Journal of Occupational Rehabilitation, Vol. 15, No. 2, June 2005 (© 2005)
DOI: 10.1007/s10926-005-1219-7

Reliability and Validity of a New Computer-Administered Pictorial Activity and Task Sort

John Mayer,[1,2] Vert Mooney,[1] Leonard Matheson,[1] Scott Leggett,[1] Joe Verna,[1] Greg Balourdas,[1] and Greg DeFilippo[1]

Development of a new pictorial activity and task sort and examination of its reliability and validity is described. The Multidimensional Task Ability Profile (MTAP) is the latest in a series of measures that use a combination of drawings and task descriptions in a self-report format to assess functional capacity. The MTAP is found to be reliable on a test-retest and split-half basis. The concurrent validity of the MTAP was examined in performance testing of lift capacity. Results demonstrate that the MTAP has good concurrent validity.

KEY WORDS: function sorts; functional capacity evaluation; self-report measures; questionnaires.

INTRODUCTION

Self-report assessment through the use of pictorial activity and task sorts is an efficient means of gathering information about functional capacity. In a brief period of time, a large amount of information can be collected that reflects the evaluee's perception of his or her strength, endurance, work capacity, tool use abilities, and other information that is valuable for rehabilitation planning. These measures often are used in conjunction with functional capacity evaluation (FCE) to cross-reference self-perceived abilities with measured abilities.

A recent review of the design characteristics and uses of pictorial activity and task sorts (1) describe the history of these instruments and their uses. The focus of this paper is on one of the applications described in this review that has the potential to facilitate measurement-driven rehabilitation services and significantly reduce the cost of FCEs. This application can also be used as a self-report assessment instrument in serial testing as a substitute for repeated performance testing to mark progress in treatment. Once the self-report instrument has been calibrated on the evaluee during the FCE, repeated measurement with the self-report instrument provides indicators of improvement without the need to do more elaborate and expensive performance testing. The information provided by the self-report instrument used in this way allows the clinicians and case manager to confirm that progressing in treatment and also indicates when follow-up performance testing is necessary. Compared

[1]US Spine & Sport Foundation, 3444 Kearny Villa Road, Suite 307, San Diego, California.
[2]Correspondence should be directed to John Mayer, US Spine & Sport Foundation, 3444 Kearny Villa Road, Suite 307, San Diego, California 92123; e-mail: jmayer2@san.rr.com.

to serial FCE, the safety, ease of use, and cost savings of serial self-report assessment has led to widespread use. Unfortunately, the psychometric studies that are necessary underpinnings of such applications are rare. The present research studied the psychometric characteristics of a new computer-administered pictorial activity and task sort, the Multidimensional Task Ability Profile (MTAP). The reliability of the MTAP on a split-half basis is studied, followed by a study of the test-retest reliability of the instrument. Subsequently, the concurrent validity is examined in concurrent performance testing of lift capacity.

LITERATURE REVIEW

Self-report questionnaires have been developed for many types of health conditions, with several developed for use in occupational rehabilitation, including the Oswestry Disability Questionnaire (ODQ), the Roland–Morris Disability Questionnaire (RMDQ), and the Arthritis Impact Measurement Scale (AIMS). The ODQ (2) and the RMDQ (3) were designed for use with adults who have low back pain. The AIMS (4) was designed for use with persons who are impaired due to arthritis. All of these instruments are in widespread use. One of the reasons for their popularity is the relative efficiency of data collection. In a few minutes, a broad array of data can be collected that pertains to the functional impairments, limitations, and symptoms experienced by the evaluee. This information can be very useful for program planning, helping clinicians make decisions about treatment. However, these pen and paper questionnaires have two important limitations for use in occupational rehabilitation. The first is that the items in each instrument enquire about a combination of impairments, functional limitations, and symptoms, rather than focusing only on functional limitations. The second is that none of these instruments has a work-related point of reference, but consider an unlimited spectrum of activities. Whether or not the evaluee can lift 20 pounds at work, for example, is still unkown.

