



Measurement Structure of the Wolf Motor Function Test: Implications for Motor Control Theory

Michelle Woodbury, *University of Florida*
Craig A. Velozo, *University of Florida*
Paul A. Thompson, *Sanford Research/USD*
Kathye Light, *University of Florida*
Gitendra Uswatte, *University of Alabama*
Edward Taub, *University of Alabama*
Carolee J. Winstein, *University of Southern California*
David Morris, *University of Alabama*
[Sarah Blanton](#), *Emory University*
Deborah S. Nichols-Larsen, *Ohio State University*

Only first 10 authors above; see publication for full author list.

Journal Title: Neurorehabilitation and Neural Repair

Volume: Volume 24, Number 9

Publisher: SAGE Publications (UK and US) | 2010-11-01, Pages 791-801

Type of Work: Article | Post-print: After Peer Review

Publisher DOI: 10.1177/1545968310370749

Permanent URL: <https://pid.emory.edu/ark:/25593/tv02t>

Final published version: <http://dx.doi.org/10.1177/1545968310370749>

Copyright information:

© The Author(s) 2010.

Accessed October 27, 2021 6:41 PM EDT



Published in final edited form as:

Neurorehabil Neural Repair. 2010 ; 24(9): 791–801. doi:10.1177/1545968310370749.

Measurement Structure of the Wolf Motor Function Test: Implications for Motor Control Theory

Michelle Woodbury, PhD^{1,2}, Craig A. Velozo, PhD^{1,2}, Paul A. Thompson, PhD³, Kathye Light, PhD², Gitendra Uswatte, PhD⁴, Edward Taub, PhD⁴, Carolee J. Winstein, PhD⁵, David Morris, PhD⁴, Sarah Blanton⁶, Deborah S. Nichols-Larsen, PhD⁷, and Steven L. Wolf, PhD⁶

¹Veterans Administration Medical Center, Gainesville, FL, USA

²University of Florida, Gainesville, FL, USA

³Sanford Research/USD, Sioux Falls, SD, USA

⁴University of Alabama at Birmingham, Birmingham, AL, USA

⁵University of Southern California, Los Angeles, CA, USA

⁶Emory University School of Medicine, Atlanta, GA, USA

⁷Ohio State University, Columbus, OH, USA

Abstract

Background—Tools chosen to measure poststroke upper-extremity rehabilitation outcomes must match contemporary theoretical expectations of motor deficit and recovery because an assessment's theoretical underpinning forms the conceptual basis for interpreting its score.

Objective—The purpose of this study was to investigate the theoretical framework of the Wolf Motor Function Test (WMFT) by (1) determining whether all items measured a single underlying trait and (2) examining the congruency between the hypothesized and the empirically determined item difficulty orders.

Methods—Confirmatory factor analysis (CFA) and Rasch analysis were applied to existing WMFT Functional Ability Rating Scale data from 189 participants in the EXCITE (Extremity Constraint-Induced Therapy Evaluation) trial. Fit of a 1-factor CFA model (all items) was compared with the fit of a 2-factor CFA model (factors defined according to item object-grasp requirements) with fit indices, model comparison test, and interfactor correlations.

Results—One item was missing sufficient data and therefore removed from analysis. CFA fit indices and the model-comparison test suggested that both models fit equally well. The 2-factor model yielded a strong interfactor correlation, and 13 of 14 items fit the Rasch model. The Rasch item difficulty order was consistent with the hypothesized item difficulty order.

Conclusion—The results suggest that WMFT items measure a single construct. Furthermore, the results depict an item difficulty hierarchy that may advance the theoretical discussion of the person ability versus task difficulty interaction during stroke recovery.

© The Author(s) 2010

Corresponding Author: Michelle Woodbury, College of Public Health and Health Professions, Department of Occupational Therapy, PO Box 100164, University of Florida, Gainesville, FL 32610, USA, mwoodbur@php.ufl.edu.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the authorship and/or publication of this article.

Keywords

stroke; rehabilitation; Rasch analysis; Wolf Motor Function Test; motor control; measurement theory

Introduction

In recent years, stroke rehabilitation has progressed as advances in neuroscience and rehabilitation have characterized the potential for the neural system to enhance recovery after stroke-related brain damage.¹ Current motor control theories suggest that movement is organized and controlled according to an adaptive interaction between a person and contextual demands.²⁻⁷ Contemporary theories also suggest that poststroke recovery is a process whereby patients regain, via practice, an ability to meet and overcome task demands.^{8,9} This process has led to the development of new rehabilitation research interventions¹⁰⁻¹³ and revision of traditional interventions¹⁴ to restore poststroke upper-extremity (UE) motor function.

Further advances in stroke rehabilitation will require that tools chosen to measure motor control are consistent with contemporary expectations of deficit, repair, and recovery.^{15,16} The theoretical underpinnings of an assessment form the conceptual basis for interpreting its score.¹⁷ The score generated by a poststroke UE motor assessment should specify movements patients can and cannot perform, quantify motor impairment by locating individuals along a continuum of ability, and inform the expected next steps in the recovery process. Information gained from such tools provides critical information to guide treatment, specify experimental effects, and/or define change. Moreover, this information provides empirical evidence to assess the validity of motor control theory itself. Therefore, there exists a pressing need to critically examine the consistency between commonly used stroke rehabilitation outcome assessments and present-day motor control theory.

In general, poststroke behavioral assessment has lagged behind advances in theory and may actually be inconsistent with contemporary conceptual frameworks. For example, Mulder and colleagues¹⁸ suggested that movement analysis tools, although technically proficient, often lack theoretical frameworks guiding their use, which could restrict clinical relevance of the results. Several recent investigations of the Fugl-Meyer UE Assessment,¹⁹ a widely used instrument in contemporary rehabilitation,²⁰ emphasize that interpretation of its score depends on whether its underlying theoretical model is supported²¹ or challenged.²²

The Wolf Motor Function Test (WMFT) is used regularly in research environments to measure outcomes of constraint-induced movement therapy^{10,13,23-26} and other task practice interventions such as robot-assisted therapy,¹¹ bilateral arm training,²⁷ and electrical stimulation-assisted task practice.²⁸ The WMFT was designed to examine voluntary UE movement through a series of tasks that follow a proposed difficulty order.^{13,23,26} Items were selected to represent common functional and nonfunctional UE movements and were sequenced so as to initially require single-joint proximal reaching movements without object grasp (eg, place forearm on table, place forearm on box) then require multiple-joint proximal and distal movements for reaching to grasp functional objects (eg, lift a soda can, fold a towel). Originally called the Emory Motor Test,²³ the assessment has been subjected to rigorous examination of its psychometric properties²⁹⁻³³ and undergone several revisions.^{13,29,34} The current WMFT retains its original item structure. However, there are no published investigations of its theoretical basis, that is, the relationship of its score to contemporary concepts of poststroke UE motor recovery.

