Department of Veterans Affairs

Translating measurement findings into rehabilitation practice: An example using Fugl-Meyer Assessment-Upper Extremity with patients following stroke

Craig A. Velozo, PhD, OTR/L;[1]* Michelle L. Woodbury, PhD, OTR/L[2]

[1]*Rehabilitation Outcomes Research Center and Brain Rehabilitation Research Center, Malcom Randall Department of Veterans Affairs (VA) Medical Center, Gainesville FL; and Department of Occupational Therapy, College of Public Health and Health Professions, University of Florida, Gainesville, FL; [2]Ralph H. Johnson VA Medical Center, Charleston SC; and Department of Health Sciences and Research, Medical University of South Carolina, Charleston, SC*

Abstract—Standardized assessments are critical for advancing clinical rehabilitation, yet assessment scores often provide little information for rehabilitation treatment planning. A keyform recovery map is an innovative way for a therapist to record patient responses to standardized assessment items. The form enables a therapist to view the specific items that a patient can or cannot perform. This information can assist a therapist in tailoring treatments to a patient's individual ability level. We demonstrate how a keyform recovery map can be used to inform clinical treatment planning for individuals with stroke-related upper-limb motor impairment. A keyform map of poststroke upper-limb recovery defined by items of the Fugl-Meyer Assessment-Upper Extremity (FMA-UE) was generated by a previously published Rasch analysis. Three individuals with stroke enrolled in a separate research study were randomly selected from each of the three impairment strata of the FMA-UE. Their performance on each item was displayed on the FMA-UE keyform. The forms directly connected qualitative descriptions of patients' motor ability to assessment measures, thereby suggesting appropriate shorter and longer term rehabilitation goals. This study demonstrates how measurement theory can be used to translate a standardized assessment into a useful, evidence-based tool for making clinical practice decisions.

Key words: assessment, clinical goal setting, Fugl-Meyer Assessment, item response theory, measurement, occupational therapy, Rasch analysis, rehabilitation, stroke, upper limb.

INTRODUCTION

Clinical rehabilitation therapists have been challenged, if not mandated, to use standardized assessments as part of clinical practice. Educational accreditation standards and practice frameworks for clinical rehabilitation disciplines such as physical and occupational therapy require that students learn and use standardized assessments [1–2]. More recently, the growing demand for evidence-based practice carries with it an implicit mandate to use reliable and valid standardized assessments as the basis of clinical decision making [3–6].

In spite of this foundational training and best-practice focus, standardized assessments are used infrequently. For example, a multicenter retrospective review of medical

charts showed that only 13 percent of 248 individuals with stroke were assessed for unilateral neglect with standardized assessments as recommended by practice guidelines [7]. A recent survey of 253 Canadian occupational therapists revealed that only 27 percent reported using standardized assessments of unilateral spatial neglect [8], and a qualitative interview with 12 Swedish occupational therapists working with individuals with brain injury revealed that the therapists were reluctant to use standardized assessments [9]. In a survey of 300 Canadian physiotherapists conducted by Abrams et al., 90 percent agreed that "health professionals should monitor the outcome of their treatment using reliable and valid tools," yet less than 30 percent reported routine use of standardized assessments [10]. Furthermore, Kay et al. reported that in spite of concerted educational efforts to motivate therapists to use published outcomes scales, 6 years later, less than half of the therapists incorporated published scales into practice [11]. There is also an inherent dissatisfaction with instruments that have widespread use. For example, in the 102 European stroke rehabilitation facilities surveyed by Torenbeek et al., while the majority of therapists used published scales, such as the Barthel Index or Functional Independence Measure (FIM™), more than 90 percent of the respondents reported dissatisfaction with these instruments [12].

For good reasons, we continue to promote the advancement of standardized assessment tools in evidence-based clinical rehabilitation [6]. Standardized assessments allow comparisons against norms, objectively document the effectiveness of interventions, and promote the translation of clinical research into clinical practice. The validity of rehabilitation constructs within the scientific community will likely depend on the quality of the tools used to measure these constructs [13]. Rigorous assessments are undeniably the key to the advancement of rehabilitation practice and science.

Clearly, standardized assessments have a critical role in clinical professions, and yet therapists are challenged in incorporating these assessments into day-to-day clinical practice. While educators, researchers, and professional organizations continue to criticize therapists for not using standardized assessments, we may need to contemplate whether or not standardized assessments have any immediate benefit to the practicing therapist. The critical question may be "What has my standardized assessment done for me (the therapist) lately?" In order to better understand the plight of standardized assessment

in practice, we need to reevaluate the assessment process and the role of measurement in this process.

Every measure begins with a qualitative experience [14–16]. The clinical evaluation incorporates in-depth patient interviews (e.g., narratives, ethnographies), observations of behaviors (e.g., developmental milestones, performance of self-care or instrumental activities of daily living tasks), and informal observations of patient participation (e.g., interacting with peers, community integration). It is the qualitative features of the therapist-patient interaction that lead to real-time interpretation of the patient's status and the formulation of a treatment plan tailored to the unique needs of the individual and his or her context. While standardized assessments provide a means to quantify some or all of the assessment process, many assessments do relatively little to immediately inform the therapist beyond the qualitative interaction. While administrators, healthcare accrediting agencies, payers, and of course, researchers use numbers to make critical clinical, financial, and scientific decisions, unfortunately, the numbers often add little value to the immediate clinical reasoning process.

The restricted use of assessment instruments in clinical practice may be a function of the measurement model underlying the development of these instruments, that is, classical test theory (CTT). CTT explains the observed score into two components, the true score and error score [17]. The majority of investigations using this measurement model are at the observed-score level (e.g., inter-rater reliability of the observed scores across raters or predictive validity of the observed scores relative to a gold standard measure). In essence, this model provides an understanding of the construct (e.g., upper-limb functioning) at the level of the whole instrument. Furthermore, the observed scores are assessment (test) and sample dependent. That is, an individual's observed score depends on whether he or she takes an easy or hard test, and reliability and validity values are highly dependent on the characteristics of the sample [18].

