

## Translating measures across the continuum of care: Using Rasch analysis to create a crosswalk between the Functional Independence Measure and the Minimum Data Set

Craig A. Velozo, PhD, OTR/L;<sup>1-2\*</sup> Katherine L. Byers, PhD, CRC;<sup>1</sup> Ying-Chih Wang, PhD, OTR;<sup>2</sup> Bryttnee Roberts Joseph, MHS, OTR/L<sup>2</sup>

<sup>1</sup>Rehabilitation Outcomes Research Center, Department of Veterans Affairs Health Services Research and Development Service and Rehabilitation Research and Development Service Center of Excellence, North Florida/South Georgia Veterans Health System, Gainesville, FL; <sup>2</sup>Department of Occupational Therapy, University of Florida, Gainesville, FL

**Abstract**—Setting-specific outcome measures present a major barrier to monitoring patient progress across the continuum of care. This study demonstrated Rasch analysis for the creation of a crosswalk between the Functional Independence Measure (FIM), which is used in inpatient rehabilitation, and the Minimum Data Set (MDS), which is used in skilled nursing facilities. To create the crosswalk, we used data from a sample of 236 patients from four Department of Veterans Affairs' facilities who had had both the FIM and the MDS administered within 7 days. The combined FIM-MDS analysis showed good internal consistency (Cronbach alpha = 0.94), with 21 of the 26 items showing acceptable fit statistics. FIM and MDS raw scores correlated at  $-0.81$  and the measures, corrected for scale direction, correlated at  $0.78$ . Future validity testing will be necessary to determine the accuracy and applicability of the crosswalk.

**Key words:** activities of daily living, crosswalk table, Functional Independence Measure, item response theory, measure, Minimum Data Set, outcomes assessment, psychometrics, Rasch analysis, rehabilitation.

### INTRODUCTION

The Veterans Health Administration (VHA) operates the largest healthcare system in the United States [1]. To evaluate the structure, processes, and outcomes of VHA

services, we must develop integrated information systems that measure functional status across the continuum of care (acute, postacute, and community care settings). While adoption of a common patient assessment tool across these settings would provide the best mechanism for the seamless monitoring of patient outcomes, the adoption of a common postacute assessment instrument has met with serious resistance [2–3]. Presently, assessment tools used in the various settings differ, and the results from these tools are not easily comparable. For inpatient rehabilitation facilities, the Functional Independence Measure (FIM<sup>TM</sup>) is the gold standard,

**Abbreviations:** ADL = activities of daily living, CTT = classical test theory, DIF = differential item functioning, FIM = Functional Independence Measure, FSOD = Functional Status and Outcomes Database, IRT = item response theory, MDS = Minimum Data Set, MDS-PAC = MDS-Post Acute Care, PECS = Patient Evaluation Conference System, SD = standard deviation, UDSMR = Uniform Data System for Medical Rehabilitation, VA = Department of Veterans Affairs, VAMC = VA medical center, VHA = Veterans Health Administration.

\*Address all correspondence to Craig A. Velozo, PhD, OTR/L; North Florida/South Georgia Veterans Health System, 1601 SW Archer Road (151-B), Gainesville, FL 32608-1197; 352-376-1611, ext 4925; fax: 352-271-4540. Email: [cvelozo@phhp.ufl.edu](mailto:cvelozo@phhp.ufl.edu)

DOI: 10.1682/JRRD.2006.06.0068

while the Minimum Data Set (MDS) is federally mandated for skilled nursing facilities in the private sector. Both of these instruments include activities of daily living (ADL) items (eating, grooming, dressing, and walking, etc.), yet the specific subsets of ADL items, the rating scales, and the instructions for scoring each activity differ by setting.

The Uniform Data System for Medical Rehabilitation (UDSMR) is the most widely used clinical database for assessing rehabilitation outcomes [4–5]. The VHA subscribes to the UDSMR, and their Functional Status and Outcomes Database (FSOD) includes all required elements of the UDSMR. The FSOD, which has been operational since 1997, was mandated for use by VHA Directive 2000-016, “Medical Rehabilitation Outcomes for Stroke, Traumatic Brain Injury, and Lower Extremity Amputation Patients” [6]. This directive requires every VHA medical center to assess functional status, enter data into the FSOD, and measure and track rehabilitation outcomes on all stroke, lower-limb amputation, and traumatic brain injury patients [6]. The core of the UDSMR and the FSOD is the FIM. The FIM consists of 18 items, 13 of which are basic ADL (eating, grooming, bathing, dressing upper body, dressing lower body, toileting, bladder management, bowel management, bed/chair/wheelchair transfer, toilet transfer, tub/shower transfer, walk/wheelchair, and stairs) with the remaining 5 based on cognitive skills (comprehension, expression, social interaction, problem solving, and memory). The FIM is presently the core of Medicare’s prospective payment system for inpatient rehabilitation facilities [7].

## Background

While the FIM is extensively used in rehabilitation settings, the MDS of the nursing home Resident Assessment Instrument is exclusively used in skilled nursing facilities for monitoring patient function and healthcare status. The Omnibus Budget Reconciliation Act of 1987 federally mandated that all nursing homes in the United States complete the MDS for Medicare prospective payment reimbursement [8]. The MDS consists of 284 items that were designed to assess the cognitive, behavioral, functional, and medical status of nursing home residents [9]. Based on past research and clinical expertise, these items were developed to reflect important indicators of the status of nursing home residents and their quality of care [10]. The MDS is administered within 14 days of a patient’s admission to the nursing home and each quarter thereafter (approximately every 90 days) or when a relevant change in a patient’s condition occurs [11].