With regard to the combination of impairments, functional limitations, and symptoms, the problem with this approach is that the resulting score on these instruments is not unambiguous, especially in the midrange. An evaluee can have a score that is elevated due to any combination of impairments, functional limitations, and symptoms. Because of this, difference scores (comparisons of scores across evaluees) and change scores (comparisons of scores over two or more points in time) are not equivalent. Although reliability studies of these instruments have generally reported moderate to good results, the combination of items tends to blunt the sensitivity and specificity of the measures, so that validity studies are less robust.

With regard to the absence of a work-related point of reference, utility of this has become increasingly apparent as occupational rehabilitation services become more focused on return to work as the primary outcome of interest. Instruments that are linked to occupational databases have an important advantage in this regard.

Self-report measures that focus on the evaluee's ability to perform work and that provide indices of work capacity were first developed in the 1970s. These instruments used pictures of work tools or activities that were sorted by the evaluee to develop a rating of the evaluee's work capacity. The first pictorial activity and task sort was the RISC Tool Sort (RTS) (5), developed to collect data about the evaluee's abilities to use various tools for computerized matching of the evaluee's profile with the Dictionary of Occupational Titles (6) database. The RTS was based on the idea that the match of the person to the job could be achieved through consideration of the person's abilities to use the work tools for

the job. The RTS used a card deck with tools depicted on each card in a sorting task to collect information about the physical functional capacity of the worker, with scores scaled in terms of the DOT "strength" physical demand factor. The usefulness of the RTS led to the development of the WEST Tool Sort (7) and the Loma Linda Activities Sort (8), both of which were card sorts focused on the tools and equipment required in jobs. On the basis of the developers' experience with these instruments, the Spinal Function Sort (SFS) and Hand Function Sort (HFS) were developed. The SFS (9) is a set of 50 items in a test booklet that are pen and ink drawings with a short task description of a person involved in a work task. The HFS (10) used items similar to the SFS to increase sensitivity to more severe functional limitations as well as limitations of the hands and upper extremities.

A NEW MEASURE

Technology has continued to evolve in occupational rehabilitation. An important step in this evolutionary process was the development of the *Functional Assessment Constructs* (FAC) taxonomy, a comprehensive set of 131 assessment constructs that are used in disability determination systems in the United States. The FAC taxonomy (11) includes 46 constructs that pertain to physical capacity and are frequently used to measure musculoskeletal disability. To tap these constructs, the Multidimensional Task Ability Profile (MTAP) was developed (12) based on items from the SFS and the HFS. After combining these items and eliminating redundant items, 99 items were submitted to an expert judgment exercise, with experts in orthopaedics, physical therapy, occupational therapy, vocational evaluation, and exercise science. This exercise was used to identify linkages to the 46 FAC constructs. When minimal coverage or voids in coverage were found, new items were proposed. On the basis of this procedure, 12 new items were constructed. Subsequently, the set of 111 items were cross-referenced to the Dictionary of Occupational Titles strength factor (13,14) and to the physical abilities section of the Fleishman Job Analysis (15), which has been adopted as the physical abilities domain in O*NET (16). The resulting 111 MTAP items are comprised of a drawing of a common functional activity of daily living or work accompanied by a short description, with a rating scale identical to the HFS. A sample MTAP item is presented in Fig. 1.

MTAP items are presented to the evaluee by computer, beginning with items that have the least physical demand. Simultaneous with the pictorial item presentation, a written text description and an auditory description are presented by the computer to the patient in the patient's primary language. After selecting a response with mouse or touch screen, the patient uses the mouse or touches screen to indicate that the response is correct, at which time the computer presents the next item. Response options are presented along a 5-point Likert scale from "Able" to "Unable." The computer performs scoring of the evaluee's responses. The weighted scoring strategy from the Spinal Function Sort is used to provide a global score that is linked to the Dictionary of Occupational Titles strength factor (13,14). Individual scale scores are provided in a profile. A sample profile is shown in Fig. 2. If the test is administered on a follow-up basis, a "report card" is provided to the patient (Fig. 3).

