1
g. Climb one flight of stairs?	5	4	3	2	1
h. Climb several flights of stairs?	5	4	3	2	1
i. Get in and out of a car?	5	4	3	2	1

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	Could not do at all
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

The following questions are about how stroke has affected your ability to participate in the activities that you usually do, things that are meaningful to you and help you to find purpose in life.

8. During the past 4 weeks, how much of the time have you been limited in...	None of the time	A little of the time	Some of the time	Most of the time	All of the time
a. Your work (paid, voluntary or other)	5	4	3	2	1
b. Your social activities?	5	4	3	2	1
c. Quiet recreation (crafts, reading)?	5	4	3	2	1
d. Active recreation (sports, outings, travel)?	5	4	3	2	1
e. Your role as a family member and/or friend?	5	4	3	2	1
f. Your participation in spiritual or religious activities?	5	4	3	2	1
g. Your ability to control your life as you wish?	5	4	3	2	1
h. Your ability to help others?	5	4	3	2	1

9. Stroke Recovery

On a scale of 0 to 100, with 100 representing full recovery and 0 representing no recovery, how much have you recovered from your stroke?

______ 100 Full Recovery

—

______ 90

—

______ 80

—

______ 70

—

______ 60

—

______ 50

—

______ 40

—

______ 30

—

______ 20

—

______ 10

________ 0 No Recovery

APPENDIX 5

The following series of tables were developed by the candidate and Dr. Lois Finch and also appear in Dr. Finch's thesis: Measuring and predicting early functioning post-stroke. January 2007.

Table 1. Model Quality Criteria for a Rasch Measure

Table 2. Item Quality Criteria for a Rasch Measure

Table 3. Person Quality Criteria for a Rasch Measure

APPENDIX 5 – Table 1 Model Quality Criteria for a Rasch Measure[*]

Quality Indicators		
Fit to the model	**Criteria**	1.Derivation / 2.Interpretation
Item trait interaction	Non significant model summary χ^2	1. The residuals derived from the difference between the observed and expected score with the expected determined by the model and the hypothesis that the data fit the model. The expected value and variance per item person interaction are calculated, summed across all items and squared to form a χ^2 per item $Z_{ni} = \frac{x_{ni} - E[X_{ni}]}{\sqrt{V[x_{ni}]}}$ summed across all items for a component χ^2 z =standardized residuals of the observed score from that predicted by the model. 2. Indicates that the level of item difficulty of the measure is consistent across subjects. It suggests the items form a linear and unidimensional measure. The difference between the observed and expected means along the continuum of the trait is smaller than expected by chance alone and the data fit the model [399, 143]
Unidimensionality		
The data fit the model	Non significant χ^2	1. See above fit section
Principal Component Analysis (PCA) of person item residuals	Variance explained by 1st component <10% [400]	2. If a measure is unidimensional there should be no meaningful correlations between item residuals. This is tested through a PCA of the residuals after the variance accounted for by the model has been removed. The amount of acceptable variance in the residual PCA analysis indicative of unidimensionality ranges from less than 10 to less than 20% [400]. The ratio of the first to the second Eigen values can be used with a larger ratio indicative of unidimensionality[394]
Principal Component Analysis of the Rasch raw items scores	Variance explained by 1st component >40% [400]	Reflects the degree of invariance across the trait [374]
ICCs A graphical representation of the items' fit to the model	Non crossing	The graphs of ICCs do not cross, but are parallel if the items (with the same number of response options) belong to a single construct [401, 402]

[*] The Quality criteria for a Rasch model are dependent on the program used for analysis. The criteria here are based on the Rasch Unidimensional Measurement Model program (RUMM2020) [143]. Additional criteria considered for analysis techniques and models are not included.