The proposed WMFT item structure implies that all items measure a single UE motor function trait. However, all items do not require interaction with functional objects; thus, the item set might be heterogeneous, measuring multiple UE motor function traits.³⁵ The number of constructs measured by an assessment (dimensionality) can be empirically investigated with factor analysis.³⁶ The WMFT item order specifies a relationship between person ability and motor tasks by describing items that individuals with less motor function should accomplish and items that individuals with more motor function should accomplish. This interaction between individuals' level of motor function and difficulty of tasks can be empirically investigated with item response theory (IRT) analyses.³⁷ IRT calibrates the sample's abilities and the test's item difficulties onto a common metric. The metric is arranged as a continuum of person ability (persons of low ability proceeding to persons of high ability) and item difficulty (items involving less challenge proceeding to items involving high challenge). Identifying an individual's ability on this metric leads to inferences about the types of tasks the individual may be able to perform. Determining the dimensionality of an item set and elucidating the interaction between person ability and task difficulty in this outcome measure can advance motor control theory.

Accordingly, the purpose of this study is to investigate the theoretical framework of the WMFT in a sample of individuals with mild to moderate poststroke UE hemiparesis using factor analysis and IRT methodologies. Specifically, this study aims to determine (1) if all WMFT items measure a single latent trait and (2) the congruency between the hypothesized and empirically determined item difficulty order.

Methods

Participants

Secondary analysis was performed on data from the Extremity Constraint-Induced Therapy Evaluation (EXCITE) multisite clinical trial.¹⁰ The trial systematically tested the effect of constraint therapy among individuals who had experienced a first-time ischemic or hemorrhagic stroke within 3 to 9 months prior to enrollment. Participants were included if they had the ability to voluntarily extend the wrist, thumb, and 2 digits at least 10° 3 times in 1 minute, transfer independently, stand from a sitting position, and remain standing for 2 minutes. Participants were excluded if they demonstrated cognitive impairment, expressed pain in any joint of the hemiparetic UE, had substantial use of the hemiparetic UE in daily life, or had insufficient stamina or unstable medical status prohibiting involvement in the treatment schedule.³⁸

Wolf Motor Function Test

The WMFT consists of 15 task-performance items and 2 strength items. Task-performance items are timed and rated by using the 6-point Functional Ability Scale (FAS, Table 1).²⁹ For the purposes of the present study, analyses were conducted only on FAS item ratings and not item performance times. The FAS describes hemiparetic UE movement during item performance, and ratings range from 0 (*no use*) to 5 (*normal*). Traditionally, FAS item ratings are summed, and the average item rating out of 75 total points is reported.²⁹

EXCITE Test Administration and Rating Procedure

The WMFT was administered by trained testers at each of 7 study sites. All testers attended a 9-day workshop, which included training in test administration procedures using a detailed step-by-step checklist. Testers engaged in hands-on supervised practice and met a predetermined criterion for adherence with standardized study methods.³⁹ Each site sent videotapes of participant WMFT performance to the EXCITE Training Core³⁸ where 4 experienced physical therapists rated each WMFT item using the FAS. Raters engaged in a

thorough training process before rating independently, which included practice rating as a group to establish a common frame of reference and agreement on detailed scoring procedures. All raters demonstrated strong interrater reliability—Intraclass correlation coefficient (ICCs) Types 2,1 and 3,1 ranging from 0.92 to 0.99.³⁹ Reliability was checked periodically throughout the trial.

Study Design

In the EXCITE trial, the WMFT was administered to the sample at baseline (ie, within 1 week of enrollment prior to randomization), immediately posttreatment, and at months 4, 8, 12, 12.5, 16, 20, and 24 posttreatment.¹⁰ For the purposes of the present study, we performed cross-sectional analysis of the FAS rating scale data obtained at baseline.

Confirmatory Factor Analysis

Confirmatory factor analysis (CFA)³⁶ examines the theoretical structure and dimensionality of an assessment. Data in this study were ordinal, and therefore, CFA was conducted (Mplus v 4.21⁴⁰) on the polychoric correlation matrix using a weighted least-squares approach.⁴¹ Evidence of adequate model fit was determined by examination of standard indices: root mean square error (RMSEA) < 0.06,⁴² Tucker-Lewis Index > 0.95,⁴³ Comparative Fit Index (CFI) > 0.95,⁴³ standardized factor loadings > 0.40,³⁶ and R^2 > 0.60.³⁶

In separate CFAs we tested 2 a priori hypotheses. The first CFA tested the hypothesis that all WMFT items contribute to measurement of a single trait: UE function. The second CFA tested the hypothesis that the WMFT comprises items representing 2 traits: factor 1 containing items 1 to 7 (items without object movement requirements) and factor 2 containing items 8 to 15 (items requiring object movements). The 2-factor model is nested within the 1-factor model, enabling factor comparison via the χ^2 difference test.³⁶ The interfactor correlation of the 2-factor model was examined. Strong interfactor correlation (>0.80) would suggest that the factors measure highly related traits.³⁶

Rasch Analysis

Rasch analysis simultaneously evaluates the difficulty of items and the ability of individuals who took the test. Winsteps software⁴⁴ and the Andrich rating-scale model⁴⁵ were used for Rasch analysis.