PURPOSE OF RESEARCH

The reliability and validity of an instrument are basic requirements of its utility. A test-retest reliability study was designed to examine the reliability of the MTAP in a sample of

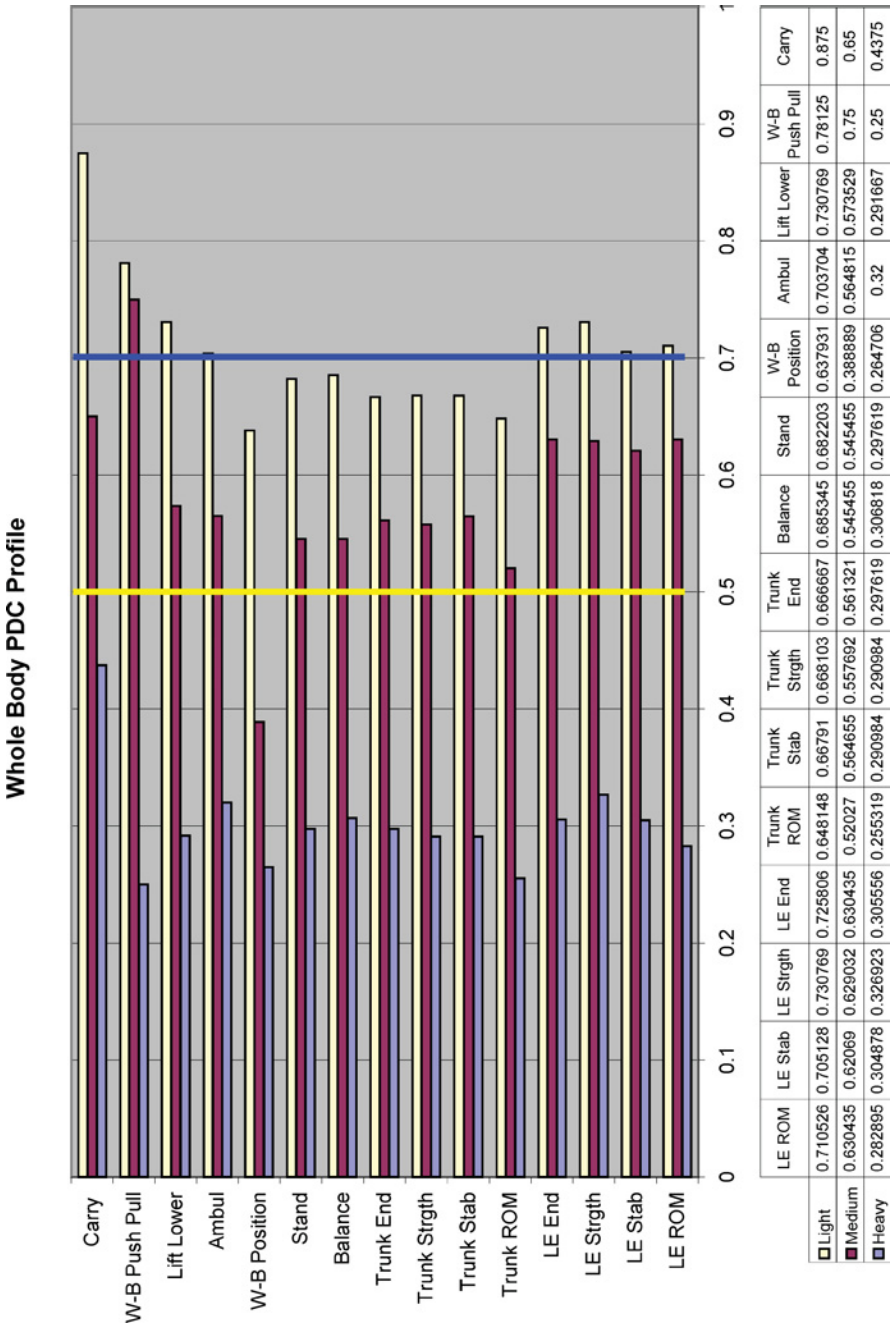
Fig. 1. Sample MTAP item ("Carry of load of laundry upstairs").

persons with musculoskeletal impairment. To examine the validity of the MTAP, a concurrent validity study was designed with this sample; comparing MTAP scores with subjects' lift capacity using a standardized measure.