Table 1 Model Quality Criteria for a Rasch Measure (cont.) *

Quality Indicators		
Fit to the model	**Criteria**	1.Derivation / 2.Interpretation
Invariance	The data fit the model	2. A Person's True ability does NOT depend on items administered. An Item's difficulty does NOT depend on the characteristics of the people taking it
Precision		
Information function	The larger the more precise $I_i(\beta)=\frac{\left[P_i'(\beta)\right]^2}{P_i(\beta)(1-P_i)(\beta)}$	I=information, P= probability correct, Beta =ability. [398,403,404] 1. It is the inverse of the standard error squared per item 2. indicates a) the precision of the estimation procedure per item, 2b) The amount of information provided by an item at an ability level i.e. it delineates the range over which an item is most useful for defining person ability
Test information function (TIF)	The larger the more precise $TIF(\beta)=\sum_{i=1}^{k} I_i(\beta)$	1. Provides an indication of the precision of the measure or the Standard Error of the measure. SEM= (TIF) $^{1/2}$. 2a)The amount of information provided by a test about ability level at each maximum likelihood estimate[398] 2b) Can be used to compare the amount of information in different measures or subsets of items within a measure

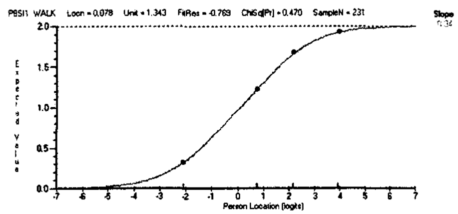
* The Quality criteria for a Rasch model are dependent on the program used for analysis. The criteria here are based on the Rasch Unidimensional Measurement Model program (RUMM2020) [143]. Additional criteria considered for analysis techniques and models are not included.

Table 2 Item Quality Criteria for a Rasch Measure[*]

Quality Indicators			
Item Fit		**Criteria**	1.Derivation / 2.Interpretation
Statistic	Standardized residuals	Per item -2 < residuals< + 2 Mean residuals close to '0' and SD of residuals close to '1' Power to detect fit is affected by number of items and sample size	1. Residuals from χ^2, squared and summed over all groups of subjects, transformed to approximate a normal distribution z =standardized residuals of the observed score from that predicted by the model then log transformed [143,149] $Z_{ni} = \frac{x_{ni} - E[X_{ni}]}{\sqrt{V[x_{ni}]}}$ 2. Items are considered to fit the model if residuals are not greater/less than ± 2 the 95% confidence interval of the normal distribution. >+ 2 can indicate irregular response patterns, noise, & multidimensionality [393] >- 2 indicates irregular response patterns, statistical dependency, and redundancy. [143]
	Chi-square	Non significant	1. The residuals derived from the observed-expected score with the expected determined by the model with the hypothesis that the data fit the model 2. The difference between the observed and expected is smaller than expected by chance alone and the item fits the model. Provides a general idea of fit. It tests the data against a perfect fit to the model not against a better fit [145]
	F-statistic	Non significant	1. A one way analysis of variance on the standardized residuals 2. A comparison between the F and χ^2 assists in determining item fit [143]

[*] The Quality criteria for a Rasch model are dependent on the program used for analysis. The criteria here are based on the Rasch Unidimensional Measurement Model program [143] Additional criteria considered for analysis techniques and models are not included.

Table 2 Item Quality Criteria for a Rasch Measure (cont.) *

	Quality Indicators		
	Item Fit	**Criteria**	**1.Derivation / 2.Interpretation**
Graph	Item characteristic curves (ICCs),	The sample is divided into groups by ability, 4 in the example below, with observed responses plotted against the predicted and represented by dots on the predicted model curve or ICC 	1. ICCs are graphical indicators of item fit. On an ICC graph the x-axis is logit ability, the y-axis the expected logit value. The observed ability is plotted against that predicted by the model. 2. ICCs indicate the location of the item and the probability of success on that item for each person's level of ability along the item's continuum. The slope of the ICC indicates the rate of change in the probability of success on that item as a function of ability. [143]

* The Quality criteria for a Rasch model are dependent on the program used for analysis.The criteria here are based on the Rasch Unidimensional Measurement Model program [143] Additional criteria considered for analysis techniques and models are not included.

Table 2. Item Quality Criteria for a Rasch Measure (cont.)