Rating Scale Diagnostics

The rating scale was examined to ensure that individuals with lower UE function received predominantly lower ratings and individuals with higher UE function received predominantly higher ratings. Three “essential” criteria, defined by Linacre et al,⁴⁶ guided this investigation. First, the number of values for each rating category was tallied. Ratings must be used a minimum of 10 times to ensure stable estimates of person ability and item difficulty. Second, the mean person ability measure for each rating category was examined. Lower rating categories should be associated with lower mean person ability measures and vice versa. Finally, the outfit statistic⁴⁷ for each category was examined. This statistic indicates whether ratings were applied appropriately. Fit statistics are reported as the mean square (MNSQ) of the category residuals (difference between actual sample ratings and ratings expected by the Rasch model).⁴⁷ Outfit MNSQ values <2.0 for each category are recommended.⁴⁶ If rating categories failed to meet these criteria, consideration was given to collapsing categories.

Test Reliability

Test reliability was examined using the index of person separation, which specifies the number of statistically significant strata into which the sample was divided.⁴⁷ Values >2.0 indicate that the assessment distinguishes persons with differing person ability levels.

Dimensionality

Exploration of dimensionality was complemented with 2 Rasch analyses: (1) goodness-of-fit statistics and (2) principal components analysis of the Rasch-generated residuals (PCAr).

Goodness-of-Fit Statistics

Under the Rasch model, low-ability persons should do well on easy items and poorly on difficult items, whereas higher-ability persons should do well on easy and difficult items. Item fit statistics evaluate the adequacy of this model. Item residuals (mean difference between observed person score and expected Rasch model score on each item) measure fit. Deviation from the ideal fit statistic value of 1.0 indicates potential departures from unidimensionality. The infit MNSQ statistic was examined because the measure is sensitive to model violations for items matching the sample ability. Large infit values (>1.40; standardized z values >2.0) were considered evidence against the hypothesis of unidimensionality.⁴⁸

Principal Components Analysis of the Rasch-Generated Residuals

PCAr uncovers dimensions in the data after the first dimension (trait) is removed. Unidimensionality is supported when the variance explained by the first dimension is >50%.⁴⁹

Item Difficulty Hierarchy

The core of the Rasch method is the relationship of person ability to item difficulty. Consistency of the Rasch-generated item difficulty hierarchy was compared with the original item order.

Results

Participants

Existing de-identified data were obtained and analyzed according to a protocol approved by the University of Florida IRB and in accordance with a confidentiality agreement between the University of Florida and Emory University. Data were available from 194 individuals to whom the WMFT had been administered within 7 days of enrollment in the EXCITE trial. Five individuals without rating-scale data were excluded; thus, the final data set included 189 participants. Demographics of the sample are reported in detail elsewhere.¹⁰

Rating Scale Diagnostics

Except for the rating category “0,” the FAS met Linacre’s essential criteria.⁴⁶ The category “0” was observed (given) only 12 times throughout the data set. In addition, disordered average person measures were noted in the progression from “0” to “1.” Persons rated “0” had higher mean person ability values (mean = -1.36 logits) than persons rated “1” (mean = -2.74 logits), indicating that “0”-rated persons had higher levels of UE function than “1”-rated persons. All category outfit values were acceptable (MNSQ < 2.0) except for the category “0” (MNSQ = 2.71), indicating idiosyncratic application of the “0” ratings. That is, the high outfit value indicated that this low rating was unexpectedly given to individuals with high UE function.

These results led to reinspection of the raw data set. On 11/12 occasions, the rating “0” was applied to 1 item (item 7, “reach and retrieve”). The 11 individuals scoring “0” on the item also scored “3,” “4,” or “5” on more difficult items, such as “lift can” and “flip cards.” Therefore, the rating “0” was likely erroneously applied to these individuals. Furthermore, 83 participants (43% of the data set) were missing responses for this item at baseline. The item also showed abnormally high missing data across the entire data set consisting of pretests, posttests, and follow-up tests. Documentation provided by those who rated the assessment revealed several reasons for the missing data: incorrect weight start position (45% of cases), incorrect participant start position (40% of cases), filming error (12% of cases), and unknown reason (2% of cases).

Based on the diagnostic results and the large proportion of missing data, the research team decided to remove this item from subsequent analysis for not conceptually contributing to UE motor function measurement. Following removal, the rating “0” was used only once in the entire data set and was therefore collapsed into the next higher category. Reanalysis showed that all remaining rating categories were performing appropriately (10+ observations, mean person measures increase monotonically across categories, outfit values < 2.0).

Test Reliability

The 14-item WMFT person separation value is high (value = 3.72; test reliability = 0.93; Cronbach’s α = .98). The assessment divided this sample into 5 statistically distinct ability levels.

Dimensionality: CFA

Two separate CFAs were conducted. A 1-factor CFA model tested the fit of all items to that single factor. In the 2-factor CFA model, the first factor included nonobject items, whereas the second factor included items requiring object grasp. Standardized factor loadings were strong for the 1-factor CFA (0.72 – 0.99 \pm 0.03) and the 2-factor CFA (factor 1: 0.82 – 0.99 \pm 0.02; factor 2: 0.80 – 0.93 \pm 0.02) models. Table 2 summarizes fit indices for these 2 models and provides the interfactor correlation for the 2-factor model. The model comparison test was highly significant ($\chi^2 = 70$; P .0001), indicating that the nested model (2-factor) is a worse fit to the data.

Dimensionality: Rasch PCAR

The Rasch PCAR supported the contention of a single factor. The variance explained by the first dimension = 73.6%.

Dimensionality: Rasch Fit Statistics

Table 3 presents item fit statistics. One item (“extend elbow side”) was marginally above the infit criteria (MNSQ = 1.44; ZSTD = 3.5). There was no conceptual reason to remove this item, and so it was retained.

Item Difficulty Hierarchy

Table 3 presents the items in descending item difficulty order and provides average item difficulty calibrations and their standard errors (logits).

Figure 1, a person–item map, presents the sample ability measured on the same metric as item difficulty. The column of numbers to the left is the measurement metric (logits). The center line represents the WMFT’s latent trait, UE motor function. The symbols to the left of the center line represent each participant’s person ability measure. Items are displayed to the

right of this line. The map is arranged so that the most able people and the most difficult items are at the top and the least able people and least difficult items are at the bottom. A curve representing the distribution of the sample's UE motor function has been overlaid onto the figure to illustrate the relationship of person ability to item difficulty. The peak of the distribution curve corresponds to the sample's mean person ability measure: -1.27 ± 0.50 logits.

