Journal of Occupational Rehabilitation
https://doi.org/10.1007/s10926-019-09842-5

Validity of the Multidimensional Task Ability Profile

Joe L. Verna[1,2] · Leonard N. Matheson[1,3] · Sharon Scherer[2] · John M. Mayer[1,4,5]

Abstract
Background The Multidimensional Task Ability Profile (MTAP) is a patient-reported outcome (PRO) measure that provides a global score linked to the physical demand characteristics of work, but needs to be validated against established measures. *Purpose* To assess the concurrent validity of the MTAP compared with the Oswestry Disability Index (ODI), Neck Disability Index (NDI), Disabilities of the Arm, Shoulder, and Hand (DASH), Lower Extremity Functional Scale (LEFS), and Short Form 12 Health-Related Quality of Life (SF-12) questionnaires. *Methods* An observational study was conducted in 157 patients undergoing musculoskeletal rehabilitation. At baseline and after 30 days of treatment, patients completed the MTAP, ODI, NDI, DASH, LEFS, and SF-12 and provided self-reported work status. *Results* At baseline and after 30 days, convergent validity between the MTAP and DASH, LEFS, NDI, and ODI was good to excellent. Concurrent validity between the MTAP and SF-12 physical component score (PCS) and mental component score (MCS) was moderate or fair, respectively. Sensitivity to change over the 30-day treatment interval was established for the MTAP, SF-12 PCS, SF-12 MCS, and LEFS. Fair to moderate predictive validity for work status was found for the MTAP, ODI, NDI, DASH, and SF-12 PCS. *Conclusions* The MTAP demonstrated adequate concurrent validity, predictive validity, and sensitivity to change compared to other PROs. For patients with various impairment types, the MTAP may be a useful omnibus measure to supplement specialty instruments such as the DASH, NDI, ODI, or LEFS.

Keywords Patient reported outcome measures · Physical function · Musculoskeletal rehabilitation

Introduction

Over the recent decades, patient-reported outcomes (PROs) of physical function have become standard tools in musculoskeletal rehabilitation [1, 2]. These measures provide useful information about a patient's functional and work capacity in a safe, efficient, and cost-effective manner [3]. U.S. governmental agencies, such as Medicare and Medicaid, recommend PROs for routine use [4]. The American Medical Association, in its Guides to the Evaluation of Permanent Impairment [5] recommends PROs to assess quality of life, function, disability, and impairment of individuals [6]. Many states' workers' compensation programs endorse the use of PROs to assist in case management as described in the Official Disability Guidelines [7]. In 2004, NIH began to fund the Patient-Reported Outcomes Measurement Information System (PROMIS) initiative to provide validated PROs for clinical research and practice [8]. Funding from NIH for this project continues and has expanded to include several languages and countries including China and Canada, underlining the legitimacy of self-report measures for both program evaluation and clinical research.

Recently, the Department of Health and Human Services (DHHS), Centers for Medicare and Medicaid Services (CMS) created a quality-improvement payment program for Medicare Part B [9]. The Merit-Based Incentive Payment (MIPS) will be introduced voluntarily in 2019 and will affect payments for services in 2021. Functional Outcome Assessment is one of the six quality reporting methods designated for Medicare providers. Defined by the Centers for Medicare and Medicaid Services as one of the "clinical

✉ John M. Mayer
johnmayer@usssf.com

1 Vert Mooney Research Foundation, San Diego, CA, USA

2 Spine & Sport, Inc, San Diego, CA, USA

3 Epic Neurorehabilitation & Psychological Services, Inc, Chico, CA, USA

4 Excellcior, LLC, Tampa, FL, USA

5 Vert Mooney Research Foundation, 3760 Convoy Street, Suite 101, San Diego, CA 92123, USA

Published online: 14 June 2019

quality measures", PROs are: "Patient completed questionnaires designed to measure a patient's physical limitations in performing the usual human tasks of living and to directly quantify functional and behavioral symptoms" [10]. Reimbursement will be increased for Medicare service providers who integrate these questionnaires into the patient's care plan and can demonstrate treatment efficacy.