Item response theory (IRT) models may provide a window to extend the applicability of assessments in clinical practice. In contrast to CTT, IRT focuses on the item rather than the test. While extensive reviews of IRT exist [19–21], critical to this study is the following characteristic of all IRT models: the attempt to measure person ability (often referred to as a latent trait) relative to item difficulty. In contrast to CTT, the values of person ability and item difficulty are sample independent [22]. In the simplest of the IRT models, the Rasch model, there is a

probabilistic relationship of person ability to item difficulty, with individuals of low ability having a high probability of failing or getting low ratings on both the easy and hard items of an assessment and individuals of high ability having higher probability (than individuals with low ability) of passing or getting high ratings on harder items and a very high probability of passing or getting high ratings on easy items [23]. This relationship of item difficulty and person ability forms the basis of generating patient evaluation forms called keyforms that can provide immediate, useful information to the therapist.

Linacre introduced the keyform as an instantaneous measure for the cognitive construct of the FIM [24]. This form is generated from the "General Keyforms" output table produced from Rasch analysis using the Winsteps software program (Winsteps; Chicago, Illinois) [25]. The form looks similar to a survey questionnaire, with the items of the survey on one side of the form and numbers corresponding to the rating scale of each item placed on the other side of the form. While the keyform is similar to a survey questionnaire, it has two additional features. First, the items of the assessment are ordered on the basis of Rasch item-difficulty calibrations. The items progress from the easiest items on the bottom of the form to the hardest items on the top of the form. Second, the ratings for the items are not lined up in straight vertical columns as on a typical survey. Instead each rating for each item is associated with a difficulty calibration. Item ratings stairstep from left to right to correspond with increasing amounts of person ability required to perform more difficult items.

Linacre with the FIM cognitive scale [24] and, later, Kielhofner et al. with the Occupational Performance History Interview-II (OPHI-II) [26] demonstrated how these keyforms could be used for evaluating patients and immediately deriving Rasch linear measures during the assessment process. This was accomplished by circling ratings (for each item) corresponding to patients' performance (i.e., on the items of the FIM keyform) or interview responses (i.e., on the items of the OPHI-II keyform). Then, by drawing a vertical line "through the bulk of the items" [24] and intersecting the x-axis, the evaluator could immediately determine the Rasch person-ability measure of the patient. Linacre and Kielhofner et al. showed that the conversion of raw scores to interval measures could be derived without specialized software programs and statistical expertise because the keyform is generated from previous analysis of data and

can be scored in real time. Similarly, Bode et al. developed a self-scoring key (keyform) for the Galveston Orientation and Amnesia Test to identify extraordinary response patterns [27]. Like Linacre and Kielhofner et al., Bode et al. recommended use of the keyform to generate instantaneous interval measures.

A further step, beyond instantaneous measurement, is the application of the keyform to clinical treatment planning. Critical to this step is the acknowledgement that the items on an instrument represent amounts of the construct under observation. That is, the items reflect an item-difficulty hierarchical structure. Avery et al. clearly demonstrate the item-difficulty structure in their presentation of their Gross Motor Ability Estimator for the Gross Motor Function Measure-66 (GMFM-66) [28]. Because childhood motor development proceeds along predictable development milestones (e.g., head control precedes sitting, which precedes crawling, which precedes walking), the structure of the keyform acquires meaning. The researchers, using a keyform, present the example of a child scoring 60/100 on the GMFM-66 [28]. This child has consistent ratings of 3 (can complete the movement task) for items such as "lifting head to midline," "sit and lowers to prone," and "stands with arms free." In addition, this child has consistent ratings of 0 (cannot complete the movement task) for items such as "jump forward 30 cm," "up steps with no rail," and "hop 10 times left foot." The "transition zone" is where the child scores a 2 (partially completes the movement task) or 1 (initiates the movement task) on items such as "high kneeling to stand," "up 4 steps holding rail," and "walk between lines." This zone represents the emerging motor ability skills of the child. Items at this level are at the "just right challenge" level; for items far below this level, the child has a high probability of succeeding, and for items far above this level, the child has a high probability of failing.

A recent study of the Fugl-Meyer Assessment-Upper Extremity (FMA-UE) offers an opportunity to create a keyform to assist neurorehabilitation therapists in planning treatment aimed at improving upper-limb motor function in persons with stroke [29]. This assessment of stroke-related upper-limb motor impairment was chosen for investigation because it is reliable and valid and often used in rehabilitation research studies [30–32]. The 33 items of the assessment are scored on a 3-point rating scale: 0 = unable to perform, 1 = partial ability to perform, and 2 = near normal ability to perform [33]. Importantly, for the purposes of the present article, the FMA-UE is potentially

useful for rehabilitation therapists interested in documenting upper-limb motor recovery during therapy.

Woodbury et al. applied Rasch analysis to 512 participants' responses to the 33 FMA-UE items [29]. The results led to a modified 30-item FMA-UE, which was shown to be a unidimensional measure of voluntary upper-limb motor ability. Furthermore, the results arranged the FMA-UE items along a hierarchical continuum according to difficulty (i.e., from least to most difficult). The item-difficulty hierarchy is a visual representation of the overall construct the assessment is intended to measure and describes a progression of this construct, i.e., defines a continuum of behaviors representing poststroke upper-limb motor recovery. The item-difficulty hierarchy challenged the traditional motor control conceptual framework underlying the FMA-UE. That is, recovery of poststroke upper-limb movement did not proceed in a strict synergistic-to-isolated and reflexive-to-integrated sequence as would be expected by the traditional framework. Instead, the Rasch-generated item-difficulty hierarchy was consistent with contemporary motor control evidence that suggests an individual's ability to perform a given arm or hand movement is a dynamic interaction between neural factors and the task-specific difficulty of the movement.