The location of each item on the metric corresponds to its average difficulty across all categories, that is, the item difficulty measure in Table 3. However, item difficulties actually range over the metric because each item has a rating scale. In Figure 1, the dotted rectangular box surrounding items illustrates this range. The boundaries of the rectangular box correspond to the step estimates (Andrich-thresholds⁴⁹) between the lowest and highest rating scale categories calculated by the analysis. For example, the person ability required to move from a rating of "1" to "2" on the easiest item ("hand to table") is -6.0 logits (bottom boundary of the rectangular box). The person ability required to move from a rating of "4" to "5" on the most difficult item ("stack checkers") is 7.0 logits (top boundary of the rectangular box).

Discussion

The purpose of this study was to test whether WMFT items (1) measure a single trait and (2) progress in difficulty per original expectations. To that end, CFA and Rasch analyses were used to examine the adequacy of the FAS, the test's dimensionality, and the empirically derived item difficulty hierarchy. Results suggest that all rating scale categories, except for "0," adequately measured poststroke UE motor function for individuals who met EXCITE trial inclusion criteria. Furthermore, results indicate that 14/15 assessment items measured a single trait. The hypothesized item order was generally supported. Results are discussed here with regard to implications for contemporary motor control theory.

Rating Scale

The rating "0," defined as "does not attempt with more-involved UE," had a low frequency of use and was used only 12 times in the data set. EXCITE inclusion criteria required that participants exhibit a minimum of 10° wrist and finger extension. Therefore, with this level of motor function, in an instructed testing situation, participants would be unlikely to fail to attempt items with the hemiparetic UE. All other ratings were used as expected.

Rating scale diagnostics revealed problems with the "0" rating category, which subsequently unmasked problems with the item "reach and retrieve." The item received the rating "0" on 11/12 occasions when the rating was used. The rating's average person measure was disordered and exhibited a high outfit MNSQ statistic, suggesting that the rating was applied in an unexpected manner. The Rasch model expected that a rating of "0" would be applied to individuals with low levels of UE function. However, the statistical results and reinspection of the raw data indicated that the rating was unexpectedly applied to individuals with high levels of UE function. The item also showed abnormally high missing data at baseline and across the entire data set. According to documentation from raters, the missing data appeared to relate to problems administering the item. But participant factors, such as increased tone and/or spasticity, could have affected item performance. However, these factors were not directly examined in the context of this test; thus, their contribution to missing data is unknown. Based on the diagnostic results and the amount of missing data, this item did not contribute to the measurement of UE motor function in this data set and was therefore removed from all subsequent analyses. All other items showed acceptable rating scale diagnostics. The potential for measurement error as a threat to test validity is a serious challenge for designers of multisite rehabilitation research clinical trials.³⁹

Acceptable rating scale diagnostics for 14/15 items suggest that the rigorous training process and standardization procedures for testers and raters across 7 study sites successfully reduced this threat in the EXCITE trial.

Test Reliability

The 14-item WMFT exhibited excellent person separation and reliability. Items divided the sample into 5 statistically distinct levels of ability. Importantly, this is a statistical stratification to examine the measurement properties of an instrument. Clinical implications of the strata require further study. The results are consistent with the body of literature demonstrating strong WMFT psychometrics.²⁹⁻³³

Dimensionality

CFA was used to test whether the WMFT measures 1 (1-factor CFA) or 2 constructs (2-factor CFA). The results were inconclusive. On one hand, WMFT multidimensionality is supported. The 2-factor CFA showed higher CFI, larger R^2 , and slightly lower RMSEA values when compared with the 1-factor CFA. This finding suggests that the 2-factor model is the best fit. If the WMFT measures 2 constructs, then the test should be separated into 2 subscales: 1 comprising items measuring nonobject movements and another comprising items measuring object-related movements. There may be pragmatic reasons to consider this division. The most optimal measurement tool is one comprising items measuring behaviors expected to change with an intervention. Thus, an intervention intended to alter motor function at the impairment level would be best measured by nonobject movement items. Similarly, an intervention intended to alter functional UE motor skills might be best measured by object-related items.

On the other hand, an equally plausible interpretation of the data supports WMFT unidimensionality. Both CFA models exhibited high CFI and TLI statistics, large factor loadings, and moderately strong R^2 values. Taken together with the significant χ^2 model comparison test, the results suggest that either factor model could have generated these data. When 2 models fit data equally well and when both are theoretically plausible, convention dictates choosing the simplest solution,³⁶ which here is the 1-factor model. Moreover, the 2-factor CFA exhibited a strong interfactor correlation. Therefore, the WMFT probably does not measure 2 completely distinct constructs but, rather, different yet overlapping aspects of the same construct. Unidimensionality is further supported by Rasch fit statistics and PCAR. There may be pragmatic reasons to consider maintaining the unity of the WMFT. For example, research motivated to explore the relationship of nonobject reaching and object-related reaching/grasping may benefit from measuring performance of both types of tasks on a common scale.

The unidimensional interpretation of our results is supported by contemporary systems and dynamic action theoretical models of motor control, which suggest that UE movement is organized through a series of highly interconnected neural networks⁵⁰ and tailored to the performance of tasks with specific behavioral goals.^{2,15} Trombly and Wu³ and Wu et al⁴ showed that movement patterns displayed during performance of nonfunctional motor tasks were distinctly different from patterns displayed during goal-oriented motor performance. These findings could indicate separate visuomotor networks underlying control of nonspecific limb transportation (reach) and goal-oriented hand-shaping (grasp) movements.⁵¹ However, contemporary literature suggests considerable coordination between reach and grasp controllers.⁵⁰ For example, reach and grasp kinematics are concomitantly affected with altered object location or orientation⁵² and temporally coupled/concurrently modulated as the arm moves toward a target.⁵³ Recent cortical stimulation studies show highly interconnected and overlapping movement representations throughout

the motor cortex.⁵⁴ Taken together, these observations suggest an interdependence and integration of the reach and grasp neuromotor processes, which could explain the strong 2-factor CFA interfactor correlation seen here and supports WMFT unidimensionality.