METHODS

Subjects

The present studies involve use of three convenience samples. The reliability study used two samples of subjects. The first sample included subjects who were patients involved in clinical treatment for musculoskeletal disorders, primarily of the spine, at a standard



	LE ROM	LE Stab	LE Strgth	LE End	Trunk ROM	Trunk Stab	Trunk Strgth	Trunk End	Balance	Stand	W-B Position	Ambul	Lift Lower	W-B Push Pull	Carry
Light	0.710526	0.705128	0.730769	0.725806	0.648148	0.66791	0.668103	0.666667	0.685345	0.682203	0.637931	0.703704	0.730769	0.78125	0.875
Medium	0.630435	0.62069	0.629032	0.630435	0.52027	0.564655	0.557692	0.561321	0.545455	0.545455	0.388889	0.564815	0.573529	0.75	0.65
Heavy	0.282895	0.304878	0.326923	0.305556	0.255319	0.290984	0.290984	0.297619	0.306818	0.297619	0.264706	0.32	0.291667	0.25	0.4375

Fig. 2. Sample MTAP profile.

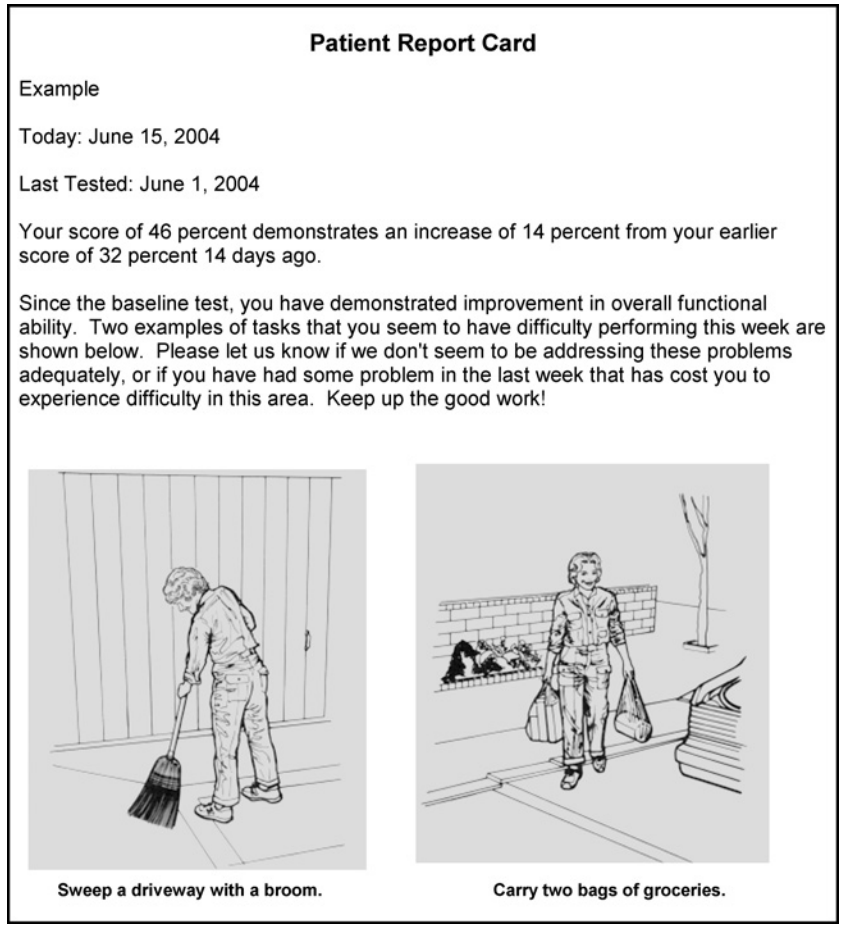


Fig. 3. Sample MTAP report card.

physical therapy center. The second sample included subjects who were patients enrolled in a medical exercise program for a variety of musculoskeletal complaints at a fitness center for older adults. Using recruitment procedures reviewed and authorized by the first author's institutional review board, 44 subjects were recruited. Subjects were tested while receiving treatment at two sites in the United States. Subject characteristics are presented in Table I.