The Rasch analysis of this sample produced a keyform. Importantly, the FMA-UE item-difficulty hierarchy was consistent across 98 percent of a 512-person sample at a single time point and was invariant in a 377-person sample across two testing occasions [29,34]. These findings strongly suggest that the FMA-UE keyform can be applied to a larger population of individuals with stroke.

The purpose of this article is to extend previous research by demonstrating how the FMA-UE keyform can be used to inform treatment planning. The present study will show how patient performance documented on a keyform can be immediately used in combination with other patient-specific information to set goals and plan treatment.

METHODS

The keyform scoring form presented in this article was generated in a previous published analysis of FMA-UE data obtained from 512 individuals, aged 69.8 ± 11.1 years. These individuals were between 0 and 145 days from their first mildly to moderately severe cortical stroke. Detailed sample demographics are presented elsewhere [29]. This keyform scoring form is illustrated in **Figure 1**. FMA-UE items are listed to the right of the figure in descending diffi-

culty order (top to bottom). The rating scale for each item (0, 1, and 2) is to the left of the figure. As can be seen, the rating scale stair-steps from the left to the right. Easier items (bottom of the figure) have ratings to the far left of the figure, and harder items (top of the figure) have ratings close to the right of the figure. This stair-stepping corresponds to the difficulty of each item. That is, the ratings for each item are placed relative to the measurement scale (in log-odd units [logits]) at the bottom of the figure, which estimates the patient's upper-limb motor ability. For example, for elbow flexion, the rating of 1 corresponds to approximately –2.0 logits, while for wrist circumduction the rating of 1 corresponds to approximately 1.0 logits. The logit is an interval-level unit of measurement that represents the log-odds ratio of the probability that an individual will successfully or unsuccessfully accomplish an item at a particular part of the rating scale [23].

In order to demonstrate the application of the keyform for clinical practice, we used the keyform scoring form constructed in the Woodbury et al. Rasch analysis [29] to display FMA-UE data from three patients with stroke enrolled in a separate prospective research study in our laboratory. A total of 55 participants were recruited from the database of the Brain Rehabilitation Research Center, a Department of Veterans Affairs Rehabilitation Research and Development Center of Excellence. Subjects were included in the study if they (1) had experienced a single, unilateral ischemic stroke at least 3 months prior; (2) demonstrated passive range of motion in affected shoulder, elbow, and wrist within functional limits; and (3) were 18 to 90 years of age. Potential subjects were excluded if they were unable to understand three-step directions; had a demonstrated orthopedic condition, pain, or impaired corrected vision that would alter the kinematics of reaching; or had experienced a brain stem or cerebellar stroke. All participants were administered the FMA-UE by trained evaluators according to standardized procedures outlined by Duncan et al. [30]. The sample was divided into three categories of upper-limb motor impairment (mild, moderate, severe) based on the aggregate FMA-UE score. Commonly used FMA-UE cutoff scores defined each category: 0 to 20 severe, 21 to 50 moderate, and 51 to 66 mild [33]. For the purposes of this article, one participant was randomly selected from each impairment level.

RESULTS

Data from the three randomly selected individuals is displayed in **Figures 2–4**. The participant's ratings for each

Rating Scale			Item Description
0	1	2	Wrist Circumduction
0	1	2	Hook Grasp
0	1	2	Shoulder Flexion to 180°, Elbow Extended
0	1	2	Spherical Grasp
0	1	2	Lateral Prehension
0	1	2	Wrist Flexion/Extension, Elbow Extended
0	1	2	Pronation-Supination, Elbow Extended
0	1	2	Wrist Stable, Elbow Extended
0	1	2	Movement with Normal Speed
0	1	2	Forearm Supination
0	1	2	Shoulder Abduction to 90°, Elbow Extended
0	1	2	Movement Without Dysmetria
0	1	2	Shoulder External Rotation
0	1	2	Wrist Stable, Elbow at 90°
0	1	2	Wrist Flexion/Extension, Elbow at 90°
0	1	2	Palmar Prehension
0	1	2	Scapular Retraction
0	1	2	Pronation-Supination, Elbow at 90°
0	1	2	Shoulder Flexion to 90°, Elbow Extended
0	1	2	Hand to Lumbar Spine
0	1	2	Shoulder Abduction
0	1	2	Elbow Extension
0	1	2	Forearm Pronation
0	1	2	Movement Without Tremor
0	1	2	Cylindrical Grasp
0	1	2	Finger Mass Extension (relaxation of flexion)
0	1	2	Scapular Elevation
0	1	2	Finger Mass Flexion
0	1	2	Shoulder Adduction with Internal Rotation
0	1	2	Elbow Flexion
−6 −4 −2 0 2 4 6			Measure (Logits)

Figure 1.
Keyform scoring form.

item were circled on the keyform. **Figure 2**, representing data from a patient with severe upper-limb motor impairment (FMA-UE score 19/60), shows near normal performance on the easiest items (bottom of figure), with progressively poorer performance as the items become more difficult. That is, the patient receives a 2 (near normal performance) on the four easiest items (elbow flexion, shoulder adduction with internal rotation, finger mass flexion, and scapular elevation) and a mixture of 0, 1, or 2 ratings (unable to perform, partial performance, and near normal performance, respectively) on the next 14 items (e.g., forearm pronation, elbow extension, and pronation/supination with elbow at 90°), and a 0 (unable to perform) on the 12 most difficult items (e.g., shoulder flexion to 180°, hook grasp, and wrist circumduction). The area boxed by the dashed line in **Figure 2** illustrates the transition zone, in which the patient's performance fluctuates between unable to perform (rating = 0), partial performance (rating = 1), and near normal performance (rating = 2).