Psychometric studies of the FIM and MDS support the use of both measures in research. Stineman et al. performed a factor analysis of the FIM that identified motor activity and cognitive skills dimensions across 20 impairment categories [12]. Additionally, they found that the internal reliability for the FIM subscales ranged from 0.86 to 0.97, with all subscales exceeding the minimum criterion for discriminate validity. The results of the factor analysis were supported by an earlier Rasch analysis of the FIM that revealed two constructs, one representing a motor dimension and the other representing a cognitive skills dimension [13]. In a meta-analysis of 11 studies, Ottenbacher et al. demonstrated that the median interrater reliability for the total FIM was 0.95 and the test-retest and equivalence reliability of the FIM were 0.95 and 0.92, respectively [14].

While the MDS is not as extensively studied as the FIM, research on the MDS suggests that it too has adequate reliability and validity for use in research studies. Early studies by Hawes et al. showed that MDS items met a standard for excellent reliability, i.e., had interclass correlations of 0.7 or higher in key areas of functional status, such as motor activities and cognitive skills [15]. Sixty-three percent of the items achieved reliability coefficients of 0.6 or higher, and 89 percent were 0.4 or higher. Some researchers have suggested that the above psychometric findings on the MDS were “inflated” because research staff members rather than clinicians administered the MDS [16]. Yet, Stineman and Maislin’s confirmatory factor analysis study derived from clinical/administrative databases of the MDS confirmed all MDS domain clusters except social quality [16]. In addition, Lawton et al. found that MDS domain scores from a clinical/administrative database correlated well with a variety of independently obtained measures of basic behavioral and mental health functions [17]. For example, the MDS ADL scale correlated at 0.79 with the Lawton and Brody Physical Self-Maintenance Scale.

One means of improving clinical care for veterans is to develop an effective and efficient method of tracking and evaluating functional status changes as veterans move across rehabilitation and skilled nursing care settings. Since the FIM is used in rehabilitation settings and the MDS is used in skilled nursing facilities, a crosswalk or translation table could be developed across these setting-specific instruments so that a score obtained from one instrument could be compared with one from the other instrument. The basis of the crosswalk methodology

proposed in this study is the similarity of the item content of the FIM and MDS. **Table 1** presents a comparison of the motor items of the FIM and the MDS. Despite ardent debates over the value of one instrument over the other, surprising similarity is found between the instruments. The foundation of this study is the hypothesis that the items of the FIM and MDS are subsets of items along a single motor construct. If this relationship is supported, we can create a crosswalk that links the two instruments.

### Context of Rehabilitation Literature

Initial attempts to link the FIM to the MDS or to MDS-like instruments have met with reasonable success. Williams et al. compared scores from FIM and rescaled

MDS motor activity and cognitive skill items (referred to as “Pseudo-FIM(E)”) of 173 rehabilitation patients admitted to six nursing homes [18]. The matching and rescaling of the MDS data were accomplished through an expert panel, with the panel judging that 12 (8 ADL activities, 4 cognitive skills) out of 18 FIM items had a corresponding MDS item. Mean ADL activity and cognitive skills subscales did not statistically differ and showed strong correlations (0.72 and 0.78, respectively). The match across items was only fair. The mean of the motor activity and cognitive skill items differed significantly, with 7 of 12 items being significantly different across the two instruments.

**Table 1.**

Comparison of Functional Independence Measure (FIM) with Minimum Data Set (MDS) items and rating scales for activities of daily living (ADL) and motor skills.

Parameter	FIM	MDS
ADL/Motor Skill	Eating	Eating
	Grooming	Personal Hygiene
	Bathing	Bathing*
	Dressing-Upper Body	Dressing
	Dressing-Lower Body	—
	Toileting	Toilet Use
	Bladder Management	Bladder Continence <sup>†</sup>
	Bowel Management	Bowel Continence <sup>†</sup>
	Bed, Chair, Wheelchair (Transfer)	Transfer
	Toilet (Transfer)	—
	Tub, Shower (Transfer)	—
	Stairs	—
	—	Bed Mobility
	Walk/Wheelchair	Walk in Room
	—	Walk in Corridor
	—	Locomotion on Unit
—	Locomotion off Unit	
Rating Scale	7 = Complete Independence (Timely, Safely)	0 = Independent
	6 = Modified Independence (Device)	—
	5 = Supervision	1 = Supervision
	4 = Minimal Assistance (Subject 75%+)	2 = Limited Assistance
	3 = Moderate Assistance (Subject 50%+)	—
	2 = Maximal Assistance (Subject 25%+)	3 = Extensive Assistance
	1 = Total Assistance (Subject 0%+)	4 = Total Dependence
	—	8 = Activity Did Not Occur During Entire 7-Day Period

\*Separate rating scale in MDS: 0 = independent, 1 = supervision, 2 = physical help limited to transfer only, 3 = physical help in part of bathing activity, 4 = total dependence, 8 = activity did not occur during entire 7 days.

<sup>†</sup>Separate rating scale in MDS: 0 = usually continent, 2 = occasionally continent, 3 = frequently incontinent, 4 = incontinent.

More recently, Buchanan et al. attempted to create a crosswalk between the FIM and the MDS-Post Acute Care (MDS-PAC), an instrument that is similar to the MDS but modified to address the needs of postacute care [19]. Thirteen ADL and five cognitive skill items were matched across the instruments. In contrast to Williams and colleagues' study, which had trained data collectors, Buchanan and colleagues had rehabilitation staff administer both instruments. Like Williams et al.'s study, mean ADL activity and cognitive skill subscales did not statistically differ and correlations were strong at 0.85 and 0.84, respectively. Again, item match was only fair across the instruments, with 11 of 18 items having poor  $\kappa$  statistics ( $<0.4$ ).

