Instantaneous Measurement and Diagnosis

John M. Linacre, Ph.D.
MESA Psychometric Laboratory
University of Chicago

The manufacture of measuring instruments is typically a large-scale, standards-based process. Their use is frequently an on-demand, local operation requiring immediate measures and measure interpretation. The FIM has been calibrated on large samples. These calibrations are used to construct the KeyFIM, a one-page data collection, measurement, and analysis device. This provides the physician the same measurement ease and immediacy as the yardstick does the carpenter. The KeyFIM incorporates the measurement replication and quality control diagnosis that the careful carpenter obtains by multiple measurements of the same unknown length.

Better Measurement

Better measuring instruments are not only more accurate and precise, they are also more immediate and intuitive. In industrial instrumentation, "better measurements, and more of them, have made it possible to interpret most data without recourse to statistical techniques" (Youden W.J., 1954).

Statistical techniques, particularly as implemented in computer programs, enable the calibration of observation instruments, such as the FIM, on large samples of patient records, representing many impairment groups and rehabilitation institutions. Collecting and analyzing large patient-record databases is an expensive and time-consuming process. Although this process yields useful information about the FIM and the patients to which it has been applied (Granger et al. 1993), it is far too slow and cumbersome to assist in the treatment of the patients whose records are in the database.

Effective use of the FIM requires that data collection, analysis, and interpretation occur almost instantaneously, preferably while the clinician is still with the patient (as with the clinical thermometer and stethoscope) or at least in a day or so (as with hospital-based laboratory tests). The increasing speed and ubiquity of computers will ultimately permit the development of artificially-intelligent systems to support the real-time analysis and interpretation of a patient's ratings on the 18 FIM items. Such interpretation will be based on the accumulated case histories of millions of patients to whom the FIM will have been administered. Nevertheless, the immediate local clinical experience of practitioners and their personal knowledge of the particular patient will always play a part in FIM interpretation.

Most of the benefits of a sophisticated computer-based system can be realized immediately with the KeyFIM, a simple, paper-and-pencil implementation of the FIM. This form combines into one graphical presentation the essential steps of data collection and measurement construction, along with a convenient layout for intuitive quality control and diagnostic interpretation.

Calibrating the Measurement System

The FIM consists of 18 items, each rated on a seven-category rating scale with each succeeding category carefully defined to represent an increasing degree of functional independence. It is designed to be administered to patients on admission to and discharge from a rehabilitation institution. Data collected from thousands of applications of the FIM have been subjected to extensive analysis. Linacre et al. (1994) report that analysis of FIM data from a measurement perspective by means of the Rasch model discloses that decomposing the 18-item FIM into 13 motor items and the 5 cognitive items produces two bases for measurement, clearly superior to the one composite original. For convenience, this paper focuses only on the FIM cognitive items, but the same considerations apply directly to the motor items.

Analysis of the FIM was conducted in the computationally convenient unit of measurement known as the Logit (log-odds unit, see Linacre, 1993, for other derivations). Though the Logit has a clear probabilistic interpretation (Wright & Stone, 1979 p. 36), its substantive interpretation depends on the use to which the measures are put. FIM measures are used in a rehabilitation setting in which clinicians expect patients to be functioning within a bounded range of the conceptually infinitely wide variable (dimension, construct) of independence. The variable is infinitely wide, because it is always possible to imagine a patient even more dependent than any encountered to date (e.g., in a deeper coma), or even more independent than any encountered (e.g., with greater physical and mental prowess). The bounded range of independence is that for which the rehabilitation setting is designed. Accordingly, it is convenient to define a measurement scale with its "0" point corresponding conceptually to the lowest level of functioning at which a patient might be administered to rehabilitation. Similarly, the "100" point is defined to be the highest level of functioning which a patient might achieve and still remain in rehabilitation. In order to maintain the interval-scale measurement characteristics of the logit (Stevens, 1951), this "0" to "100" scale is a linear transformation of the logit scale. For clarity in substantive use, the new units of measurement are called FIMITs (Linacre, 1995).

FIM Cognitive Items		
	Item Name	FIMIT calibration
N.	Auditory Comprehension	42
O.	Verbal Expression	40
P.	Social Interaction	46
Q.	Problem Solving	55
R.	Memory	52

Table 1. FIM Cognitive Items, condensed from FIM Guide (1993).