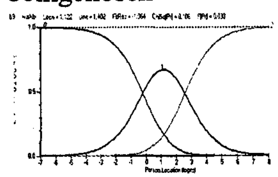
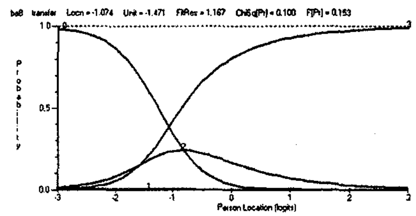
	Quality Indicators	Criteria	
Ordered Response Options			**1.Derivation / 2.Interpretation**
Statistic	Threshold values	Thresholds ordered from low to high based on numeric response options	2. A disordered response option results when more able people do not have a greater probability of successfully responding to a more difficult level of an item than the less able. [164 405 406]
Graph	Category Characteristic Curves (CCCs)	Ordered response options are seen as a series of hills Each option has a probability of beingchosen 	1. The horizontal axis represents ability the vertical axis represents the probability of endorsing a response option. Each curve represents the threshold between response levels 0: between 0 and 1; 1 between 1 and 2.[143] 
Reliability			
	Reliability index	Ranges from 0-1 with 1 representing perfect reliability and 0 no reliability	1. An indication of the consistency of the item responses in the sample r=1-(MSEi)/(Mean Variance of Items); MSEi = $\Sigma(SE^2)/N$ where σ is the estimated variance of the item. A ratio of the adjust item variance to the observed item variance in logits 2. Reliability of the item hierarchy; if the items were given to a different population of person with same attributes [393 148] the hierarchy would remain the same
	Separation index[149 148 407]	Ranges from 0 to ∞ and is interpreted as a Cronbach's alpha: Acceptable: 1.5 or α=0.7; good: 2.0 or α =0.8; and excellent: 3.0 or α= 0.9	1. Separation Index or G = $\sqrt{\dfrac{\textit{reliability}\ \text{coefficient}}{(1-\textit{reliabiltycoefficient})}}$ 2. Indicates the spread of person ability or item difficulty in standard error units. The larger the index the better the differentiation is between subjects and item difficulty. It aids in quantifying the construct and facilitates the measurement of change
	Strata		1 Derived from the separation index: Strata = (4G+1)/3 2 Number of statistically different levels, separated by 3 standard errors, of item difficulty that can be identified [148 393]

Table 2 Item Quality Criteria for a Rasch Measure (cont.)

Quality Indicators		
Precision	**Criteria**	**1.Derivation / 2.Interpretation**
Information function	The larger the more precise $I_i(\beta) = \frac{\left[P_i'(\beta)\right]^2}{P_i(\beta)(1-P_i)(\beta)}$	I=information P= probability correct Theta =ability 1. It is the inverse of the standard error squared per item 2. Indicates a) the precision of the estimation procedure, b) The amount of information provided by an item at an ability level [398,403,404,408] The information statistic indicates where the item contributes the most information along the continuum.
Statistical independence of the items	Standardized Fit statistics >-2.0, residual inter-item correlations>0.3	1. See above for determination of standardized residuals 2. Ability is based only on ability and not influenced by other factors. The answer to one items is not influenced by the answers to any other item [398,403,404]
Differential Item Function (DIF) or Item bias.	1. Separate calibration t-test non-significant [138] 2. Between group item fit statistic [409] likelihood ratio chi-square in RUMM non-significant [410] two-way ANOVA of residuals with people divided by ability and divided within that group by the factor	Differential Item Functioning or item bias (DIF) indicates that each item works in the same way for different subpopulations of the sample that are compared. [143] 1. t-test based on 2 separate calibrations of the same item on 2 subpopulations of interest e.g. male, female. $t = \frac{d_{i1} - d_{i2}}{(s_{i1} + s_{i2})^{1/2}}$ d_{i1} = difficulty of item I in subpopulation 1ie male d_{i2} = difficulty of item I in subpopulation 2 ie female s_{i1} = standard error for d_{i1} s_{i2} = standard error for d_{i2} * Multiple comparisons for a single item raise questions about the appropriateness of the Type I error rates. 2. This statistic is based on subpopulation residuals after the variance for the items have been calibrated (Between group item fit statistic criteria based on the WINSTEPS programme [138] Per item -2 < residuals< + 2)