While initial attempts at linking the FIM and the MDS-like instruments are encouraging, these studies are limited by their methodologies. Expert-panel matching requires an item-to-item match across instruments. Differences in the instruments at the item-level imply that comparing scores across instruments is difficult, if not impossible [20]. Furthermore, conversion systems based on item-to-item matching are based on similar items across instruments and disregard nonmatching items from one or the other instruments that may contribute to the construct. This elimination of items is a threat to the instrument's psychometric integrity, especially reliability and measurement precision [21]. In addition to the rigorous criterion of item-to-item match, the raw-score or classical test theory (CTT) methodologies used in previous studies are limited by sample and test dependency. That is, test psychometrics, while commonly presented as a characteristic of the test, are dependent on the sample from which those psychometrics were derived [22]. Finally, when CTT methodologies are used, an individual's score for a particular construct (e.g., ADL) is dependent on the particular test or instrument; a test with easy items will generate higher scores than a test with difficult items.

While raw-score or CTT methodologies are sample and test dependent, item response theory (IRT) is sample and test independent. IRT relates the observable responses to items on a test (correct or incorrect) to the unobservable latent trait that the test is intended to measure [23]. The item response function postulates that respondents with more ability have a higher probability of providing correct responses (or higher ratings) to items than patients with less ability. This function allows the estimation of the latent trait, independent of the particular sample of items selected for the test [23]. That is, an examinee theoretically should have the same ability

across various samples of test items and across various tests intended to measure the same trait [23]. Using IRT, we can view each healthcare instrument as representing different samples of test items that all measure a particular latent trait (e.g., ADL) and therefore generate the same or similar ability measures. By connecting the CTT-generated scores of instruments to their IRT-generated measures, we should be able to readily link or create crosswalks between instruments that are intended to measure the same latent trait, such as ADL ability.

While IRT has been gaining tremendous popularity in evaluating healthcare instruments, it has only recently been applied to linking measures in healthcare. Fisher et al. were the first to use Rasch analysis, the least complex of the IRT models,<sup>\*</sup> to link two global measures of function, the FIM and the Patient Evaluation Conference System (PECS) [24]. They found that separate FIM- and PECS-generated measures for 54 rehabilitation patients correlated at 0.91. Furthermore, they demonstrated that Rasch analysis could be used to create a common unit of measurement, which they referred to as the "rehabit." Smith and Taylor replicated this finding with a sample of approximately 500 patients [25]. This common unit of measurement allows the translation of raw scores generated from one instrument to those generated from another instrument. Since the results of Rasch analysis are sample and test independent, these tables or algorithms can be used for all future and past instrument-to-instrument score conversions.

Fisher et al. used this same methodology to create a common unit of measurement across the 10 physical function items of the Medical Outcome Scale 36-Item Short Form and the Louisiana State University Health Status Instrument [26]. Difficulty estimates for a subset of similar items from the two instruments correlated at 0.95, which again suggests that the items from the two scales measured the same latent trait. In a sample of 1,926 patients, McHorney and Cohen applied a two-parameter IRT model to link three forms of a survey with 206 physical functioning items through 71 items common across the three forms [27]. In a similar study, with a sample of

---

<sup>\*</sup>Considerable debate exists over whether the Rasch measurement model is a subcategory of IRT or a separate measurement theory in and of itself (*Source*: Andrich D. Controversy and the Rasch model: A characteristic of incompatible paradigms? *Med Care*. 2004;42(1 Suppl):17-16. [PMID: 14707751](https://pubmed.ncbi.nlm.nih.gov/14707751/)). Statistically, Rasch or the one-parameter IRT model differs from two- and three-parameter IRT models in that it holds the item discrimination and guessing parameters constant.

4,566 respondents from the Asset and Health Dynamics Among the Oldest Old study, McHorney used a two-parameter IRT model to link three sets of physical functioning items (total of 39 items) through 16 common items [28]. Both studies demonstrated that items intended to measure the same latent trait can be linked and placed on a common metric.

While a growing number of studies use IRT methodologies to link healthcare instruments, we have found no studies that have applied these methodologies to link the FIM and MDS. In addition, a limitation of previous FIM-MDS linking studies was that research-trained data collectors or healthcare staff administered both instruments. However, the positive findings of these prospective studies may be inflated since the data collectors were not blinded to the intent of the study, i.e., to link the FIM and MDS. This study demonstrates Rasch analysis for linking the motor components of the FIM and the MDS through the use of existing Department of Veterans Affairs (VA) databases. In addition to taking advantage of the test- and sample-free characteristics of IRT methodologies, we have “naturally blinded” the study from data-collector bias by using existing VA databases.

## METHODS

### Analysis

Common person-item equating was employed in linking and creating the crosswalk between the FIM and MDS. This methodology requires (1) sufficient commonality of items across the two instruments and (2) a “linked sample,” a sample of participants who have been administered both instruments. **Table 1** presents the items of the FIM along with analogous ADL activity items from the MDS and their respective rating scales. At the item-label level, the similarities across the two instruments are evident. Basic self-care items such as eating, grooming, bathing, dressing, bladder/bowel management, transfers, and ambulation are common to the FIM and the MDS. These similarities provide the initial “face validity” that the two instruments are measuring the same construct. As stated earlier, IRT linking methodologies do not require a one-to-one correspondence between items, definitions, or rating scales. All that is required is a commonality across the items of both instruments that suggests the instruments are representative of the same underlying construct.

The second element necessary for creating a crosswalk across the FIM and MDS is the existence of a linked sample—a sample of patients that had both FIM and MDS data. Initial inquiries to VA medical centers (VAMCs) nationwide identified four facilities with FIM-MDS linked data that were interested in participating in the study. The FIM and MDS data reside in two databases at the VA’s Austin Automation Center. Data from both databases were downloaded from the four VA facilities and merged on the basis of Social Security numbers. To minimize the effect that change in a patient’s condition could have on FIM and MDS scores, we restricted the amount of time that elapsed between administrations of the assessments to 7 days. For purposes of this initial linking study, we collapsed the sample across all demographic and diagnostic groups. The study was approved by the institutional review board at the University of Florida and by the Research and Development Subcommittee for Clinical Investigations at the VAMC in Gainesville, Florida.