What is important though is that the CFA results were not definitive. Both CFA models exhibited RMSEA values exceeding the cutoff criterion, which could signal the presence of other factor structures and/or the presence of “noise” in the data. Fit might be improved by adding factors or testing alternative models. However, to our knowledge, there are no alternative theoretically postulated WMFT factor structures. Therefore, no other solutions were considered here.

Item Difficulty Hierarchy

The analysis elucidated how WMFT items related to one another. In general, the Rasch-derived item difficulty hierarchy was consistent with original item difficulty expectations put forth by Wolf et al.²³ Items progress in control requirements from single joint to multijoint. The easiest items require control of primarily 1 joint, whereas the most difficult items require control and coordination of multiple joints. In addition, items progress in difficulty as the information processing requirements of the items increase. The easiest items involve 1-step actions, such as placing the limb at a location, whereas the more difficult items involve multistep actions such as reaching and grasping, or reaching and lifting.

Historically, the motor control literature has supported an association between task difficulty and task demand.⁶ Fitts⁵⁵ demonstrated an increase in task difficulty with increasing speed/accuracy requirements. Others have shown that increasing the amount of precision required of the action,⁵⁶ amount of intertrial movement variability,⁵⁷ number of performance steps,⁵⁸ and biomechanical limb load² amplify task difficulty. The Rasch WMFT item difficulty hierarchy is consistent with this literature. The most challenging items have high accuracy/precision requirements (eg, stack checkers, flip cards), whereas the easiest items have low accuracy/precision requirements (eg, hand to table, forearm to table).

Item—Person Match

Traditionally, the WMFT is scored by summing item ratings and reporting an overall mean or median value.²⁹ However, a mean or median score does not elucidate the interaction of individuals with specific items. In contrast, Rasch analysis calibrates person ability onto the same metric as item difficulty, thereby allowing WMFT scores (ie, person ability measures) to be interpreted with regard to the probability of being able to achieve/not achieve specific items. The map illustrates which items were harder and which items were easier for participants to accomplish. Individuals of very low ability (those with low WMFT person measures) were matched to the least difficult items at the lowest part of the rating scale. Individuals of very high ability (those with higher WMFT person measures) were matched to the most difficult items at the highest part of the rating scale.

In a recent study, Bogard and colleagues³⁴ examined the association between the magnitude of improvement on any 1 WMFT item and the change score of the whole test (ie, improved motor function) with regression methods. Their analysis of EXCITE data showed 6 tasks (hand to table, hand to box, reach and retrieve, lift can, lift pencil, and fold towel) that influenced the overall WMFT for those treated within 3 to 9 months poststroke; 6 other tasks (extend elbow weight, hand to box, lift can, lift pencil, turn key, and fold towel) influenced the overall WMFT score for those treated 1 year poststroke. The authors concluded that these items were most sensitive to improvements on the entire assessment.

The present study may suggest an alternative framework for investigating the interaction of person ability and item performance, which may affect how the WMFT score is interpreted.

The item map illustrates the concept that successful performance of more difficult items requires a greater “amount” of person ability, that is, greater UE motor function. As seen on the item map, persons with low WMFT scores were likely to exhibit control of 1 or 2 joints to position the arm on various surfaces; persons with higher scores were likely to exhibit control of multiple joints for functional task performance. The map suggests that UE motor recovery corresponds with an increased probability of successfully performing functional tasks requiring more skilled movement control.

The item map is consistent with the results of Lin et al⁵⁹ and Fritz et al⁶⁰ who found fine-motor dexterity to be an important predictor of success with constraint therapy—that is, obtaining a higher score on the WMFT. Indeed, the item map illustrates the match between individuals with more UE motor function (higher WMFT scores) and tasks requiring fine-motor control (eg, lift paper clip, lift pencil). Moreover, the item map shows that the WMFT may not be appropriate to measure individuals having less UE motor function than the sample because the test may not contain enough “easy” items.

Implications for Motor Control Theory

Testing the effects of poststroke UE interventions requires accurate measurement of motor function, yet the interpretation of an assessment score is affected by its conceptual framework. Therefore, we examined the theoretical framework of the WMFT because this test is increasingly used to measure rehabilitation outcomes. Specifically, we investigated whether the WMFT is consistent with contemporary motor control theories (eg, systems theories such as dynamical action and ecological theory) that stress the importance of understanding motor control as a dynamic process arising from the continuous interaction of the person (motor, cognitive, and perceptual processes) with the task being performed.²⁻⁷

The results suggest that the WMFT’s item structure is consistent with contemporary theories as noted throughout this discussion and summarized here. For example, the strong 2-factor CFA interfactor correlation may suggest the integration of the visuomotor neural pathways underlying reach and grasp^{2,15,50,52,54,61}; the item difficulty hierarchy may suggest a progression of task difficulty, proceeding from more simple (single joint) to more complex (multiple joint), and further may be related to the information processing requirements of the task (ie, amount of precision, number of performance steps, and biomechanical load).^{2,6,7,23,56-58}

Of course, the challenge in connecting the results to the field of motor control theory is that the present investigation is a macro-level study of human functional behavioral patterns. Although we acknowledge this limitation, the present findings provide a basis for future studies aimed at more directly investigating neural and/or biomechanical factors contributing to the person-to-task interaction. For example, ongoing electromyographic and motion analysis studies in our lab are identifying critical neuromechanical factors underlying performance of specific items. In summary, the results of this study depict an item difficulty hierarchy that may be the basis for connecting a person’s WMFT score to a location along a continuum of UE motor function and may advance the theoretical discussion of the interaction of person ability with task difficulty during recovery.

Limitations

There are several limitations to this study that affect this study’s generalizability. First, the sample size ($n = 189$) may be considered too small. Guidelines specifying appropriate CFA sample size range from 5 to 10 participants/ item⁶² to a minimum of 200 participants.⁶³ Similarly, estimates guiding appropriate Rasch sample size range from 50⁶⁴ to 200 participants.³⁷ Additionally, only 20% to 25% of the stroke population may meet EXCITE

inclusion criteria⁶⁵; therefore, the results cannot be generalized beyond the characteristics of this sample.