Several PROs are used to assess physical function and the consequences of musculoskeletal impairment, such as the Oswestry Disability Index (ODI) [11], Neck Disability Index (NDI) [12], Disabilities of the Arm, Shoulder, and Hand (DASH) [13], Lower Extremity Functional Scale (LEFS) [14], and Short-Form 12 health-related quality of life questionnaire (SF-12) [15].

The items in most PROs, such as those described above, are entirely text, using language to convey meaning to the patient. The use of text alone to convey meaning may lead to language, culture, and literacy barriers to comprehension, which can result in inaccurate self-reporting and thereby limit the utility of questionnaires [16]. As an alternative to traditional questionnaires, pictorial activity test questionnaires are available, which contain items that combine both text and simple drawings to convey meaning [2, 3, 16, 17]. The use of a picture of an activity in combination with a short text caption describing the activity enables better comprehension and a lower level of ambiguity than text alone [18], thus improving accuracy [16], which is especially important when the measure is used with non-English-speaking persons and those who may have limited literacy [18].

The Multidimensional Task Ability Profile (MTAP) [3, 18–20] was developed through the use of the Rasch rating scale approach to Item Response Theory [3, 19–22], which provides item calibration to maximize the instrument's precision, proportional evaluation of the instrument's total score to facilitate normative comparisons, and two measures of internal consistency to screen for inconsistent self-report, the INFIT and OUTFIT scores [19]. The MTAP is composed of items depicting activities of daily living and work across multiple dimensions. The evaluee rates ability to perform each item on a five-level ordinal scale from "Able" to "Unable". Items are presented in a standardized order from less to more physically demanding [3] in both the computer-administered and paper and pencil versions. Scores are summed and have been linked to gender-specific and age-level normative data and to standard measures of performance such as the physical demand characteristics system used to categorize the five general levels of strength and energy demands of occupations in the Dictionary of Occupational Titles [23].

The MTAP was developed from a larger set of items that were administered to approximately 20,000 patients in outpatient occupational physical therapy clinics. Approximately 70% of this sample was primarily English speaking and was administered the MTAP questionnaire in English, while approximately 30% was primarily Spanish speaking and was administered the questionnaire in Spanish. The original version of the MTAP (MTAP 1.0) consisted of 111 items. Psychometric studies found that the first version of the MTAP was reliable and valid [3, 19], with good test–retest reliability, split–half reliability, and strong correlation to the physical demands of occupations performed by healthy people [3, 19].

The current version of the MTAP (MTAP 2.0) contains 50 items, which were selected through factor analysis and Rasch analysis of the original 111-item MTAP instrument [19]. A psychometric study indicated that MTAP 2.0 is reliable when administered in English and Spanish [18]. While some of the psychometric properties of the MTAP have been described, the relationship of the MTAP to commonly-used PROs has not been assessed.

The purpose of this study was to examine the concurrent validity of the MTAP 2.0 compared with seven frequently-used PROs in adult patients with musculoskeletal disorders enrolled in an exercise-focused treatment program.

Methods

Study Design

An observational study of repeated measures was conducted with a convenience sample of patients from three outpatient physical therapy centers in Southern California that use a measurement-driven treatment model to manage musculoskeletal dysfunction across numerous conditions. In this model, treatment is informed by PROs, such as the MTAP and other measures described in this paper, along with clinical observations and physical performance testing. Assessments are conducted serially and the subsequent treatment plan is adapted according to performance. Interventions include focused and progressive exercise, incorporating other therapeutic modalities, as needed.

Participants

Patients were recruited by word of mouth and posters to include an approximately equal number of males and females in proportion to the ethnic and racial characteristics of Southern California. One-hundred fifty-seven individuals with musculoskeletal disorders being treated at outpatient physical rehabilitation centers were enrolled in this study (provided consent). The sponsoring organization's institutional review board approved the experimental protocol. Each candidate for enrollment provided written informed

consent prior to participation. Participants were in various stages of care upon enrollment.