While considerable debate exists over the necessary sample size for IRT analyses, Wang and Chen, using Monte Carlo techniques, showed that sample size had virtually no effect on item bias for Rasch polytomous (rating-scale) analyses performed with Winsteps<sup>®</sup> (Winsteps, Chicago, Illinois) [29]. In contrast, item number did influence item bias, with tests of 20 or more items showing bias that was statistically significant, but negligible in magnitude. The combined analysis in the present study consisted of 26 items. In a pilot analysis we conducted of the FIM motor scale with an independent random sample of 200 patients, item error range was small, between 0.06 and 0.09 logits. A summary of the steps for creating the crosswalk follows.

### Step 1: Convert MDS Ratings to Match FIM Ratings

Prior to performing the Rasch analysis, we took several steps to increase the conceptual consistency of the FIM and MDS rating scales. One inconsistency between the instruments is that the MDS includes a separate designation for “activity did not occur,” while the FIM incorporates this situation into a rating of requiring “total assistance” (**Table 1**). Using a procedure adapted by Jette and colleagues, we recoded the MDS rating of “activity did not occur” to the FIM rating of “total dependence” [30]. The rationale underlying this decision was that a probable explanation for an activity not occurring was that the activity could not be performed [30].

Another inconsistency between the FIM and the MDS is that their rating scales progress in different directions; i.e., “independence” is reflected by a higher score on the FIM and a lower score on the MDS. Furthermore, the FIM and MDS rating scales have different ranges, 1 to 7 and 0 to 4, respectively, with the FIM ratings of “3 moderate assistance” and “6 modified independence” without an analogous rating category on the MDS (Table 2). To compensate for the rating-scale differences, we reverse-scored the MDS and recoded its rating scale to qualitatively match that of the FIM (Table 2). For example, “4 total dependence” on the MDS was recoded to match “1 total assistance” on the FIM.

### Step 2: Remove Invalid Data

In generating a crosswalk between instruments, we must assure that the conversion system is based on valid data. While a number of criteria exist for determining data validity, one expectation is that patients should have similar scores on similar measures. For example, a patient with an overall score that represents independence on the FIM should obtain a similar score on the MDS. We employed the method used by Linacre for eliminating invalid data by creating a scatter plot of FIM and MDS measures for each patient and removing person measures that were outside the 95 percent confidence interval around the identity line.

### Step 3: Generate FIM-MDS Cocalibrated Item and Rating-Scale Measures

Following elimination of invalid person measures, our next critical step in creating the crosswalk was to place the FIM and MDS items and rating-scale calibrations on a common linear scale with the same local origin. We accomplished this by including both FIM and

MDS items in a Rasch (one-parameter IRT) partial credit model analysis using the Winsteps<sup>®</sup> program [13]. The combined analysis provided cocalibrated item measures and rating-scale measures (“step measures”), which we then used as anchors in separate FIM and MDS analyses.

### Step 4: Anchor Separate FIM and MDS Rasch Analyses to Item and Step Measures from Cocalibrated Analysis

We generated the final crosswalk of FIM raw scores to MDS raw scores from two Rasch analyses of the FIM and MDS data, both of which were anchored on the cocalibrated item and step measures. Each of these analyses produced a table, which connected total FIM and total MDS raw scores to person measures.

### Step 5: Generate Crosswalk Tables

Since person-ability measures generated from the separate FIM and MDS Rasch analyses were anchored to the item and step measures from the cocalibrated analysis (see steps 3 and 4), the raw scores of the FIM become linked to the raw scores of the MDS. By matching the raw scores from each instrument to the same “linked” person measure, we created a FIM-MDS raw-score conversion table whereby total FIM raw scores can be translated into total MDS raw scores and total MDS raw scores can be translated into total FIM raw scores.

## RESULTS

Data were downloaded from four VA facilities and merged on the basis of Social Security number. The criteria of “7 or less days” between assessment administrations resulted in 254 unique patient records with linked FIM and MDS data. Eighteen patients with FIM-MDS measures that fell outside the 95 percent confidence interval around the scatter plot identity line were eliminated from all further analyses, which resulted in a final sample of 236 patient records. The days between the administration of the FIM and the MDS ranged from 0 to 7, with a mean  $\pm$  standard deviation (SD) of  $3.7 \pm 1.9$  days. The major diagnostic groups included stroke (25%), orthopedic (22.5%), medically complex (11.4%), and amputation (8.5%) (Table 3). These primary diagnostic classifications were retrieved from the FIM data set. Patients ranged in age from 30 to 96 years, with a mean  $\pm$  SD age of  $71.24 \pm 11$  years.

**Table 2.** Minimum Data Set (MDS) to Functional Independence Measure (FIM) rating conversion based on qualitative meanings of categories.

MDS Rating Description	MDS Rating $\rightarrow$ FIM Rating	FIM Rating Description
Total Dependence	4 $\rightarrow$ 1	Total Assistance
Extensive Assistance	3 $\rightarrow$ 2	Maximum Assistance
—	3	Moderate Assistance
Limited Assistance	2 $\rightarrow$ 4	Minimal Assistance
Supervision	1 $\rightarrow$ 5	Supervision
—	6	Modified Independence
Independent	0 $\rightarrow$ 7	Complete Independence

**Table 3.**

Primary diagnostic classification retrieved from Functional Independence Measure data set.

Diagnosis	Frequency	Percent
Stroke	59	25.0
Orthopedic Disorders	53	22.5
Medically Complex	27	11.4
Amputation	20	8.5
Neurologic Conditions	14	5.9
Debility	13	5.5
Cardiac	11	4.7
Pain Syndromes	10	4.2
Brain Dysfunction	8	3.4
Pulmonary Disorders	6	2.5
Other Disabling Impairments	6	2.5
Arthritis	4	1.7
Major Multiple Trauma	3	1.3
Burns	2	0.8
Total	236	100.0

Ninety-six percent of the sample was male, and 45 percent of the sample was married.