Acknowledgments

The authors wish to thank Megan Briggs, OTS, College of Public Health and Health Professions, Department of Occupational Therapy, University of Florida for her help with data preparation.

Funding

This material is based upon work supported by the Office of Research and Development Rehabilitation Research and Development Service, Department of Veterans Affairs (Career Development-2 Award No. B-6332W, PI: Woodbury). The EXCITE trial was funded by NIH Grant R01 HD 37606 (PI: Wolf).

References

1. Nudo R. Adaptive plasticity in motor cortex: implications for rehabilitation after brain injury. *J Rehabil Med.* 2003; 41:7–10. [PubMed: 12817650]
2. Shumway-Cook, A.; Woollacott, MH. *Motor Control: Translating Research into Clinical Practice.* 3rd ed.. Lippincott Williams & Wilkins; Philadelphia, PA: 2006.
3. Trombly CA, Wu CY. Effect of rehabilitation tasks on organization of movement after stroke. *Am J Occup Ther.* 1999; 53:333–344. [PubMed: 10427675]
4. Wu C, Trombly CA, Lin K, Tickle-Degnen L. A kinematic study of contextual effects on reaching performance in persons with and without stroke: influences of object availability. *Arch Phys Med Rehabil.* 2000; 81:95–101. [PubMed: 10638883]
5. Rosenbaum DA, Cohen RG, Dawson AM, et al. The posture-based motion planning framework: new findings related to object manipulation, moving around obstacles, moving in three spatial dimensions, and haptic tracking. *Adv Exp Med Biol.* 2009; 629:485–497. [PubMed: 19227517]
6. Scott SH, Norman KE. Computational approaches to motor control and their potential role for interpreting motor dysfunction. *Curr Opin Neurol.* 2003; 16:693–698. [PubMed: 14624078]
7. Shadmehr R, Krakauer JW. A computational neuroanatomy for motor control. *Exp Brain Res.* 2008; 185:359–381. [PubMed: 18251019]
8. Krakauer JW. Motor learning: its relevance to stroke recovery and neurorehabilitation. *Curr Opin Neurol.* 2006; 19:84–90. [PubMed: 16415682]
9. Carr, JH.; Shepherd, RB. *Movement Science: Foundations for Physical Therapy in Rehabilitation.* 2nd ed.. Aspen; Gaithersburg, MD: 2000.
10. Wolf SL, Winstein CJ, Miller JP, et al. 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. [PubMed: 17077374]
11. Lo AC, Guarino P, Krebs HI, et al. Multicenter randomized trial of robot-assisted rehabilitation for chronic stroke: methods and entry characteristics for VA ROBOTICS. *Neurorehabil Neural Repair.* 2009; 23:775–783. [PubMed: 19541917]
12. Woodbury ML, Howland DR, McGuirk TE, et al. Effects of trunk restraint combined with intensive task practice on poststroke upper extremity reach and function: a pilot study. *Neurorehabil Neural Repair.* 2009; 23:78–91. [PubMed: 18812433]
13. Taub E, Miller NE, Novack TA, et al. Technique to improve chronic motor deficit after stroke. *Arch Phys Med Rehabil.* 1993; 74:347–354. [PubMed: 8466415]
14. Howle, J. *Neuro-Developmental Treatment Approach: Theoretical Foundations and Principles of Clinical Practice.* Neuro-Developmental Treatment Association; Laguna Beach, CA: 2002.
15. Mulder T, Zijlstra W, Geurts A. Assessment of motor recovery and decline. *Gait Posture.* 2002; 16:198–210. [PubMed: 12297260]
16. Levin MF, Kleim JA, Wolf SL. What do motor “recovery” and “compensation” mean in patients following stroke? *Neurorehabil Neural Repair.* 2009; 23:313–319. [PubMed: 19118128]

17. McDowell, I.; Newell, C. The theoretical and technical foundations of health measurement. In: McDowell, I.; Newell, C., editors. *Measuring Health: A Guide to Rating Scales and Questionnaires*. University Press; New York, NY: 1996. p. 10-46.
18. Mulder T, Nienhuis B, Pauwels J. The assessment of motor recovery: a new look at an old problem. *J Electromyogr Kinesiol*. 1996; 6:137-145. [PubMed: 20719671]
19. Fugl-Meyer AR, Jaasko L, Leyman I, Olsson S, Steglind S. The post-stroke hemiplegic patient: a method for evaluation of physical performance. *Scand J Rehabil Med*. 1975; 7:13-31. [PubMed: 1135616]
20. Van Wijck FM, Pandyan AD, Johnson GR, Barnes MP. Assessing motor deficits in neurological rehabilitation: patterns of instrument usage. *Neurorehabil Neural Repair*. 2001; 15:23-30. [PubMed: 11527276]
21. Crow JL, Harmeling-van der Wel BC. Hierarchical properties of the motor function sections of the Fugl-Meyer assessment scale for people after stroke: a retrospective study. *Phys Ther*. 2008; 88:1554-1567. [PubMed: 18927197]
22. Woodbury ML, Velozo CA, Richards LG, Duncan PW, Studenski S, Lai SM. Dimensionality and construct validity of the Fugl-Meyer Assessment of the upper extremity. *Arch Phys Med Rehabil*. 2007; 88:715-723. [PubMed: 17532892]
23. Wolf SL, Lecraw DE, Barton LA, Jann BB. Forced use of hemiplegic upper extremities to reverse the effect of learned nonuse among chronic stroke and head-injured patients. *Exp Neurol*. 1989; 104:125-132. [PubMed: 2707361]
24. Morris D, Crago J, DeLuca S, Pidikiti R, Taub E. Constraint-induced (CI) movement therapy for motor recovery after stroke. *Neurorehabilitation*. 1997; 9:29-43.
25. 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. [PubMed: 10066856]
26. Blanton S, Wolf SL. An application of upper-extremity constraint-induced movement therapy in a patient with subacute stroke. *Phys Ther*. 1999; 79:847-853. [PubMed: 10479785]
27. Richards LG, Senesac CR, Davis SB, Woodbury ML, Nadeau SE. Bilateral arm training with rhythmic auditory cueing in chronic stroke: not always efficacious. *Neurorehabil Neural Repair*. 2008; 22:180-184. [PubMed: 17660456]
28. Kowalczewski J, Gritsenko V, Ashworth N, Ellaway P, Prochazka A. Upper-extremity functional electric stimulation-assisted exercises on a workstation in the subacute phase of stroke recovery. *Arch Phys Med Rehabil*. 2007; 88:833-839. [PubMed: 17601461]
29. Morris DM, Uswatte G, Crago JE, Cook EW, Taub E. The reliability of the Wolf Motor Function Test for assessing upper extremity function after stroke. *Arch Phys Med Rehabil*. 2001; 82:750-755. [PubMed: 11387578]
30. Whittall J, Savin DN Jr, Harris-Love M, Waller SM. Psychometric properties of a modified Wolf Motor Function Test for people with mild and moderate upper-extremity hemiparesis. *Arch Phys Med Rehabil*. 2006; 87:656-660. [PubMed: 16635628]
31. Wolf SL, Catlin PA, Ellis M, Archer AL, Morgan B, Piacentino A. Assessing the Wolf Motor Function Test as an outcome measure for research in patients after stroke. *Stroke*. 2001; 32:1635-1639. [PubMed: 11441212]
32. Wolf SL, Thompson PA, Morris DM, et al. The EXCITE Trial: attributes of the Wolf Motor Function Test in patients with subacute stroke. *Neurorehabil Neural Repair*. 2005; 19:194-205. [PubMed: 16093410]
33. Ang JH, Man DW. The discriminative power of the Wolf Motor Function Test in assessing upper extremity functions in persons with stroke. *Int J Rehabil Res*. 2006; 29:357-361. [PubMed: 17106357]
34. Bogard K, Wolf S, Zhang Q, Thompson P, Morris D, Nichols-Larsen D. Can the Wolf Motor Function Test be streamlined? *Neurorehabil Neural Repair*. 2009; 23:422-428. [PubMed: 19276293]
35. Daly JJ, Ruff RL. Construction of efficacious gait and upper limb functional interventions based on brain plasticity evidence and model-based measures for stroke patients. *Scientific-WorldJournal*. 2007; 7:2031-2045. [PubMed: 18167618]