It is important to note that decisions about where to draw the transition zone boundaries (i.e., edges of the dotted box) are based on the pattern of the patient's unique responses to the items. The boundary reflects the point at which a patient begins to transition from one rating scale category to the next rating scale category. Since the rating of any item is based on a probability, the edges of the transition zone are equivocal. However, the overall response pattern provides the necessary information for treatment planning.

Figures 3–4 present data from patients with moderate and mild upper-limb motor impairment, respectively (**Figure 3**: FMA-UE score 36/60, and **Figure 4**: FMA-UE score 56/60). Relative to the "severe" patient (**Figure 2**), the individual with moderate upper-limb motor impairment (**Figure 3**) shows near normal performance (rating = 2) on five of the six easiest items and partial performance (rating = 1) on four of the nine most difficult items. The area boxed by the dashed line in **Figure 3** (i.e., items from movement without tremor to forearm supination) illustrates the transition zone, in which the patient's performance fluctuates between partial performance (rating = 1) and near normal performance (rating = 2). Relative to both the severe and moderate patients, the individual with mild upper-limb motor impairment (**Figure 4**) shows near normal performance (rating = 2) on the vast majority of the assessment
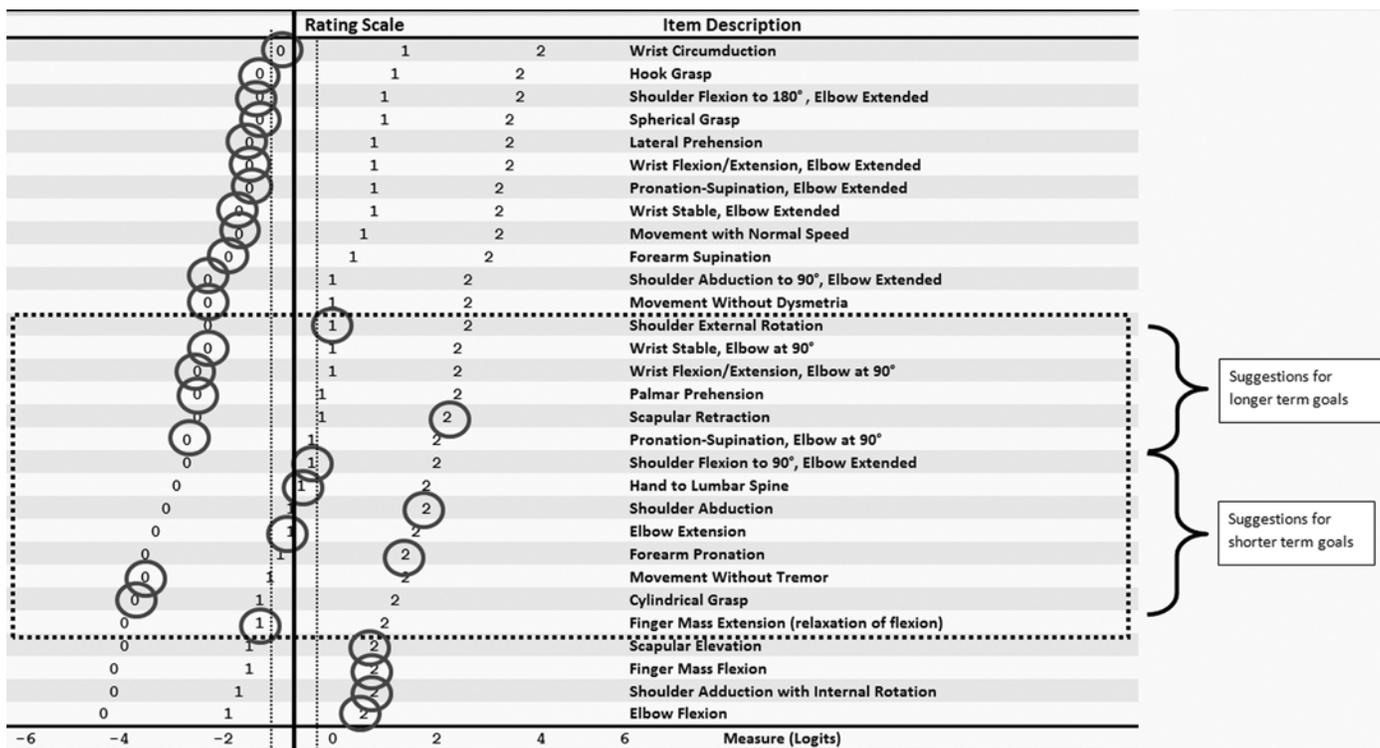
Rating Scale | **Item Description**

Item Description
Wrist Circumduction
Hook Grasp
Shoulder Flexion to 180°, Elbow Extended
Spherical Grasp
Lateral Prehension
Wrist Flexion/Extension, Elbow Extended
Pronation-Supination, Elbow Extended
Wrist Stable, Elbow Extended
Movement with Normal Speed
Forearm Supination
Shoulder Abduction to 90°, Elbow Extended
Movement Without Dysmetria
Shoulder External Rotation
Wrist Stable, Elbow at 90°
Wrist Flexion/Extension, Elbow at 90°
Palmar Prehension
Scapular Retraction
Pronation-Supination, Elbow at 90°
Shoulder Flexion to 90°, Elbow Extended
Hand to Lumbar Spine
Shoulder Abduction
Elbow Extension
Forearm Pronation
Movement Without Tremor
Cylindrical Grasp
Finger Mass Extension (relaxation of flexion)
Scapular Elevation
Finger Mass Flexion
Shoulder Adduction with Internal Rotation
Elbow Flexion

Suggestions for longer term goals

Suggestions for shorter term goals

−6   −4   −2   0   2   4   6   Measure (Logits)

Figure 2.
Patient (C31) with severe upper-limb motor impairment (Fugl-Meyer Assessment-Upper Extremity score = 19/60, Rasch Ability Measure = −1.14 ± 0.33 logits).

items, with partial performance (rating = 1) on only the four most difficult items. The area boxed by the dashed line in **Figure 4** illustrates the transition zone, in which the patient's performance fluctuates between partial performance (rating = 1) and near normal performance (rating = 2).