In general, the combined FIM-MDS measure showed good person- and item-level psychometric properties. Person reliability (Cronbach  $\alpha$ ) was 0.94. The average item infit of the combined instrument was 1.1 (ideal is 1.0), with 21 of the 26 items showing infit and outfit mean-square statistics below 1.7, which is the recommendation for clinical scales (misfitting items were FIM stairs, MDS walk in room, MDS walk in corridor, MDS locomotion off unit, and MDS bladder) [31]. Point-measure correlations for the items ranged between 0.54 and 0.84. The average item difficulty (mean  $\pm$  SD = 0.00  $\pm$  0.56 logits) was well-matched with the mean of person measures (mean  $\pm$  SD = 0.01  $\pm$  0.9 logits).

The **Figure** is a map that presents person measures (“#” represents two patients and “.” represents one patient) to the far left, FIM item measures in the middle, and MDS item measures to the far right. Linear measures, in natural logarithm odds units (logits), are represented along the central axis. The distribution of person calibrations (higher represents more ability and lower represents less ability) were normally distributed and matched well with item difficulties of the FIM and MDS (i.e., no ceiling or floor effects). In general, item calibrations from the FIM and MDS were distributed as has been reported in previous literature for global functional measures [13,27].

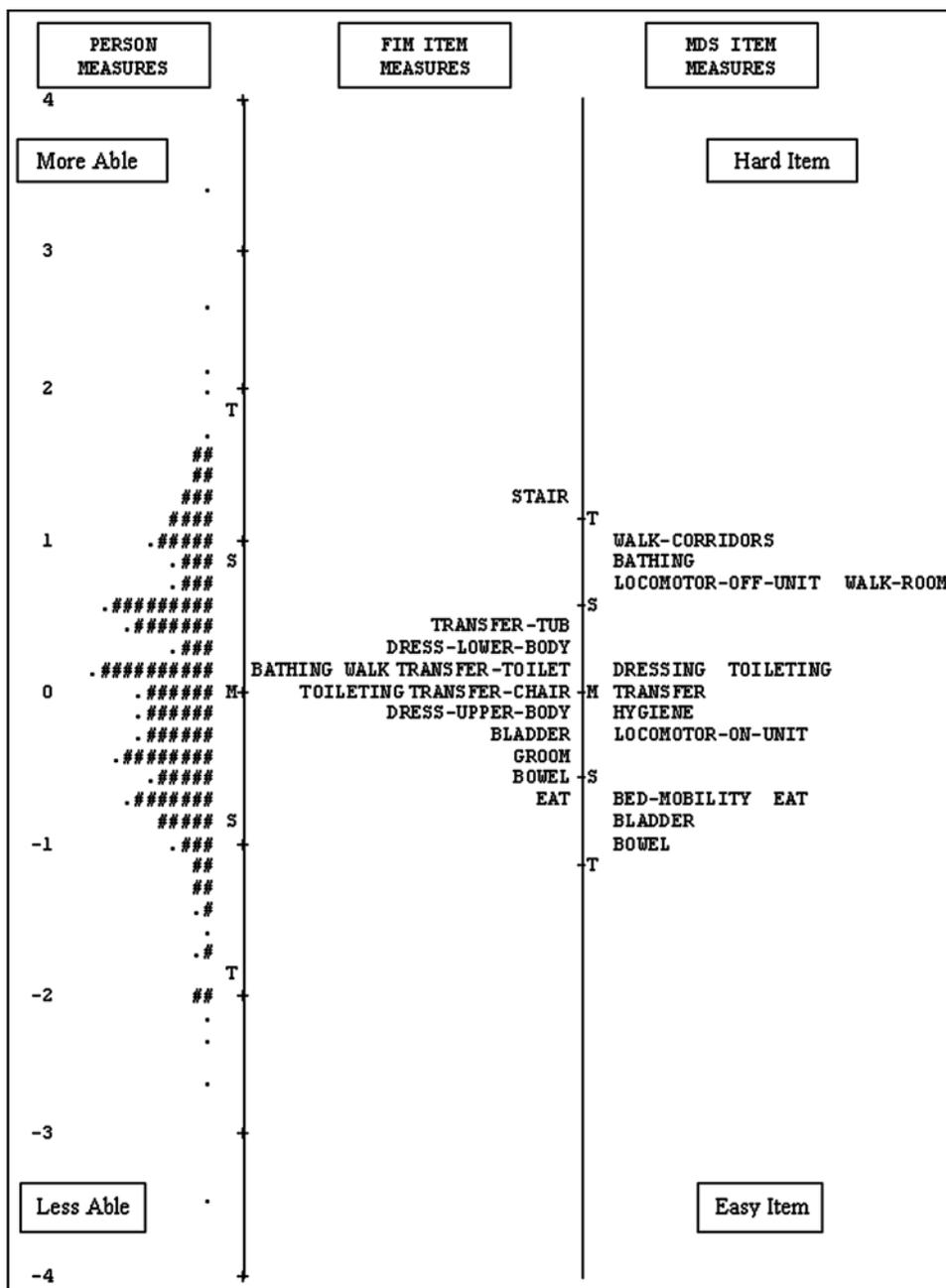
For example, the less challenging items of the FIM and MDS, such as eating and bowel/bladder function, had the lowest calibrations, while more challenging items, such as FIM stairs and MDS walk in corridor, had the highest calibrations. All the items of the FIM and MDS calibrated within 2.26 logits (range [mean  $\pm$  standard error]  $-0.95 \pm 0.05$  to  $1.31 \pm 0.06$  logits). A number of items that had similar definitions across the two instruments demonstrated similar calibrations. For example, the corresponding FIM-MDS items for eating, toileting, and transfer (transfer to chair and transfer) were all within 0.05 logits. The corresponding FIM-MDS items that were furthest apart were bathing (0.68 logits) and bladder function (0.57 logits).