36. Kline, R. Principles and Practices of Structural Equation Modeling. 2 nd ed.. Guilford; New York, NY: 2005.
37. Hambleton, RK.; Swaminathan, H.; Rogers, HJ. Fundamentals of Item Response Theory. Sage; Newbury Park, CA: 1991.
38. Winstein CJ, Miller JP, Blanton S, et al. Methods for a multisite randomized trial to investigate the effect of constraint-induced movement therapy in improving upper extremity function among adults recovering from a cerebrovascular stroke. *Neurorehabil Neural Repair*. 2003; 17:137–152. [PubMed: 14503435]
39. Morris DM, Taub E, Macrina DM, Cook EW, Geiger BF. A method for standardizing procedures in rehabilitation: use in the extremity constraint induced therapy evaluation multisite randomized controlled trial. *Arch Phys Med Rehabil*. 2009; 90:663–668. [PubMed: 19345784]
40. [Accessed April 21, 2010] Mplus Version 4.21. www.statmodel.com
41. Flora DB, Curran PJ. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol Methods*. 2004; 9:466–491. [PubMed: 15598100]
42. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling*. 1999; 6:1–55.
43. Marsh HW, Balla JR, McDonald RP. Goodness-of-fit indices in confirmatory factor analysis: the effect of sample size. *Psychol Bull*. 1988; 103:391–410.
44. WINSTEPS Rasch Measurement Computer Program. Version 3.61.2. Chicago, IL: 2006. Winsteps.com
45. Andrich D. A rating formulation for ordered response categories. *Psychometrika*. 1978; 43:561–573.
46. Linacre JM. Understanding Rasch measurement: optimizing rating scale category effectiveness. *J Appl Meas*. 2002; 3:85–106. [PubMed: 11997586]
47. Wright, BD.; Stone, MH. Best Test Design. Mesa Press; Chicago, IL: 1979.
48. Bond, TG.; Fox, CM. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. Erlbaum; Mahwah, NJ: 2001.
49. Linacre, JM. A User’s Guide to Winsteps and Ministeps RaschModel Computer Programs. Chicago, IL: 2006. Winsteps.com
50. Tretriluxana J, Gordon J, Winstein CJ. Manual asymmetries in grasp pre-shaping and transport-grasp coordination. *Exp Brain Res*. 2008; 188:305–315. [PubMed: 18437369]
51. Jeannerod M. Mechanisms of visuomotor coordination: a study in normal and brain-damaged subjects. *Neuropsychologia*. 1986; 24:41–78. [PubMed: 3517680]
52. Desmurget M, Grea H, Prablanc C. Final posture of the upper limb depends on the initial position of the hand during prehension movements. *Exp Brain Res*. 1998; 119:511–516. [PubMed: 9588786]
53. Sangole AP, Levin MF. Palmar arch dynamics during reach-to-grasp tasks. *Exp Brain Res*. 2008; 190:443–452. [PubMed: 18641977]
54. Melgari JM, Pasqualetti P, Pauri F, Rossini PM. Muscles in “concert”: study of primary motor cortex upper limb functional topography. *PLoS One*. 2008; 3:e3069. [PubMed: 18728785]
55. Fitts PM. The information capacity of the human motor system in controlling the amplitude of movement. 1954. *J Exp Psychol Gen*. 1992; 121:262–269. [PubMed: 1402698]
56. Fleishman EA. Relating individual differences to the dimensions of human tasks. *Ergonomics*. 1978; 21:1007–1019. [PubMed: 365521]
57. Gentile AM. Implicit and explicit processes during acquisition of functional skills. *Scand J Occup Ther*. 1998; 5:7–16.
58. Wood RE. Task complexity: definition of the construct. *Organ Behav Hum Decis Process*. 1986; 37:60–82.
59. Lin KC, Huang YH, Hsieh YW, Wu CY. Potential predictors of motor and functional outcomes after distributed constraint-induced therapy for patients with stroke. *Neurorehabil Neural Repair*. 2009; 23:336–342. [PubMed: 18984830]