We should note that this overall scoring pattern is based on the probability of a patient's response to items. That is, there is not an absolute pattern of first passing all the easy items, then partially passing all the moderately difficult items, and then failing the next set of all the most difficult items. For example, the patient with severe upper-limb motor impairment (**Figure 2**) was unable to perform several easier items (e.g., cylindrical grasp, movement without tremor) while being successful or partially successful on several more challenging items (e.g., scapular retraction, shoulder external rotation). Similarly, the patient with moderate upper-limb motor impairment (**Figure 3**) fluctuated between partial and near normal performance on the easiest items, then scored a 0 (unable to perform) on a relatively easy item (e.g., movement without tremor).

The linear measure of each person can be estimated in each of the figures. Linacre suggested drawing a vertical line through the "bulk" of the circled ratings [24]. For example, for the patient with severe upper-limb motor impairment (**Figure 2**), the majority of the ratings are unable to perform (rating = 0) and then fluctuate between partial performance (rating = 1) and near normal performance (rating = 2). For the patient with moderate upper-limb motor impairment (**Figure 3**), the bulk of the ratings are between partial performance level (rating = 1) and near normal performance level (rating = 2). For the patient with mild upper-limb motor impairment (**Figure 4**), the bulk of the ratings are at the near normal performance level (rating = 2). In **Figures 2–4**, the solid vertical line represents the actual person-ability measure as derived by Rasch analysis and the dotted vertical lines represent the 95 percent confidence interval surrounding that measure.

DISCUSSION

The purpose of this article was to demonstrate how the keyform recovery map, which was derived from the Rasch analysis of a standardized assessment, could provide
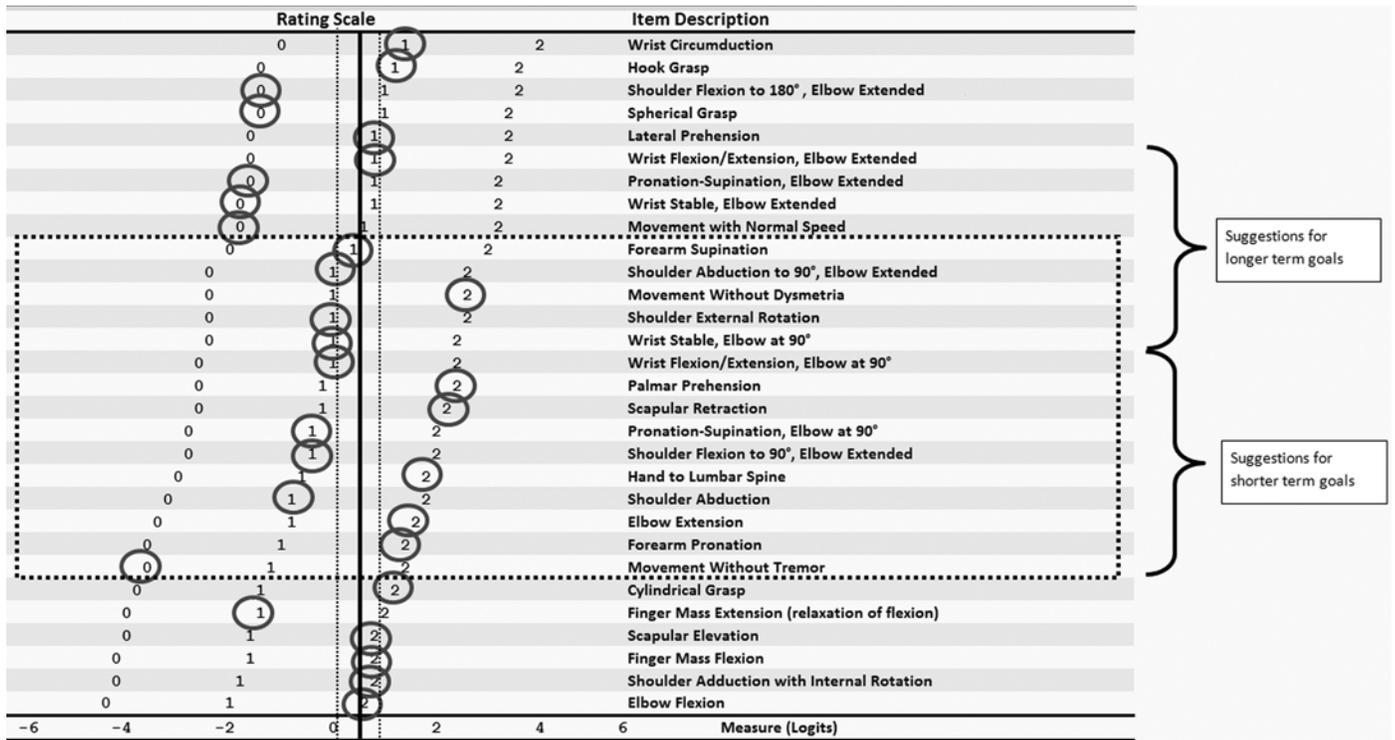
Figure 3.
Patient (C25) with moderate upper-limb motor impairment (Fugl-Meyer Assessment-Upper Extremity score = 36/60, Rasch Ability Measure = 0.61 ± 0.32 logits).

useful information for the practitioner. The FMA-UE keyform maps the relationship between the ratings of a patient's upper-limb motor performance and the items of the FMA-UE. Item ratings show a pattern of what the patient can do, can partially do, and cannot do. The pattern of item ratings reveals a transition zone, in which the patient's performance fluctuates between two ratings, therefore, pointing the way toward appropriate upper-limb motor recovery therapy goals. Finally, the keyform offers a means to estimate linear measures of individual patients' upper-limb motor ability.