FIM Levels		
	NO HELPER	FIMIT Step Calibration
7.	Complete Independence	24
6.	Modified Independence	8
	HELPER	
5.	Supervision	1
4.	Minimal Assistance	-5
3.	Moderate Assistance	-11
2.	Maximal Assistance	-17
1.	Total Assistance	-

Table 2. FIM Rating Scale, condensed from FIM Guide (1993).

Expected Measures on FIM Cognitive Items								
Item Name	Level:	1	2	3	4	5	6	7
N.	Auditory Comprehension	8	24	34	41	49	61	82
O.	Verbal Expression	5	22	31	39	47	59	80
P.	Social Interaction	11	27	37	44	52	64	85
Q.	Problem Solving	20	37	46	53	61	73	94
R.	Memory	18	34	44	51	59	71	92

Table 3. Expected FIMIT measures for each Level on each FIM Cognitive Items.

Tables of corresponding values of FIM raw scores and FIM measures (in FIMITs) are given in Heinemann et al. (1994), as well as item calibrations in logits. For the purposes of constructing the KeyFIM, the Cognitive score-to-measure conversion table (op. cit., Table 4) was recomputed based on a random sample of 15,439 relevant patient records from the Uniform Data System (UDS) database using the BIGSTEPS computer program (Linacre & Wright, 1991). For the purposes of constructing the KeyFIM, a useful substantive range was obtained when the linear conversion is 12.5 FIMITS per logit. Table 1 contains FIMIT calibrations for the FIM item difficulties for this sample. Table 2 contains FIMIT calibrations for the adjacent category (step) calibrations. Table 3 contains the expected FIMIT measure corresponding to each possible rating on each FIM item. Since the expected measure for an extreme category is infinite, i.e., out of the operational range of the FIM, a Bayesian adjustment is made so that, for the extreme categories 1 and 7, the measures corresponding to expected FIM ratings of 1.25 and 6.75 are listed.

For most IGCs (except 1.1, 2, 12)		
FIM raw score on 5 cognitive items	FIMIT measure	FIMIT S.E.
5	0	17
6	8	12
7	17	9
8	22	7
9	25	6
10	28	6
11	30	5
12	32	5
13	34	5
14	36	5
15	38	5
16	40	4
17	41	4
18	43	4
19	44	4
20	46	4
21	47	4
22	49	4
23	51	5
24	52	5
25	54	5
26	56	5
27	58	5
28	61	6
29	63	6
30	67	6
31	70	7
32	75	8
33	81	10
34	91	13
35	100	18

Table 4. FIM raw scores to FIMIT measures conversion table.

Table 4 contains a FIM cognitive raw score to FIMIT measure conversion table. This covers most impairment group codes (IGCs), except groups 1.1 (left-hemisphere stroke), 2 (brain dysfunction), and 12 (congenital deformity).

Constructing the KeyFIM

The measures and calibrations presented in Tables 1-4 are sufficient to draw the KeyFIM shown in Figure 1. To explain its features and demonstrate its use, the analysis of two patient records is described here.

Figure 1. KeyFIM data collection and analysis sheet.