The validity study used a sample of subjects in clinical treatment who had painful musculoskeletal impairments involving the spine. Using recruitment procedures reviewed

Table I. Demographic Characteristics of Sample 1 and Sample 2

	Sample 1 [Mean (SD)]		Sample 2 [Mean (SD)]	
Variable	Male ($n = 11$)	Female ($n = 9$)	Male ($n = 8$)	Female ($n = 16$)
Age (years)	39.8 (12.7)	44.1 (19.6)	71.3 (9.3)	68.3 (6.0)
Height (cm)	169.4 (8.4)	161.0 (10.2)	177.3 (6.4)	161.8 (5.3)
Weight (kg)	82.2 (12.2)	62.9 (10.8)	80.0 (11.1)	69.1 (8.9)

Table II. Demographic Characteristics of Sample 3

Variable	Male ($n = 40$)	Female ($n = 20$)
Age (years)	39.6 (9.9)	42.4 (9.7)
Height (cm)	175.3 (9.1)	167.4 (5.8)
Weight (kg)	91.2 (14.2)	76.2 (21.6)

and authorized by the first author's institutional review board, 60 subjects were recruited. Subjects were tested while receiving treatment at three sites throughout the United States. Subject characteristics are presented in Table II.

Procedures

Reliability Study

After completing the informed consent process, each subject was oriented to the MTAP computer by a clinician and was shown how to use the mouse or touch screen to respond to each item. Each subject then completed the MTAP without additional assistance. Subjects required 12–15 min to complete the process, with no omitted items. After a period of 2 days, each subject completed the MTAP once again. At this time, each subject also received the SFS. The MTAP was administered before the SFS. All subjects completed all test administrations.

Validity Study

After completing the informed consent process, each subject completed the MTAP. Immediately followed by the MTAP test, each subject completed the EPIC Lift Capacity (ELC) test (17). The ELC test is a standardized evaluation of lift capacity that uses six stages of progressive lifts. The first three of the six subtests from the ELC were summed to derive the ELC score.

Data Analysis

To examine the internal consistency of the MTAP, split-half reliability was analyzed using an intraclass correlation procedure comparing the responses to even items with the responses to odd items.

To examine the stability of the MTAP over time, test-retest reliability was analyzed with a 2-day interval using both an intraclass correlation procedure and a Pearson product-moment correlation for the interval data, and the Spearman rho procedure for ordinal data. The use of the intraclass correlation and the Pearson product-moment correlation is necessary due to the nonspurious time-linked changes that are likely to occur with such measures while subjects are in treatment.

To examine the concurrent validity of the MTAP in terms of an established similar measure, the SFS score was compared to the MTAP score using the Pearson product-moment correlation. To examine the validity of the MTAP in terms of an established performance test, the sum of the maximum acceptable weights from each of the "occasional" subtests

Table III. Response to MTAP and Lift Capacity Test

Variable	Mean (SD)	
	Male ($n = 40$)	Female ($n = 20$)
MTAP total RPC score[a]	265.7 (91.4)	223.2 (79.3)
MTAP lift score[b]	10.4 (8.0)	6.4 (4.6)
ELC lift (kg)[c]	43.8 (31.2)	25.9 (21.5)

[a] Range from 0 to 444, with higher scores indicating more ability.
[b] Sum of weighted scores for MTAP items that involve lifting; 8 total; highest possible score = 32.
[c] Sum of maximum acceptable weight for the three occasional lifts in the ELC test.

from the ELC test was compared to the MTAP total score and the MTAP score for only those items that represent lifting tasks (36 total) using the Pearson product-moment correlation.

RESULTS

The reliability of the MTAP was considered first. After examining the responses of Samples 1 and 2 and finding no significant differences between samples in MTAP or SFS scores, the subjects were combined. Internal consistency of the MTAP was measured using the split-half procedure and was found to be ICC $r = 0.99$ ($p < 0.05$). Stability of the MTAP over time was studied with a 2-day interval and resulted in a Pearson product-moment correlation of $r = 0.95$ ($p < 0.05$). The intraclass correlation for the same data was ICC $r = 0.95$ ($p < 0.05$).