The pattern of item responses is probabilistic, not deterministic. That is, patients do not pass all the easy items, then partially pass all the moderate items, and then fail all the most difficult items. As can be seen in the figures, a response to any given item sometimes does not fit the overall pattern. These findings support using the probabilistic Rasch model versus the deterministic Guttman scale model [18]. These findings also support the observations that variability exists in human motor performance. That is, while the motor pattern across individuals has a consistent pattern (i.e., easy items to hard items),

variability in an individual's performance reflects real-life clinical presentation. In contrast to a total score, the keyform preserves both the consistency (pattern) and uniqueness (variability) of an individual's motor performance.

One of the advantages of applying the FMA-UE keyform to clinical practice is that the item-difficulty hierarchy is empirically derived. Woodbury et al. demonstrated that 98 percent of their 512-patient sample supported this item hierarchical structure, and this hierarchy remained stable across two time points in a 377-patient sample [29,34]. These findings strongly suggest that the keyform can be applied with confidence to the larger population of individuals with mild to moderate cortical stroke. From a clinical perspective, these findings indicate that the FMA-UE keyform can be used as an evidence-based method to assist the therapist in planning treatment and monitoring patient progress.

Ratings plotted on the FMA-UE keyform show a transition zone occurring above easier items on which patients show near normal performance (rating = 2) and below more difficult items on which the patients show partial performance (rating = 1) or the inability to perform
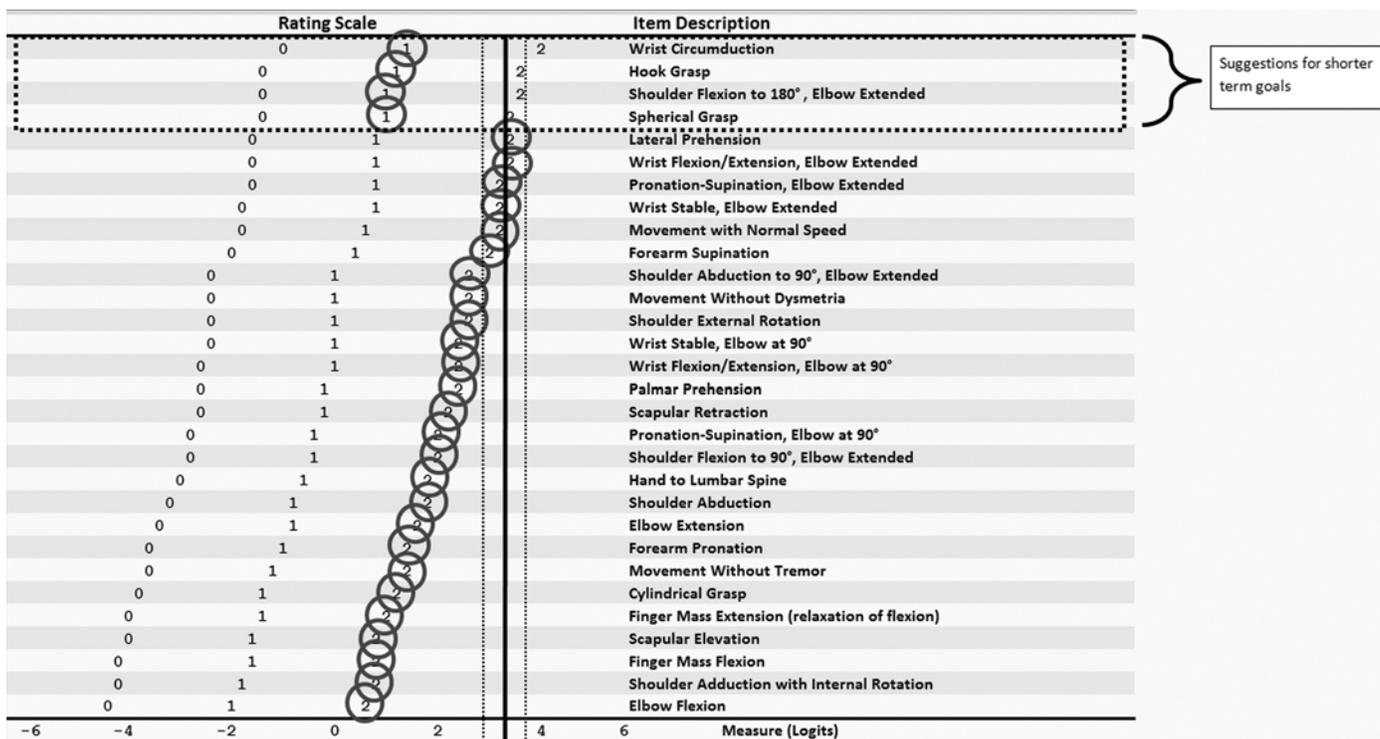
Figure 4.
Patient (C17) with mild upper-limb motor impairment (Fugl-Meyer Assessment-Upper Extremity score = 56/60, Rasch Ability Measure = 3.55 ± 0.55 logits).