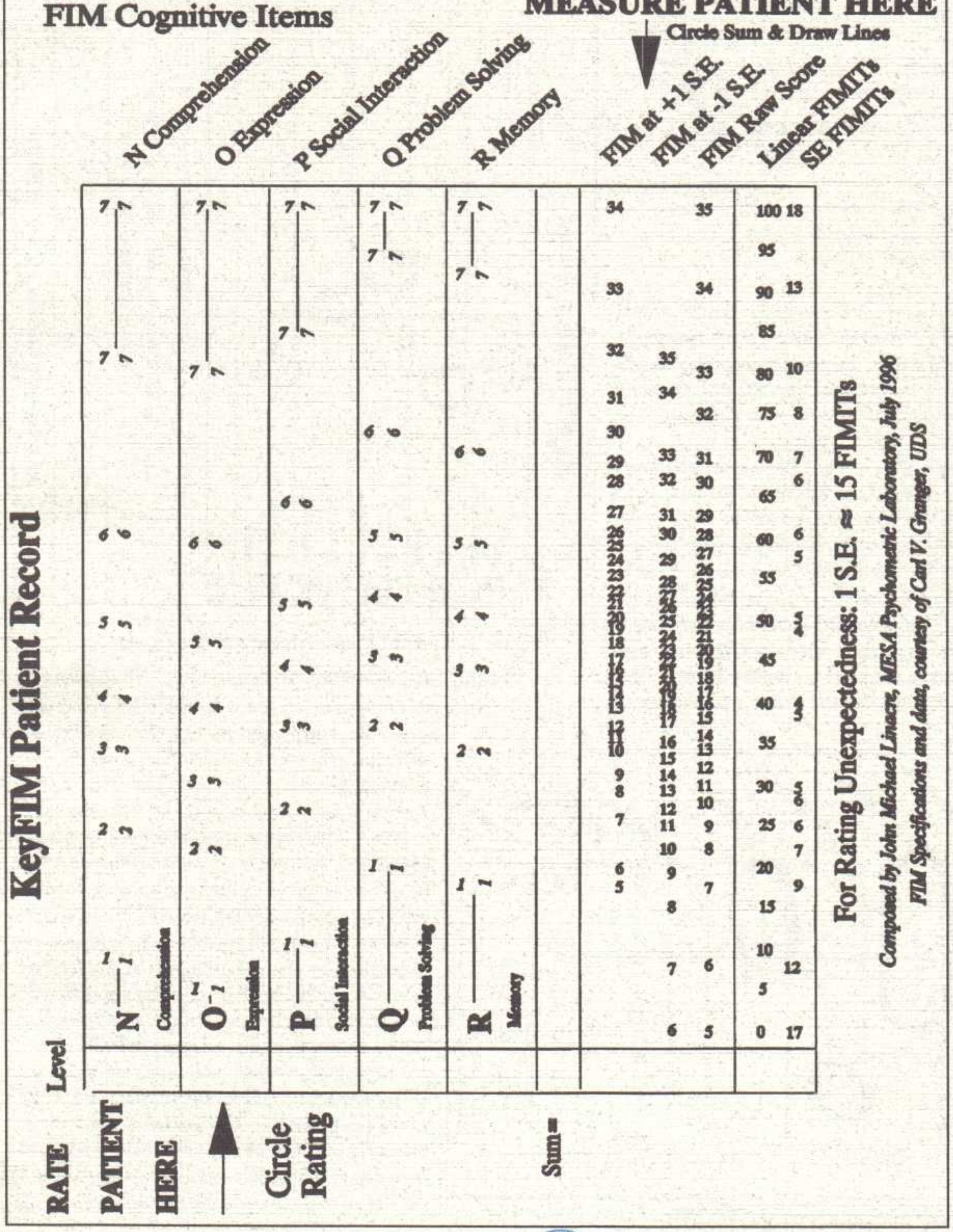

Figure 2 shows an actual patient record from IGC group 13, "Other Impairments." The KeyFIM has been turned on its side and the FIM levels recorded for each of the 5 cognitive items: 3 on item N. Comprehension, 3 on item O Expression, etc. The FIM ratings total 16. The corresponding levels are circled in the body of the KeyFIM. Data collection is now completed.

Figure 3 depicts the analysis stage. The KeyFIM is rotated, and a line drawn through the FIM raw score of 16 in each of three columns. The column "FIM at +1 S.E." indicates a high measure corresponding to one standard error of measurement above the expected measure. Continuing the line, by eye, to the "Linear FIMITs" column, indicates that a high measure corresponding to a raw score of 15 is about 45 FIMITs. The column, "FIM at -1S.E.," indicates a low measure one standard error below the expected measure. The "Linear FIMITs" column indicates that this is about 35 FIMITs. The third column, "FIM Raw Score," indicates that the expected measure for a score of 16 is about 40 FIMITs. The right-most column indicates that the standard error of this mea-

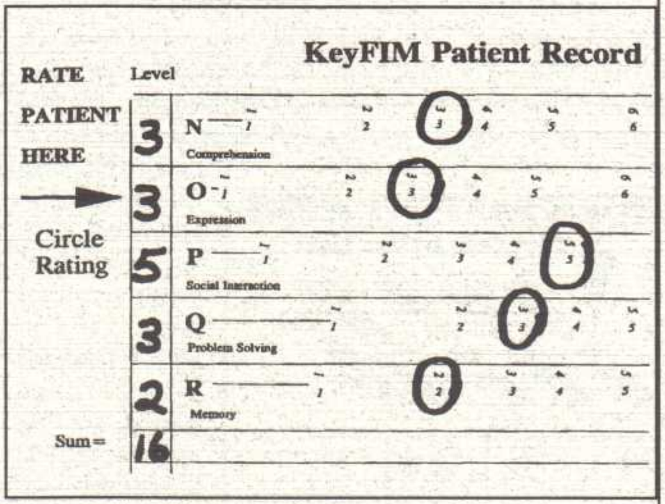

Figure 2. KeyFIM data collection.

sure is about 4 FIMITs, i.e., about the range 35-45 as illustrated. The legend on the right of the Figure states "For Rating Unexpectedness: 1 S.E. = 15 FIMITs." Based on conventional statistical testing, observations located further than 30 FIMITs from the mean line would be suspect, but here the most outlying, "5" on Social Interaction, is only 15 FIMITs away.