**Table 4** is the crosswalk or conversion table of FIM raw scores to MDS raw scores. The raw scores are linked through the person measures generated from separate FIM and MDS analyses, both of which were anchored at the cocalibrated item and step measures. The FIM raw scores and the MDS raw scores are linked through the person measures generated from each analysis. For example, a FIM raw score of 50 corresponds to a MDS raw score of 28.5 because both of these raw scores were associated with the common person ability measure of  $-0.14$  logits. Similarly, a MDS raw score of 10 corresponds to a FIM raw score of 78 since both raw scores were associated with a person ability measure of 1.12 logits.

The relationships between FIM raw scores and MDS raw scores as well as FIM measures and MDS measures were analyzed. The raw scores correlated at  $-0.81$  and the measures correlated at 0.78. Note that the negative correlation of the raw scores is a function of the greater “independence” on motor activities being represented by higher ratings on FIM and lower ratings on the MDS.

## DISCUSSION

This study demonstrated Rasch analysis for linking the motor components of the FIM and the MDS through the use of existing VA databases. We described six steps for linking the FIM and MDS: (1) choosing items from both the FIM and MDS, (2) equating FIM and MDS rating scales in terms of directionality and range, (3) removing invalid data, (4) running a cocalibrated Rasch analysis with both FIM and MDS scores that places all items and ratings on the same linear scale, (5) running separate Rasch analyses anchored on the cocalibrated item and rating-scale



**Figure.** Map representing person measures (“#” represents two patients and “.” represents one patient) to far left, Functional Independence Measure (FIM) item measures in middle, and Minimum Data Set (MDS) item measures to far right. Linear measures, in natural logarithm odds units (logits), are represented along central axis. Distribution of person calibrations (higher represents more ability and lower represents less ability) were normally distributed and matched well with item difficulties of FIM and MDS (i.e., no ceiling or floor effects). M = mean, S = 1 and T = 2 standard deviations.

values that link the person measures derived from the two instruments, and (6) creating a raw-score conversion table by matching FIM and MDS raw scores to similar measures derived from the anchored analyses.

Overall, the psychometrics of the cocalibrated analysis indicated that the motor activity items of the FIM and MDS appear to be measuring the same construct. The internal consistency of the combined instrument was good,

**Table 4.**

Functional Independence Measure (FIM)–Minimum Data Set (MDS) crosswalk conversion table based on raw scores.

FIM	Logit	MDS	FIM	Logit	MDS	FIM	Logit	MDS
13	-3.79	51.5	40	-0.48	33.0	67	0.51	19.0
14	-2.76	50.0	41	-0.44	32.5	68	0.56	18.0
15	-2.25	49.0	42	-0.41	32.0	69	0.60	17.5
16	-1.98	48.0	43	-0.38	31.5	70	0.65	17.0
17	-1.79	47.0	44	-0.34	31.0	71	0.70	16.0
18	-1.65	46.0	45	-0.31	30.5	72	0.75	15.0
19	-1.54	45.5	46	-0.27	30.0	73	0.81	14.5
20	-1.45	45.0	47	-0.24	29.5	74	0.86	13.5
21	-1.36	44.0	48	-0.21	29.5	75	0.92	13.0
22	-1.29	43.0	49	-0.17	29.0	76	0.99	12.0
23	-1.22	43.0	50	-0.14	28.5	77	1.05	11.0
24	-1.16	42.0	51	-0.11	28.0	78	1.12	10.0
25	-1.10	41.0	52	-0.07	27.5	79	1.19	9.5
26	-1.05	41.0	53	-0.04	27.0	80	1.27	8.5
27	-1.00	40.0	54	0.00	26.5	81	1.36	7.5
28	-0.95	39.5	55	0.03	26.0	82	1.45	6.5
29	-0.91	39.0	56	0.07	25.5	83	1.55	6.0
30	-0.86	38.5	57	0.11	25.0	84	1.67	5.0
31	-0.82	38.0	58	0.14	24.0	85	1.80	4.0
32	-0.78	37.5	59	0.18	23.5	86	1.96	3.5
33	-0.74	37.0	60	0.22	23.0	87	2.16	2.5
34	-0.70	36.5	61	0.26	22.5	88	2.40	2.0
35	-0.66	36.0	62	0.30	22.0	89	2.76	1.0
36	-0.62	35.0	63	0.34	21.5	90	3.38	0.5
37	-0.59	34.5	64	0.38	20.5	91	4.52	0.0
38	-0.55	34.0	65	0.42	20.0			
39	-0.51	33.5	66	0.46	19.5			

as indicated by good person separation (Cronbach  $\alpha = 0.94$ ), good point-measure correlations (0.54–0.84), and good fit statistics for 21 of the 26 items. Only one item from the FIM (stairs) and four items from the MDS (3 locomotor items and the bladder item) had poor fit statistics. The tendency of ambulation items and bowel/bladder items to have high fit statistics is well documented in the literature [32–33]. These findings indicate that ambulation/locomotion items and incontinence items may represent a construct separate from other motor items.

In the combined analysis, the items from the FIM and MDS showed a similar, logical order of difficulty, with items requiring less physical ability (e.g., eating, bowel, bladder, grooming/hygiene) showing less difficulty than items requiring more physical ability (e.g., stairs, walk-

ing off unit, walking room). This finding again replicates the item hierarchy shown in individual analyses of the FIM, MDS, and other similar global measures of function [13,32,34]. These findings can be interpreted to mean that the motor activities of the FIM and MDS represent a subset of items (from many items) that can be used to measure global physical functioning ability.