(rating = 0). This transition zone can be the basis for goal setting and treatment planning. Because items in the transition zone are at the "just right challenge" level for the individual, the position of items in relationship to this zone reflect the patient's expected next steps in the poststroke upper-limb motor recovery process. Motor behaviors (items) within the transition zone may suggest appropriate shorter term functional goals, while motor behaviors (items) above the transition zone may form the foundation for appropriate longer term functional goals.

For example, in **Figure 3**, the patient is transitioning between partial and near normal ability for motor behaviors (items) such as elbow extension, shoulder flexion to 90° with elbow extended, and shoulder external rotation. Reasonable short-term goals would be directed at functional activities involving reaching away from the body, such as bathing (washing/drying body parts), dressing (obtaining clothing from storage area), and/or job performance (completing work-related desktop activities).

Figure 3 shows that this patient is having more difficulty with shoulder abduction to 90° with elbow extended, wrist flexion/extension with elbow extended, and wrist cir-

cumduction. Reasonable long-term goals would be directed at functional activities involving more extreme reach and the coordinated use of multiple joints, such as dressing (fastening and adjusting clothing and shoes), home management (yard/garden work), and shopping (selecting and purchasing items). These treatment goals are by no means intended to be exhaustive. In the hands of an innovative rehabilitation therapist with excellent task-analysis skills, the keyform can provide a framework for myriad patient-tailored goals.

While the ideal of research is to translate findings into practice, this lofty goal is often inhibited by the scientific methods available. Rasch methodologies present an innovative way to think about our standardized assessments. For the first time, a methodology is available that connects assessment scores and ratings to the qualitative content of an instrument. This connection not only allows us to reevaluate traditional expectations of poststroke upper-limb recovery but also maps recovery as a progression of item difficulties. A premise of this article is that standardized assessments have offered the practicing therapist little benefit for making day-to-day clinical practice

decisions. The keyform recovery map provides a framework to facilitate goal setting and treatment planning by restructuring the assessment using Rasch analysis. Efforts such as these may show promise for closing the gap between research and practice in rehabilitation.

LIMITATIONS

There are several limitations to this study. First, regarding instantaneous measurement, as proposed by Linacre and Kielhofner et al., keyforms have considerable error in determining person measures, especially for individuals at the extremes of the distribution. For example, for the individual with mild upper-limb motor impairment (**Figure 4**), using Linacre's suggestion for determining a person's measure [24] by placing a vertical line through the "bulk" of the responses, the rater would likely place the vertical line to the left of the actual measure (left of the vertical line depicted in the figure), possibly outside of the 95 percent confidence interval (dotted vertical lines). This imprecision of the keyform is a function of increased error at the extremes of the instrument (i.e., at the ceiling and floor), especially when attempting to measure individuals of very high or very low ability. Relative to the individual depicted in **Figure 4**, if the instrument had more difficult items (i.e., items for which the individual showed partial performance), a more accurate placement of the vertical line and determination of the individual's measure could be achieved.

In addition, as with all statistical values, the location of the demarcations of the transition zone and decisions for shorter term and longer term goals is imprecise (i.e., cannot be specified to an individual item on the item hierarchy). Moreover, the previous suggestions for interventions are hypothetical. That is, these suggested interventions have not been applied to clinical populations within clinical practice. A promising clinical study would be to determine whether an intervention generated through keyforms is more or less effective than traditionally derived clinical interventions.

CONCLUSIONS

In summary, keyforms provide an innovative way to immediately apply the findings from a standardized assessment to clinical practice. This methodology incorporates evidence-based practice (e.g., empirically derived FMA-UE item-difficulty hierarchy) with state-of-the-art measurement theory (i.e., IRT). We should note that the demonstration given here is not intended to suggest that clinically useful keyforms can be generated from any assessment. The FMA-UE is a well-developed, psychometrically sound instrument that supports a logical item-difficulty hierarchy. Standardized assessments with these characteristics, or that are designed with these characteristics in mind, should serve as candidates for creating keyforms that will support day-to-day clinical practice.

CLINICAL MESSAGE

- Evaluation forms, such as keyforms, may assist therapists in making day-to-day clinical decisions.
- Keyforms can be created using IRT statistical methods, which connect the score to patients' performance on specific items.
- Well-developed standardized assessments, especially those with a logical item-difficulty structure, may be promising candidates for generating keyforms.

ACKNOWLEDGMENTS

REFERENCES

1. American Physical Therapy Association. Guide to physical therapist practice. Second edition. American Physical Therapy Association. Phys Ther. 2001;81(1):9–746.
[PMID: 11175682]

2. Occupational therapy practice framework: Domain and process. Am J Occup Ther. 2002;56(6):609–39.
[PMID: 12458855]
http://dx.doi.org/10.5014/ajot.56.6.609

3. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: What it is and what it isn't. BMJ. 1996;312(7023):71–72. [PMID: 8555924]

4. Salter K, Jutai JW, Teasell R, Foley NC, Bitensky J, Bayley M. Issues for selection of outcome measures in stroke rehabilitation: ICF activity. Disabil Rehabil. 2005;27(6):315–40.
[PMID: 16040533]
http://dx.doi.org/10.1080/09638280400008545

5. Bennett S, Bennett JW. The process of evidence-based practice in occupational therapy: Informing clinical decisions. Aust Occup Ther J. 2000;47(4):171–80.
http://dx.doi.org/10.1046/j.1440-1630.2000.00237.x

6. Law M. Evidence-based rehabilitation: A guide to practice. Thorofare (NJ): Slack; 2002.

7. Menon-Nair A, Korner-Bitensky N, Wood-Dauphinee S, Robertson E. Assessment of unilateral spatial neglect post stroke in Canadian acute care hospitals: Are we neglecting neglect? Clin Rehabil. 2006;20(7):623–34.
[PMID: 16894806]
http://dx.doi.org/10.1191/0269215506cr974oa