In this example there are no observations in extreme categories, but these require special treatment. A rating at an extreme level "1" or "7" corresponds to an infinite range of performance away from the next most extreme category. Accordingly, this is shown by a "—" on the KeyFIM. Thus for "7" on N. Comprehension, the KeyFIM has "7—7." This means that any location along the "—" is a reasonable location for the rating to be marked on the form. In practice, ring around the entire region, as in Figure 4, and choose the point on the line most consistent with the other ratings for measurement and fit analysis purposes.

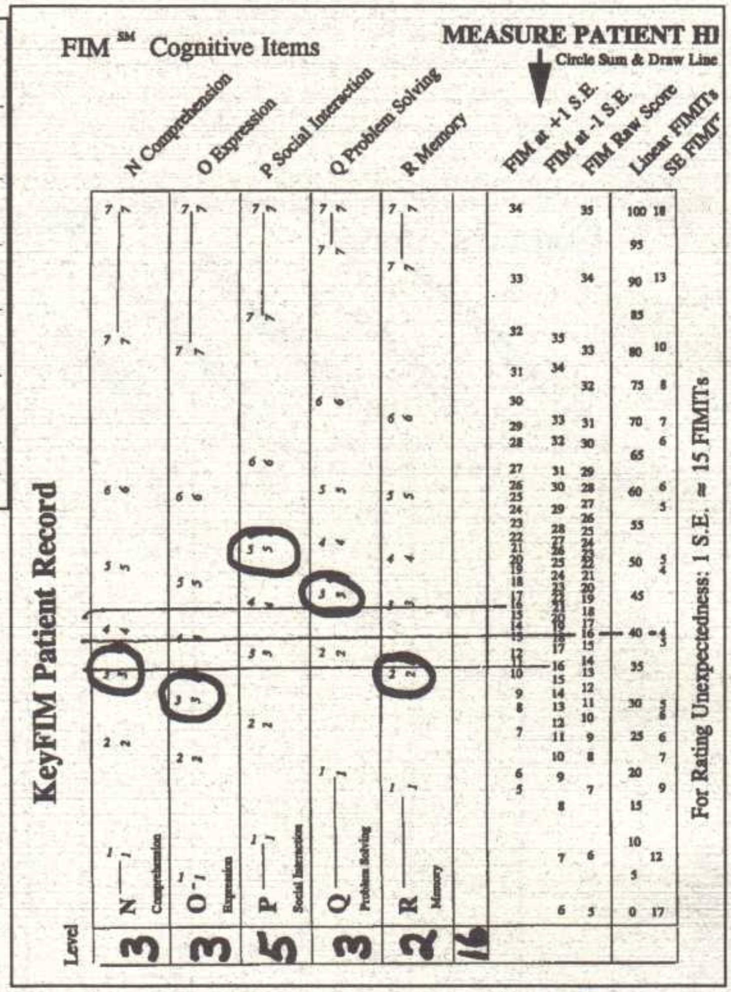

Figure 3. KeyFIM Measurement and Fit Diagnosis.