60. Fritz SL, Light KE, Patterson TS, Behrman AL, Davis SB. Active finger extension predicts outcomes after constraint-induced movement therapy for individuals with hemiparesis after stroke. *Stroke*. 2005; 36:1172–1177. [PubMed: 15890987]
61. Sangole AP, Levin MF. A new perspective in the understanding of hand dysfunction following neurological injury. *Top Stroke Rehabil*. 2007; 14:80–94. [PubMed: 17573314]
62. Bryant, FB.; Yarnold, PR. Principal components analysis and exploratory and confirmatory factor analysis. In: Grimm, LG.; Yarnold, PR., editors. *Reading and Understanding Multivariate Analysis*. American Psychological Association; Washington, DC: 1995. p. 99-136.
63. Gorsuch, RL. *Factor Analysis*. Lawrence Erlbaum; Hillsdale, NJ: 1983.
64. Linacre JM. Sample size and item calibration stability. *Rasch Meas Trans*. 1994; 7:328.
65. Taub E, Crago J, Uswatte G. Constraint-induced movement therapy: a new approach to treatment in physical rehabilitation. *Rehabil Psychol*. 1998; 43:152–170.

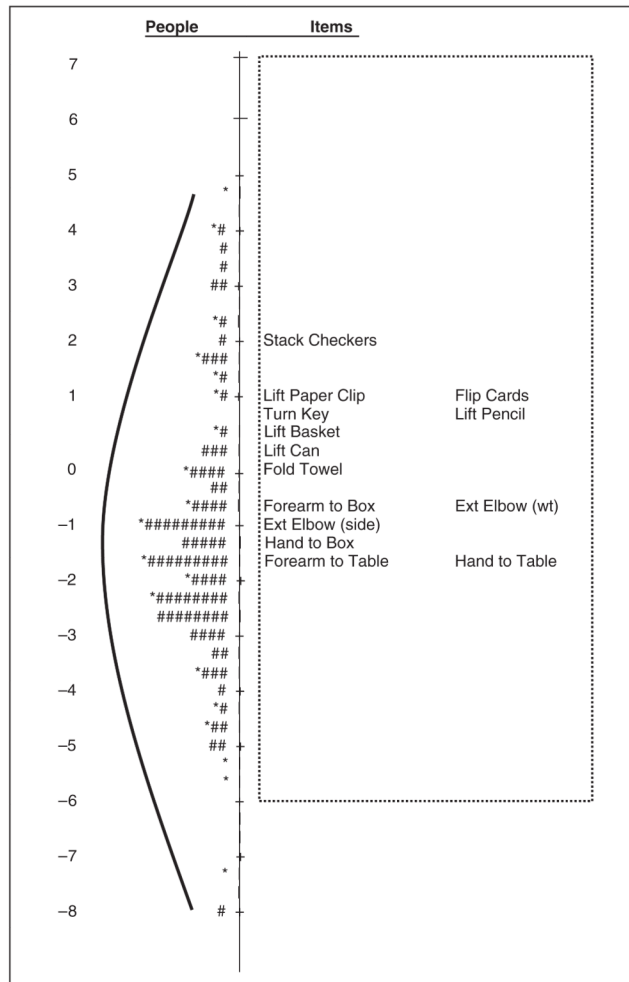


Figure 1.
 Person-item map^a
^aExt, extend; wt, weight.

Table 1

The Functional Ability Rating Scale

Rating	Description
0	Does not attempt with more-involved upper extremity (UE)
1	More-involved UE does not participate functionally; however, attempt is made to use the UE. In unilateral tasks the less-involved UE may be used to move the more-involved UE
2	Attempts to use more-involved UE but requires assistance of the less-involved UE for minor readjustments or change of position, or requires more than 2 attempts to complete, or accomplishes very slowly. In bilateral tasks, the more-involved UE may serve only as a helper
3	Attempts to use more-involved UE, but movements are influenced to some degree by synergy or are performed slowly or with effort
4	Attempts to use more-involved UE; movement is close to normal, ^a but slightly slower; may lack precision, fine coordination, or fluidity
5	Attempts to use more-involved UE; movements appear to be normal ^a

^aFor the determination of “normal,” the less-involved UE can be used as an available index for comparison, with premorbid UE dominance taken into consideration.

Table 2

Confirmatory Factor Analysis Results

	One-Factor Model	Two-Factor Model
χ^2 Value	173.06	102.53
Degrees of freedom	35	36
TLI	0.98	0.99
CFI	0.96	0.98
RMSEA	0.14	0.09
Standardized factor loading estimates \pm average SE	0.72 – 0.99 \pm 0.03	Factor 1: 0.82 – 0.99 \pm 0.02 Factor 2: 0.80 – 0.93 \pm 0.02
Model average R^2	0.68	0.75
Correlation between factors	NA	0.87

Abbreviations: TLI, Tucker Lewis Index; CFI, Comparative Fit Index; RMSEA, root mean square error of approximation; SE, standard error of the measure.

Table 3

Item Statistics: Items Are Ordered From Most Difficult (Top) to Least Difficult (Bottom)

Item	Item Difficulty Measure (logits)	Item Difficulty SE	Infit	
			MNSQ	ZSTD
Stack checkers	1.91	0.13	1.10	1.0
Flip cards	1.04	0.13	0.61	-4.5
Lift paper clip	0.99	0.13	1.10	1.0
Turn key	0.96	0.13	0.84	-1.7
Lift pencil	0.92	0.13	0.99	-0.1
Lift basket	0.81	0.13	1.38	3.3
Lift can	0.27	0.13	1.09	0.8
Fold towel	0.15	0.13	0.76	-2.4
Forearm to box	-0.61	0.14	0.84	-1.5
Extend elbow (weight)	-0.81	0.14	1.32	2.6
Extend elbow (side)	-1.04	0.14	1.44 ^a	3.5 ^a
Hand to box	-1.30	0.14	0.79	-2.0
Forearm to table	-1.51	0.14	0.81	-1.7
Hand to table	-1.77	0.14	0.62	-4.0

Abbreviations: SE, standard error of the measure; MNSQ, mean square of the item residuals; ZSTD, transformation of the mean square of the residuals to its standardized form.

^aExceeds cutoff criterion (MNSQ < 1.4, ZSTD < 2.0).