Concurrent validity of the MTAP was studied in relation to both the SFS and the ELC test. With Sample 1 and Sample 2 combined, the Pearson product-moment correlation between SFS and MTAP was $r = 0.89$ ($p < 0.05$). For Sample 3, the data that describe the responses to the MTAP and the ELC are presented in Table III.

As expected, and consistent with prior research with both instruments, male scores are substantially greater than female scores with both tests.

Using the data from Sample 3, three different analyses were conducted, the first considering the interval scores from the MTAP with the SFS score, the second considering the MTAP total score with the ELC test total, and the third considering the MTAP score from the "lift" subscale with the ELC total score (Table IV). A second set of analyses using this sample compared the MTAP total score and lift subscale score with the physical demand characteristics (PDC) rating of each subject from the ELC, a frequently used ordinal ranking method. The data that describe these comparisons are presented in Table V.

The comparison between the two self-report measures, MTAP and SFS, yields a correlation coefficient of $r = 0.91$ ($p < 0.05$), which was expected because of the use in the

Table IV. Pearson Correlations Between MTAP Scores, SFS Score, and Lift Capacity Score

Comparison	r
MTAP percent × SFS percent ($n = 60$)	0.91
MTAP total × ELC lift ($n = 60$)	0.77
MTAP lift × ELC lift ($n = 57$)	0.89

Table V. Spearman Correlations Between MTAP Scores and Lift Capacity PDC Rating

Comparison	ρ
MTAP total × ELC PDC ($n = 60$)	0.77
MTAP lift × ELC PDC ($n = 57$)	0.83

MTAP of SFS items. The primary difference between the two is the greater number of items in the MTAP and the computerized administration of the MTAP, while the SFS is a paper and pencil test. The comparisons between both MTAP scores and each subject's ELC performance indicate high levels of concurrent validity, with the stronger correlation found between the MTAP subscale and the ELC performance. This is somewhat surprising, given the smaller number of items in the subscale (36 vs. 111). When subjected to a Fisher r to z transformation, the difference between the two correlation coefficients is not statistically significant at $p < 0.05$.

The comparison between the MTAP scores and each subject's rating on the PDC chart was of interest because the comparison of person to occupation in occupational rehabilitation clinics often employs this ordinal ranking scale. The present study demonstrated that each correlation coefficient is degraded slightly, probably as a consequence of the error introduced through the use of this second-order measurement system based on the subject's ELC performance. Translation of ELC test performance into the ordered ranking system is likely to introduce error variance, which slightly degrades the correlation between MTAP scores and ELC performance. However, the resulting Spearman rho correlation demonstrated excellent validity.

DISCUSSION

This series of studies examined the reliability and validity of a new computer-administered pictorial activity task sort instrument, the MTAP, in samples of adults involved in occupational rehabilitation. Reliability on both a split-half basis and a test-retest basis was found to be excellent. Validity was found to be good on a concurrent validity basis; comparing MTAP total scores and lift subscale scores with performance on both a standardized lift capacity test and an established pictorial activity task sort.

The good psychometric properties of the MTAP make it a potentially useful tool to augment functional capacity evaluation in occupational rehabilitation. Given the high level of test-retest reliability, the sensitivity of the instrument to change across time in response to treatment is likely to be high. This will require formal analysis. Given the good concurrent validity, the use of the instrument as a substitute for aspects of a functional capacity evaluation, such as a lift capacity test, is reasonable to consider. This also will require formal analysis. In practice at the clinics in which the data were collected, progress in therapy is tracked through serial administration of the MTAP every 2 or 3 weeks, with administration of the MTAP at discharge along with selected subtests from a standard functional capacity evaluation. This appears to be a useful and efficient application, the validity of which requires formal study.