8. Menon-Nair A, Korner-Bitensky N, Ogourtsova T. Occupational therapists' identification, assessment, and treatment of unilateral spatial neglect during stroke rehabilitation in Canada. Stroke. 2007;38(9):2556–62. [PMID: 17673707]
http://dx.doi.org/10.1161/STROKEAHA.107.484857

9. Holmqvist K, Kamwendo K, Ivarsson AB. Occupational therapists' descriptions of their work with persons suffering from cognitive impairment following acquired brain injury. Scand J Occup Ther. 2009;16(1):13–24.
[PMID: 18609240]
http://dx.doi.org/10.1080/11038120802123520

10. Abrams D, Davidson M, Harrick J, Harcourt P, Zylinski M, Clancy J. Monitoring the change: Current trends in outcome measure usage in physiotherapy. Man Ther. 2006;11(1):46–53. [PMID: 15886046]
http://dx.doi.org/10.1016/j.math.2005.02.003

11. Kay T, Myers A, Huijbregts M. How far have we come since 1992? A comparative survey of physiotherapists' use of outcome measures. Physiother Can. 2001;53:268–75.

12. Torenbeek M, Caulfield B, Garrett M, Van Harten W. Current use of outcome measures for stroke and low back pain rehabilitation in five European countries: First results of the ACROSS project. Int J Rehabil Res. 2001;24(2):95–101.
[PMID: 11421397]
http://dx.doi.org/10.1097/00004356-200106000-00002

13. Corcoran M. Defining and measuring constructs. Am J Occup Ther. 2007;61(1):7–8.

14. Thurstone LL. A method of scaling psychological and educational tests. J Education Psychol. 1925;16(7):433–51.
http://dx.doi.org/10.1037/h0073357

15. Thurstone LL. The scoring of individual performance. J Education Psychol. 1926;17(7):446–57.
http://dx.doi.org/10.1037/h0075125

16. Bezruczko N. Faulty thinking by educational researchers. Rasch Meas Trans. 1990;4(3):114–15.

17. Hambleton RK, Jones RW. Comparison of classical test theory and item response theory and their applications to test development. Education Meas Issues Prac. 1993;12(3):38–47.
http://dx.doi.org/10.1111/j.1745-3992.1993.tb00543.x

18. Nunnally JC, Bernstein IH. Psychometric theory. 3rd ed. New York (NY): McGraw-Hill; 1994.

19. Fayers PM. Applying item response theory and computer adaptive testing: The challenges for health outcomes assessment. Qual Life Res. 2007;16 Suppl 1:187–94.
[PMID: 17417722]
http://dx.doi.org/10.1007/s11136-007-9197-1

20. Jette AM, Haley SM. Contemporary measurement techniques for rehabilitation outcomes assessment. J Rehabil Med. 2005;37(6):339–45. [PMID: 16287664]
http://dx.doi.org/10.1080/16501970500302793

21. Reise SP, Waller NG. Item response theory and clinical measurement. Annu Rev Clin Psychol. 2009;5:27–48.
[PMID: 18976138]
http://dx.doi.org/10.1146/annurev.clinpsy.032408.153553

22. 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(1):83–90. [PMID: 9926224]

23. Wright BD, Stone MH. Best test design. Chicago (IL): Mesa Press; 1979.

24. Linacre JM. Instantaneous measurement and diagnosis. In: Smith RM, editor. Physical medicine and rehabilitation state of the art reviews. Vol 11: Outcome measurement. Philadelphia (PA): Hanley & Belfus; 1997. p. 315–24.

25. Winsteps Rasch measurement computer program [computer program]. Version 3.61.2. Chicago (IL): Winsteps.com; 2006.

26. Kielhofner G, Dobria L, Forsyth K, Basu S. The construction of keyforms for obtaining instantaneous measures from the Occupational Performance History Interview rating scales. OJTR. 2005;25(1):(23–32).

27. 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(1):637–55. [PMID: 10745181]
http://dx.doi.org/10.1097/00001199-200002000-00004

28. 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(5):697–705. [PMID: 12736885]

29. 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(6):715–23. [PMID: 17532892]
http://dx.doi.org/10.1016/j.apmr.2007.02.036

30. Duncan PW, Propst M, Nelson SG. Reliability of the Fugl-Meyer Assessment of sensorimotor recovery following cerebrovascular accident. Phys Ther. 1983;63(10):1606–10. [PMID: 6622535]

31. Gladstone DJ, Danells CJ, Black SE. The Fugl-Meyer Assessment of motor recovery after stroke: A critical review of its measurement properties. Neurorehabil Neural Repair. 2002;16(3):232–40. [PMID: 12234086]
http://dx.doi.org/10.1177/154596802401105171

32. Van Wijck FM, Pandyan AD, Johnson GR, Barnes MP. Assessing motor deficits in neurological rehabilitation: Patterns of instrument usage. Neurorehabil Neural Repair. 2001; 15(1):23–30. [PMID: 11527276]
http://dx.doi.org/10.1177/154596830101500104

33. Fugl-Meyer AR, Jääskö L, Leyman I, Olsson S, Steglind S. The post-stroke hemiplegic patient. 1. A method for evaluation of physical performance. Scand J Rehabil Med. 1975; 7(1):13–31. [PMID: 1135616]

34. Woodbury ML, Velozo CA, Richards LG, Duncan PW, Studenski S, Lai SM. Longitudinal stability of the Fugl-Meyer Assessment of the upper extremity. Arch Phys Med Rehabil. 2008;89(8):1563–69. [PMID: 18674991]
http://dx.doi.org/10.1016/j.apmr.2007.12.041