Further evidence of the compatibility of the motor activity FIM and MDS scores and measures was evidenced by their correlations. The FIM-MDS raw scores and FIM-MDS anchored measures correlated at  $-0.81$  and  $0.78$ , respectively. The correlations from the present study were slightly higher than the  $0.72$  correlation found by Williams et al. in their comparison of the FIM with their rescaled motor activity MDS (Pseudo-FIM(E)) [18],

and slightly lower than the correlation of 0.85 found by Buchanan et al. between the FIM and the MDS-PAC motor scales [19]. In some respects, since the present study was a secondary analysis of clinical/administrative data, we were a bit surprised that our correlations were as high as those achieved in prospective studies. Since the purpose of the study, to link instruments, may be difficult to keep “blind” from data collectors in prospective studies, one may expect higher correlations between the FIM and MDS (or MDS-PAC). We should note that prospective studies of other crosswalked instruments have shown correlations in the 0.90s [24].

Since the present study is an initial demonstration of Rasch procedures for linking the FIM to MDS, a number of modifications may result in a more precise crosswalk. Selecting patients with FIM-MDS assessments less than 7 days apart may reduce variability in scores from patient change across assessment administrations. We may find an “optimal” time period between FIM and MDS assessments that produces the most accurate conversion table. A more unidimensional form of the instrument (i.e., removing ambulation and incontinence items) may result in a more precise crosswalk. Furthermore, different item calibrations that may be generated across different diagnostic groups (differential item functioning [DIF]), may lead to degradation of the crosswalk. Improvement in the translation of scores across the FIM and MDS may occur through removal of items with DIF or creation of crosswalks specific to diagnostic groups. We should note that while removal of items may improve the crosswalk, these changes may negate the underlying psychometrics of the original measures. Finally, the critical test of the crosswalk is whether the conversion system generates scores that are comparable with actual scores in a validation sample. We are presently addressing each of the above challenges with larger VA data sets.

This study demonstrated a methodology for connecting scores from similar healthcare instruments and represents a conceptual change in instrument linking. That is, items from similar instruments represent subsets of items in the pool of all items that represent a particular construct, e.g., motor construct. Furthermore, when using IRT methodologies, we do not need a one-to-one match between items and rating scales to convert scores from one instrument to another.

## CONCLUSIONS

The implications of translating scores across healthcare instruments may be profound. Crosswalks would allow clinicians or researchers to follow patients longitudinally across the continuum of care, independent of the setting-specific measures used to monitor outcomes. Facility-level outcomes could be compared, even when the facilities used different outcomes measures. Furthermore, crosswalks may permit findings to be more readily compared across studies that use different outcomes measures intended to measure the same construct. Of course, the viability of this methodology in healthcare is still in its infancy and its applicability is dependent on replications and validity studies.

## ACKNOWLEDGMENTS

We thank Dr. Christa Hojlo and Clifford Marshall, MS, for clinical consultations. We also thank Dr. Richard Smith for statistical consultation.

This material was based on work supported by the University of Florida Research Opportunity Fund (grant 33070612) and also resources and the use of facilities at the Rehabilitation Outcomes Research Center, VA Health Services Research and Development Service and Rehabilitation Research and Development Service Center of Excellence, North Florida/South Georgia Veterans Health System.

The authors have declared that no competing interests exist.

## REFERENCES

1. Reker DM, Reid K, Duncan PW, Marshall C, Cowper D, Stansbury J, Warr-Wing KL. Development of an integrated stroke outcomes database within the Veterans Health Administration. *J Rehabil Res Dev*. 2005;42(1):77–91. [\[PMID: 15742252\]](#)
2. Grabois M. Open letter from ACRM to HCFA on proposed Medicare PPS. American Congress of Rehabilitation Medicine. Health Care Financing Administration. *Arch Phys Med Rehabil*. 2001;82(4):567–69. [\[PMID: 11310452\]](#)
3. DeJong G. Post-Acute Care: Hearing before the Subcomm. on Health of the House Comm. on Ways and Means, 109th Cong., 1st Sess. (Jun. 16, 2005).