One likely reason for the high levels of reliability in the MTAP is the large number of items included in the instrument. As the number of items in an instrument increases,

random error becomes less important. In the classical testing model, the proportion of true score variability relative to the total observed variability is directly related to the number of items in the instrument. The more the number items, the less the importance of random error. Unfortunately, the use of a large number of items is limited by the time required for processing each item and the evaluee's motivation, ability to sustain attention, and energy. In the studies described herein, the subjects' performance indicated that seven MTAP items to nine items can be processed per minute, over a 12–15-min test session, with no omitted items. This compares with three items to five items per minute with the SF-36 (18) or four items per minute with the Beck Depression Inventory II (19). The improvement in per-item efficiency while reliability is maintained may be a function of the combination of pictorial and text task descriptions in each item along with computerized item presentation. The reasons for efficiency of item processing require additional study.

SUMMARY

The psychometric properties of a new computer-administered pictorial activity task sort, the Multidimensional Task Ability Profile (MTAP) were studied in samples of adults receiving exercise therapy and physical therapy as part of their rehabilitation from musculoskeletal injuries. All subjects were able to complete all items in the 111-item instrument in 12–15 min. Comparison of MTAP scores collected on a test-retest basis demonstrated excellent reliability. Comparison of MTAP scores with responses to an established self-report test instrument and performance on a standardized lift capacity test demonstrated good concurrent validity. The high levels of reliability and concurrent validity suggest that the sensitivity and specificity of the MTAP may make the instrument useful for tracking change over time and outcome from treatment programs. This will require additional research.

ACKNOWLEDGEMENT

The authors participated in the development of the Multidimensional Task Ability Profile and received benefits from its use.

REFERENCES

1. Matheson L. History, design characteristics, and uses of the pictorial activity and task sorts. *J Occup Rehabil* 2004; 14(3): 175–195.
2. Fairbank J, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980; 66(8): 271–273.
3. Roland M, Morris R. A study of the natural history of back pain: Development of a reliable and sensitive measure of disability in low-back pain. *Spine* 1983; 8(2): 141–144.
4. Meenan R, Gertman P, Mason J. Measuring health status in arthritis: The arthritis impact measurement scales. *Arthritis Rheum* 1980; 23(2): 146–152.
5. Matheson L. *RISC Tool Sort*. Woodland Hills, CA: Rehabilitation Institute of Southern California, 1979.
6. U.S. Department of Labor. *Dictionary of occupational titles*, 4th edn. Washington, DC: U.S. Department of Labor, 1977.
7. Matheson L. *WEST Tool Sort*. Signal Hill, CA: Work Evaluation Systems Technology, 1981.
8. Anzai D. *Loma Linda Activity Sort*. Fort Bragg, CA: Work Evaluation Systems Technology, 1984.
9. Matheson L, Matheson M, Grant J. Development of a measure of perceived ability. *J Occup Rehabil* 1993; 3(1): 15–30.

10. Matheson L, Kaskutas V, Mada D. Development and construct validation of the Hand Function Sort. *J Occup Rehabil* 2001; 11(2): 75–86.
11. Guadino E, Matheson L, Mael F. Development of the functional assessment taxonomy. *J Occup Rehabil* 2001; 11(3): 155–175.
12. Mooney V, Matheson L. *Multidimensional task ability profile*. San Diego, CA: Mind Trust, 2003.
13. U.S. Department of Labor. *Dictionary of occupational titles, classification of jobs*, 4th supplement edn., Vols. I and II. Washington, DC: U.S. Department of Labor, 1986.
14. U.S. Department of Labor. *Handbook for analyzing jobs*. Washington, DC: Manpower Administration, 1972.
15. Fleishman E, Reilly M. *Administrators guide: Fleishman's job analysis survey*. Palo Alto, CA: Consulting Psychologists Press, 1992.
16. Peterson N, Mumford M, Borman W, Fleishman E. *An occupational information system for the 21st century: The development of O*NET*. Washington, DC: American Psychological Association, 1999.
17. Matheson L, Mooney V, Grant J, Affleck M, Hall H, Melles T, Lichter R, McIntosh G. A test to measure lift capacity of physically impaired adults: Part I. Development and reliability testing. *Spine* 1995; 20(19): 2130–2134.
18. Ware J. *SF-36 Health Survey*. Boston, MA: Health Institute New England Medical Center, 1998.
19. Beck A, Steer R, Brown G. *Beck Depression Inventory II*. Dallas, TX: Harcourt, 1996.