Sample Size Calculation/Power Analysis

The study enrolled from a sample of convenience in a pragmatic clinical setting across specific time periods. Thus, a formal power analysis was not conducted to determine sample size.

After providing consent, the participant completed a basic health history questionnaire and the Older Adult Health and Mood Questionnaire [24], in their native language (English or Spanish), and was interviewed by the investigator to determine eligibility.

Inclusion criteria were: active patient of an outpatient physical rehabilitation center being treated for a musculoskeletal disorder; age 18–70 years; literate (verbal and written) in English or Spanish language; and able to provide written informed consent. Exclusion criteria were: currently diagnosed with depression or a psychiatric condition or attaining a score of ten or higher on the Older Adult Health and Mood Questionnaire, indicating likelihood of clinical depression [24]; INFIT or OUTFIT scores on the baseline MTAP test above the "inconsistent" cut-point of 1.50 [25]; or unable or unwilling to complete the study's outcome questionnaires. Exclusion for depression was undertaken because its positive relationship to catastrophizing self-perceptions [26–29] were expected to distort the responses to the self-report measures.

Procedures

Administration of the MTAP

Immediately following the screening procedures, each participant completed the MTAP 2.0 in their native language (English or Spanish) as described elsewhere [3, 18]. Briefly, the MTAP 2.0 is a computer-administered and scored PRO comprised of 50 items consisting of drawings with text captions depicting a wide range of physical work tasks and activities of daily living (Fig. 1). For each item, the MTAP uses a five-level ordinal rating scale ("Able," "Slightly Restricted," "Restricted," "Very Restricted," "Unable"), with an additional response for "Don't Know." Items are presented one at a time, starting with items that have the least physical demand. Missed items are not allowed with the computer interface. The mean time to complete the MTAP is approximately 6 to 8 min [18]. Test number (i.e. baseline, follow-up) and language of administration (e.g., English, Spanish) add variation to the administration time. Once the participant completes the test, the computer automatically scores the responses with the following scoring rubric: Able = 4, Slightly Restricted = 3, Restricted = 2, Very Restricted = 1, Unable and Don't Know = 0. Item scores are summed over the instrument to calculate a global score ranging from 0 to 200, with higher scores indicating more ability. Research on the responses patterns for MTAP rating scale [19] indicated that it is appropriate to weight the Unable and Don't Know responses as 0 for three reasons. First, evaluees usually are able to estimate ability; Don't Know responses are rare. Second, the computer administration system prompts the evaluee who selects Don't Know to make an estimate of ability. Third, the Rasch rating scale model used to develop the MTAP [21, 30] provide a confirmable estimate of the evaluee's ability with an incomplete response set.

The confirmable estimate of ability that is provided by the Rasch item response theory approach includes two additional scores that indicate the degree to which the pattern of responses is what would be expected based on the participant's ability. The INFIT score is sensitive to responses that are different from the expected near the participant's ability level; changes in the rating scale responses should reflect the interaction between the participant's ability and