Table 3 Person Quality Criteria for a Rasch Measure*

Quality Indicators	Person	1.Derivation / 2.Interpretation
Fit	**Criteria**	
Standardized residuals	Per person - 2 < residuals<+ 2 Mean residuals close to '0' and SD close to '1'. Affected by sample size and spread of ability	1. Residuals from χ^2, squared and summed over all groups of subjects, transformed to approximate a normal distribution z =standardized residuals of the observed score from that predicted by the model then log transformed[143, 149] $Z_{ni} = \frac{x_{ni} - E[X_{ni}]}{\sqrt{V[x_{ni}]}}$ 2. Persons are considered to fit the model if residuals are not greater/less than ± 2 the 95% confidence interval of the normal distribution. >+ 2 indicates irregular response patterns, or noise, data entry errors [393] >- 2 indicates irregular response patterns, statistical dependency of responses. [143]
Chi-square	Non significant	1. The residuals derived from observed-expected score with the expected determined by the model with the hypothesis that the data fit the model 2. The difference between the observed and expected is smaller than expected by chance alone and the person fits the model

* The Quality criteria for a Rasch model are dependent on the program used for analysis. The criteria here are based on the Rasch Unidimensional Measurement Model program [143]. Additional criteria considered for analysis techniques and models are not included.

Table 3. Person Quality Criteria for a Rasch Measure (cont.) *

Quality Indicators Fit	Person Criteria	1.Derivation / 2.Interpretation
Reliability index	Ranges from 0-1 with 1 representing a perfect reliability and 0 no reliability	1. An indication of the consistency of the responses in the sample r = 1- (MSE_p)/ (Mean Variance of persons), $MSE_p = \Sigma(SE^2)/N$ where σ is the estimated variance 2. Reliability of the of person hierarchy if the same people were given a different test of the same construct [148,393]
Separation index	Similar to Cronbach's alpha Acceptable 1.5 good 2.0, and excellent 3.0	1. Separation Index or G= $\sqrt{\frac{reliability\ \text{coefficient}}{(1 - reliabilty coefficient)}}$ [149 148 407] 2. Indicates the spread of person ability in standard error units. The larger the index the better the differentiation is between subjects and item difficulty, aids in quantifying the construct and facilitates the measurement of change [411 148]
Strata		1 derived from the separation index: STRATA=(4G+1)/3 2 Number of statistically different levels, separated by 3 standard errors, of person ability that can be identified [148 393]
Invariance		2. A Person's True ability does NOT depend on items administered An Item's difficulty does NOT depend on the people taking it

* The Quality criteria for a Rasch model are dependent on the program used for analysis. The criteria here are based on the Rasch Unidimensional Measurement Model program [143]. Additional criteria considered for analysis techniques and models are not included.

APPENDIX 6 - Additional statistical considerations

Polychoric vs. Pearson correlations

The polychoric correlation coefficient was introduced by Pearson as a more adequate choice when both variables are dichotomous or ordinal but both are assumed to reflect underlying continuous variables.[412] The estimate is based on the assumption of an underlying continuous bivariate normal distribution.[412]

In this particular study, ordinal data is used and thus factor analysis (or principal component analysis (PCA)) should be done, in theory, using a polychoric correlation as an input.[38] When using a polychoric correlation, the only assumption imposed on the data is its ordinal properties.

However, there were some disadvantages to using polychoric correlations in this particular study. First, to estimate the polychoric correlation, SAS uses an iterative, maximum likelihood method to estimate the polychoric correlation. When missing values are present, certain pairs of variables do not form at least a 2x2 table and thus the correlation cannot be computed. Also, a maximum likelihood approach demands larger sample sized to yield accurate results.

According to Coenders and Saris,[413] if the variables are categorized with approximately equally-spaced thresholds, polychoric correlations should be used, but Pearson correlations should produce similar results. However, if non-normal underlying variables are categorized with equally-spaced thresholds, then Pearson correlations should be preferred. Also, Nandakumar and colleagues [414] found that Pearson correlations for factor analyses produced at least as good results as polychoric correlations and that polychoric correlation may be more adequate for four-category response options than dichotomously scored items.

Each one of the particular studies in this thesis were very different in terms of sample size, the distribution of the variables and also, the items were varied in their scoring responses (dichotomous and polytomous items), making the choice between polychoric and Pearson correlations very difficult.

Because the debate is still ongoing as to whether polychoric or Pearson correlation coefficients should be used in which instances, both were performed for each one of the studies in this thesis. Both methods yielded the exact same variables as not fitting on the first factor and thus increased our confidence in the results obtained.