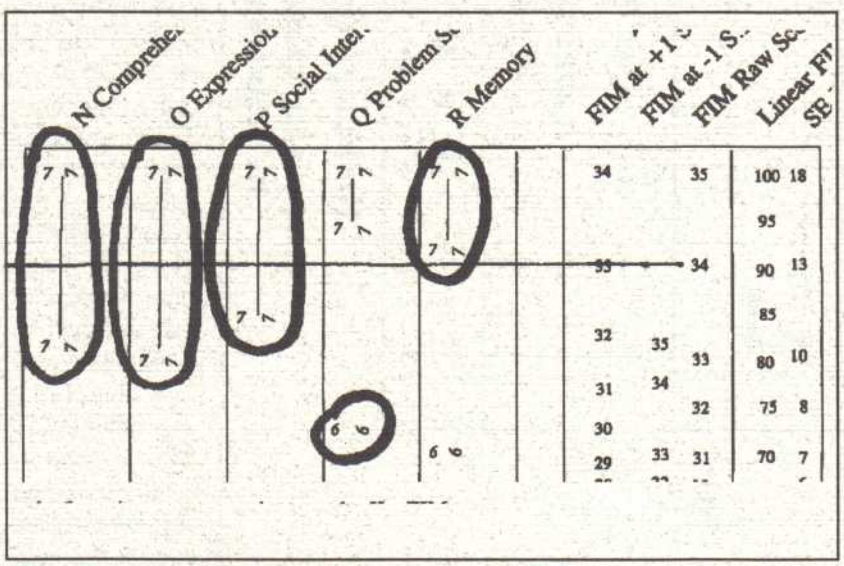

Figure 4. Locating extreme ratings on the KeyFIM.

Instantaneous Measurement and Diagnosis

Since each FIM item provides a locally independent measure of functional independence, they can be used as the basis for an intuitive, rather than statistical, measurement process. Figure 5 provides an example of another actual patient record. Here the observation to item R. Memory has been deliberately omitted — as though it were not yet recorded, perhaps never to be. There is no "complete" raw score, so the horizontal lines cannot be drawn directly. Intuitively, it is clear that the patient's typical level of independence is described by the higher ratings. A line has been drawn by eye through these, yielding a general independence of 58 FIMITs. The S.E. of this measure will be greater than the indicated 5 FIMITs due to the missing observation and discrepant rating pattern, treating the precision of this measure as 8 FIMITs would be reasonable. The low rating of "2" on Expression is at 20 FIMITs, about 38 FIMITs below the typical level. 38 FIMITs is twice the rating S.E. of 15 FIMITs, so that this rating is statistically unexpected.

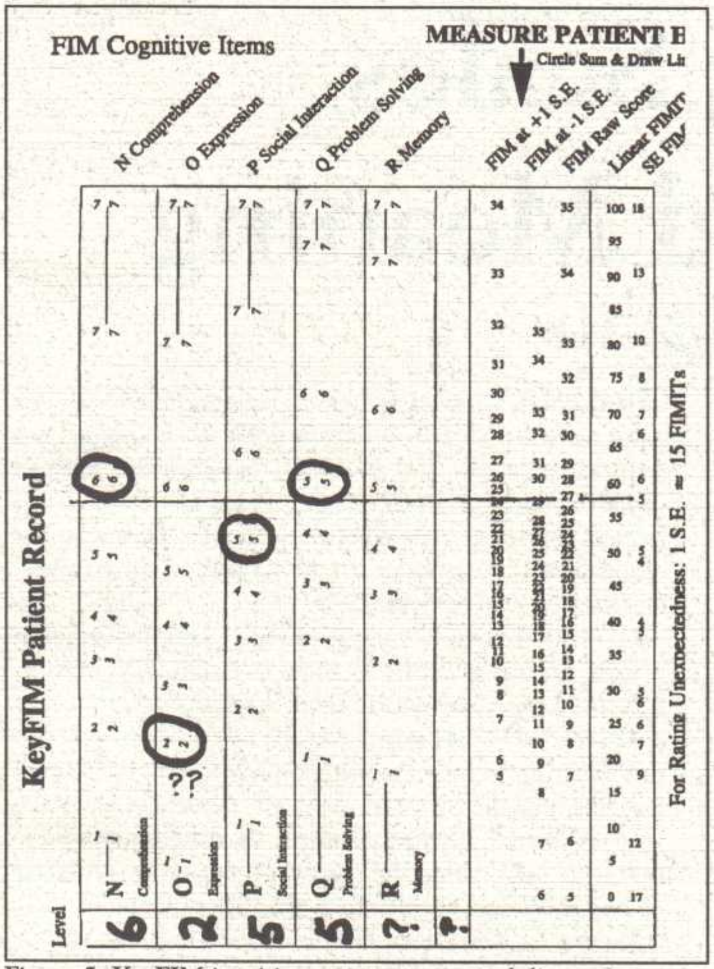

Figure 5. KeyFIM intuitive measurement and diagnosis.

More important for practice, however, is that it is obviously an outlier according to Leonard "Jimmy" Savage's "intra-ocular traumatic test." For clinical practice, it is this rating that will motivate the patient's immediate therapy.