4. Granger CV, Hamilton BB. UDS Report. The Uniform Data System for Medical Rehabilitation Report of First Admissions for 1990. *Am J Phys Med Rehabil.* 1992; 71(2):108–13. [\[PMID: 1532719\]](#)
5. Fiedler RC, Granger CV. Uniform data system for medical rehabilitation: Report of first admissions for 1995. *Am J Phys Med Rehabil.* 1997;76(1):76–81. [\[PMID: 9036916\]](#)
6. Department of Veterans Affairs (VA), Veterans Health Administration (VHA). VHA Directive 2000–016: Medical rehabilitation outcomes for stroke, traumatic brain injury, and lower extremity amputation patients. Washington (DC): VA, VHA; 2000.
7. Carter GM, Buntin MB, Hayden O, Paddock SM, Relles DA, Ridgeway G, Totten ME, Wynn BO. Analyses for the initial implementation for Medicare's inpatient rehabilitation prospective payment system. RAND Monograph Report No.: MR-1500-CMS. Santa Monica (CA): RAND; 2002. p. 1–360.
8. Rantz MJ, Zwiygart-Stauffacher M, Popejoy LL, Mehr DR, Grando VT, Wipke-Tevis DD, Hicks LL, Conn VS, Porter R, Maas M. The minimum data set: No longer just for clinical assessment. *Nurs Home Med.* 1999;7(9):354–60.
9. Morris JN, Hawes C, Fries BE, Phillips CD, Mor V, Katz S, Murphy K, Drugovich ML, Friedlob AS. Designing the national resident assessment instrument for nursing homes. *Gerontologist.* 1990;30(3):293–307. [\[PMID: 2354790\]](#)
10. Casten R, Lawton MP, Parmelee PA, Kleban MH. Psychometric characteristics of the Minimum Data Set I: Confirmatory factor analysis. *J Am Geriatr Soc.* 1998;46(6):726–35. [\[PMID: 9625189\]](#)
11. Department of Veterans Affairs (VA), Veterans Health Administration (VHA). VHA Directive 2005-060: Implementation of the Medicare prospective payment system (PPS) assessment form (MPAF). Washington (DC): VA, VHA; 2005.
12. Stineman MG, Shea JA, Jette A, Tassoni CJ, Ottenbacher KJ, Fiedler RC, Granger CV. The Functional Independence Measure: Tests of scaling assumptions, structure, and reliability across 20 diverse impairment categories. *Arch Phys Med Rehabil.* 1996;77(11):1101–8. [\[PMID: 8931518\]](#)
13. Linacre JM, Heinemann AW, Wright BD, Granger CV, Hamilton BB. The structure and stability of the Functional Independence Measure. *Arch Phys Med Rehabil.* 1994; 75(2):127–32. [\[PMID: 8311667\]](#)
14. Ottenbacher KJ, Hsu Y, Granger CV, Fiedler RC. The reliability of the Functional Independence Measure: A quantitative review. *Arch Phys Med Rehabil.* 1996;77(12):1226–32. [\[PMID: 8976303\]](#)
15. Hawes C, Morris JN, Phillips CD, Mor V, Fries BE, None-maker S. Reliability estimates for the Minimum Data Set for nursing home resident assessment and care screening (MDS). *Gerontologist.* 1995;35(2):172–78. [\[PMID: 7750773\]](#)
16. Stineman MG, Maislin G. Clinical, epidemiological, and policy implications of Minimum Data Set validity. *J Am Geriatr Soc.* 2000;48(12):1734–36. [\[PMID: 11129771\]](#)
17. Lawton MP, Casten R, Parmelee PA, Van Haitsma K, Corn J, Kleban MH. Psychometric characteristics of the Minimum Data Set II: Validity. *J Am Geriatr Soc.* 1998;46(6): 736–44. [\[PMID: 9625190\]](#)
18. Williams BC, Li Y, Fries BE, Warren RL. Predicting patient scores between the Functional Independence Measure and the Minimum Data Set: Development and performance of a FIM-MDS “crosswalk.” *Arch Phys Med Rehabil.* 1997;78:48–54. [\[PMID: 9014957\]](#)
19. Buchanan JL, Andres PL, Haley SM, Paddock SM, Zaslavsky AM. Evaluating the planned substitution of the Minimum Data Set-Post Acute Care for use in the rehabilitation hospital prospective payment system. *Med Care.* 2004;42(2):155–63. [\[PMID: 14734953\]](#)
20. Rogers JC, Green Gwinn SM, Holm MB. Comparing activities of daily living assessment instruments: FIMt, MDS, OASIS, MDS-PAC. *Phys Occup Ther Geriatr.* 2001;18(3): 1–25.
21. McHorney CA. Generic health measurement: Past accomplishments and a measurement paradigm for the 21st century. *Ann Intern Med.* 1997;127(8 Pt 2):743–50. [\[PMID: 9382391\]](#)
22. Thompson B, Vacha-Haase T. Psychometrics is datametrics: Test is not reliable. *Educ Psychol Meas.* 2000;60(2): 174–95.
23. Hambleton RK. Principles and selected applications of item response theory. In: Linn RL, editor. *Educational measurement.* 3rd ed. New York (NY): Macmillan; 1989. p. 147–200.
24. Fisher WP Jr, Harvey RF, Taylor P, Kilgore KM, Kelly CK. Rehabits: A common language of functional assessment. *Arch Phys Med Rehabil.* 1995;76(2):113–22. [\[PMID: 7848069\]](#)
25. Smith RM, Taylor PA. Equating rehabilitation outcome scales: Developing common metrics. *J Appl Meas.* 2004; 5(3):229–42. [\[PMID: 15243171\]](#)
26. Fisher WP Jr, Eubanks RL, Marier RL. Equating the MOS SF36 and the LSU HSI Physical Functioning Scales. *J Outcome Meas.* 1997;1(4):329–62. [\[PMID: 9661727\]](#)
27. McHorney CA, Cohen AS. Equating health status measures with item response theory: Illustrations with functional status items. *Med Care.* 2000;38(9 Suppl):II43–59. [\[PMID: 10982089\]](#)
28. McHorney CA. Use of item response theory to link 3 modules of functional status items from the Asset and Health Dynamics Among the Oldest Old Study. *Arch Phys Med Rehabil.* 2002;83(3):383–94. [\[PMID: 11887121\]](#)
29. Wang WC, Chen CT. Item parameter recovery, standard error estimates, and fit statistics of the Winsteps program

- for the family of Rasch models. *Educ Psychol Meas.* 2005; 65(3):376–404.
30. Jette AM, Haley SM, Ni P. Comparison of functional status tools used in post-acute care. *Health Care Financ Rev.* 2003;24(3):13–24. [[PMID: 12894632](#)]
31. Wright BD, Linacre JM. Reasonable mean-square fit values. *Rasch Meas Trans.* 1994;8(3).
32. Velozo CA, Magalhaes LC, Pan AW, Leiter P. Functional scale discrimination at admission and discharge: Rasch analysis of the Level of Rehabilitation Scale-III. *Arch Phys Med Rehabil.* 1995;76(8):705–12. [[PMID: 7632124](#)]
33. Nilsson AL, Sunnerhagen KS, Grimby G. Scoring alternatives for FIM in neurological disorders applying Rasch analysis. *Acta Neurol Scand.* 2005;111(4):264–73. [[PMID: 15740579](#)]
34. Silverstein B, Fisher WP, Kilgore KM, Harley JP, Harvey RF. Applying psychometric criteria to functional assessment in medical rehabilitation: II. Defining interval measures. *Arch Phys Med Rehabil.* 1992;73(6):507–18. [[PMID: 1622298](#)]

Submitted for publication June 15, 2006. Accepted in revised form December 19, 2006.