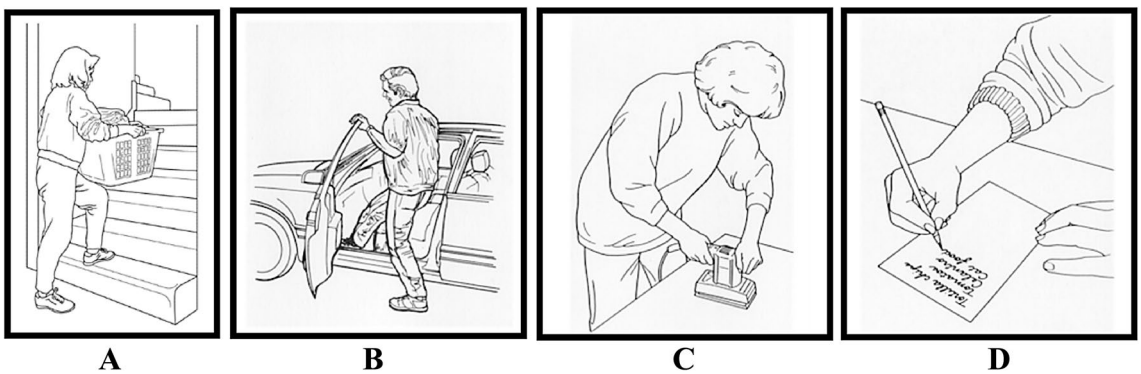

Fig. 1 Sample MTAP items: **a** carry a full laundry basket up one flight of stairs; **b** get into an automobile driver's seat; **c** use an electric sander to smooth a table top; and **d** write a shopping list with a pencil

the difficulty of the item. The OUTFIT score is sensitive to responses that are different from expected responses with items that are outliers compared to the participant's ability level; unexpected responses indicate misunderstanding of the testing process and/or misreading of items due to literacy problems, boredom, or fatigue. Scores above 1.50 on either INFIT or OUTFIT are considered suspicious of inconsistent responding [21, 30].

Administration of Other PROs

After completing the MTAP, the participant rated current pain intensity on a 10 cm visual analog scale (VAS) (0 = no pain, 10 = worst imaginable pain) [31] and the SF-12 [32]. Depending on the presenting condition, the participant also completed one or more of the following condition-specific PROs in their native language (English or Spanish)—DASH, LEFS, ODI, or NDI. The MTAP and VAS were completed electronically on a standard computer monitor in each clinic and the other PROs were completed on paper.

On a separate day approximately 2–3 days (Visit 2) after Visit 1, the participant returned to the study site and completed the MTAP and VAS. On a separate day approximately 30 days (Visit 3) after Visit 1, the participant returned to the study site and completed the identical tests as Visit 1—the MTAP, VAS, and SF-12, and depending on the presenting condition, also completed the DASH, LEFS, ODI, or NDI.

Intervention

Between baseline (Visit 1) and 30-day follow-up (Visit 3), the participant underwent an active outpatient physical therapy program for musculoskeletal conditions consisting of rehabilitative exercises, manual therapies, and adjunctive modalities using a measurement-driven approach [20]. MTAP results were available to the patient and providers.

Telephone Interview

Approximately 1–9 months (mean ± SD: 155 ± 80 days; range 32–260 days) after Visit 3, the participant completed a telephone interview (led by study personnel) to assess their work status through the following two questions: (1) "Were you working when you first started the study?" (2) "Are you working now?"

Data Management and Analysis

The MTAP and VAS were automatically scored by the computer system. Data from the other PROs were manually entered into an electronic spreadsheet and scored according to the instrument's scoring specifications. After auditing and arithmetic manipulation, the data were exported to SPSS (SPSS Statistics for Windows, Released 2017, Version 25.0, IBM Corp., Armonk, NY), which was used to perform statistical analyses, including basic descriptive analyses and analyses to examine temporal stability, concurrent validity, predictive validity, and responsiveness, as described below.

Temporal Stability

MTAP global scores and VAS ratings were compared on a test–retest basis between Visit 1 and Visit 2.

Responsiveness/Sensitivity

For each measure, global scores were compared across the time points (Visit 1 to Visit 3) to measure effect size.

Concurrent Validity

MTAP global scores were compared to the VAS and the global scores of the other measures at each time point (Visit 1 and Visit 3). Pearson product moment correlation coefficients were calculated. Convergent and discriminate validity: MTAP global scores were compared to physical-task-focused patient reported outcome measures as well as to patient reported outcome measures that focus on other constructs, such as mood, pain, and mental function.

Predictive Validity

Global scores for each measure at each time point (Visit 1 and Visit 3) were compared to reported work status.