In this example, measurement and fit diagnosis proceeded successfully and immediately despite incomplete data and the inability to use a "complete" raw score as the basis of analysis. Further, fit analysis and diagnosis could have proceeded successfully even without any formal statistical tests.

Conclusion

The KeyFIM is an example of how any rating instrument can be presented as a self-measuring form, supporting intuitive measurement and fit diagnosis. Its format encourages the practitioner to evaluate the ratings as they are being collected, so avoiding obvious data entry errors and misunderstandings. With a little experience, the practitioner can perform measurement and fit analysis in the same immediate, effortless and routine way that useful measurements are obtained from bathroom scales and clinical thermometers. The KeyFIM and instruments like it further blur the artificial distinction between physical and psychological measurement.

Notes:

1. "The term accuracy usually denotes in some sense the closeness of the measured values to the true value, taking into consideration both precision and bias. Bias [is] defined as the difference between the limiting mean [of observations] and the true value" (Ku H.H. 1967). See also "Use of the Terms Precision and Accuracy as Applied to the Measurement of a Property of Material" (ASTM Designation, E177-61T, 1961).

2. The Bayesian adjustment for extreme scores and ratings employs this line of reasoning: the KeyFIM would not have been administered to the patient if there were no chance that the patient might have been observed in a non-extreme category. Accordingly, the observation in the extreme category was barely enough to qualify as extreme. For extreme scores, this corresponds to an unobservable raw score that is 0.5 raw score points away from the extreme, i.e., a raw score of 35 out of 35 is treated as a score of 34.5, and a raw score of 5 out of 35 is treated as a score of 5.5. For individual ratings, performances in the range 1.5 to 2.5 would be observed as ratings of level "2." Ratings less than 1.5 would be observed in the extreme level of "1." Consequently any performance from 1 to 1.5 is observed as "1," and a "1" is treated as an "average" rating of 1.25 for the purposes of locating the category on the KeyFIM. Similarly, a "7" is treated as a 6.75. References:

Granger C.V., Hamilton B.B., Linacre J.M. Heinemann A.W., Wright B.D. (1993) Performance profiles of the Functional Independence Measure. American Journal of Physical Medicine and Rehabilitation 72:2 April 84-89.

FIM Guide (1993) Guide for the Uniform Data Set for Medical Rehabilitation (Adult FIM). Version 4.0. Buffalo, New York: State University of New York at Buffalo.

Heinemann A.W., Linacre J.M., Wright B.D., Hamilton B., Granger C.V. (1994) Measurement characteristics of the Functional Independence Measure. Topics in Stroke Rehabilitation 1(3) p.1-15. Fall.

Ku H.H. (1967) Statistical Concepts in Metrology. Chapter 2 in Handbook of Industrial Metrology. American Society of Tool and Manufacturing Engineers. p. 20-50. New York: Prentice-Hall.

Linacre J.M. (1993) Why logistic ogive and not autocatalytic curve? Rasch Measurement Transactions 6:4 p. 260-261.

Linacre J.M. (1995) KeyFIM -Self-Measuring Score Form. Rasch Measurement Transactions 9:3 p. 453-4.

Linacre J.M., Heinemann A.W., Wright B.D., Granger C.V., Hamilton B.D. (1994) The structure and stability of the Functional Independence Measure. Archives of Physical Medicine and Rehabilitation 75(2), 127-132, Feb.

Linacre J.M., Wright B.D. (1991) BIGSTEPS Rasch measurement computer program. Chicago: MESA Press.

Stevens S.S. (Ed). (1951) Handbook of experimental psychology. New York: Wiley

Wright B.D., and Stone M.H. (1979) Best Test Design. Chicago IL.: MESA Press.

Youden W.H. (1954) Instrumental Drift. Science 120:3121 p. 627-631. October 22, 1954.

John Michael (Mike) Linacre, Ph.D., M.A., C.D.P., C.C.P.

Dr. Linacre is Associate Director of the Measurement, Evaluation and Statistical Analysis (MESA) Psychometric Laboratory at the University of Chicago. After obtaining a degree in Mathematics from Cambridge University in 1967, he engaged in computer-related activities in England, Japan, Australia, and the USA. In 1981, he worked with Prof. Benjamin Wright to develop the Rasch analysis computer program, Microscale. In 1986, Mike moved to the University of Chicago and obtained a Ph.D. in psychometrics. Since then he has conducted research, taught classes and continued the development of Rasch computer programs, most recently Facets and WINSTEPS.