Statistical significance was accepted at alpha < 0.05. The quality of Pearson correlations between measures was considered to be "poor" for $r < 0.25$, "fair" for $r = 0.25$ to < 0.50, "moderate" for $r = 0.50$ to < 0.75, and "good to excellent" for $r > 0.75$ [33].

Results

Enrollment

Out of the 157 participants who were enrolled in the study (provided consent), 156 completed eligibility screening assessments, 35 subsequently were excluded due to symptoms indicating likely clinical depression as measured by the Older Adult Health and Mood Questionnaire, 6 were excluded due to inconsistent MTAP results based on INFIT or OUTFIT scores above 1.50, and 6 were lost to follow-up. Visit 1 baseline assessments were completed by 117 participants and 110 completed the Visit 2 assessments two to

3 days later, and 99 completed the 30-day follow-up (Visit 3) assessments.

Baseline characteristics of the participants are shown in Tables 1 and 2. Of the enrolled sample, 30.6% of participants had a chief complaint in the thoracic-lumbar spine, 12.1% of participants had a chief complaint in the cervical spine, 30.6% of participants had a chief complaint in the upper extremity, and 26.8% of participants had a chief complaint in the lower extremity.

Temporal Stability

The test–retest reliability of MTAP 2.0 (comparison of Visit 1 to Visit 2) using these data was found to be good to excellent ($r = 0.95$, $p < 0.0001$). The test–retest reliability of VAS was also good to excellent ($r = 0.88$, $p < 0.0001$).

Responsiveness/Sensitivity

Table 3 presents the global score data for each instrument at the Visit 1 baseline, at Visit 2 two days later, and at Visit 3 that occurred 30-days after Visit 1. The effect sizes comparing Visit 1 with Visit 3 for the MTAP, VAS, SF-12 physical component score (PCS), SF-12 mental component score (MCS), and LEFS are of sufficient magnitude to represent a significant difference (all $p < 0.05$). The DASH, NDI, and ODI scores over the 30-day interval did not demonstrate significant differences (all $p > 0.05$).

Concurrent Validity

Pearson correlation coefficients (r) between the MTAP and the other PROs at baseline and 30-day follow-up time points are depicted in Table 4. The baseline relationships between the MTAP and both the Older Adult Health and Mood Questionnaire and current pain intensity measured by the Visual Analog Scale were inverse, in the "fair" range, in the expected direction; higher scores on the MTAP correlated with lower scores on the latter measures. After 30 days, the inverse relationship between the MTAP and VAS improved to the "moderate" range. The relationship between the MTAP and SF-12 PCS remained in the "moderate" range at both baseline and after 30 days. Similarly, the relationship between the MTAP and the SF-12 mental remained in the "fair" range at both baseline and after 30 days. Relationships between the MTAP and the DASH, LEFS, NDI, and ODI were in the expected direction in the "good to excellent" range at both baseline and after 30 days.

Table 1 Baseline demographic characteristics of participants—continuous variables

Variable	Mean	Standard deviation	Range (min–max)
Age (year)	44.8	12.5	19–69
Duration of condition (year)	2.3	4.5	0–47
Body Mass Index (kg/m^2)	29.5	5.8	18.4–46.6
OAHMQ (0–24)	6.3	5.6	0–22

n = 157; *OAHMQ* Older Adult Health and Mood Questionnaire

Table 2 Baseline demographic characteristics of participants—categorical variables

Variable	n	%
Sex		
Female	75	47.8
Male	82	52.2
Primary language		
English	92	58.6
Spanish	65	41.4
Region of chief complaint		
Spine—Cervical	19	12.1
Spine—Thoracic	6	3.8
Spine—Lumbar	42	26.8
Extremity—upper	48	30.6
Extremity—lower	42	26.8
Payer/insurance type		
Medicare	7	4.5
Personal injury	12	7.6
Private/HMO/PPO	30	19.1
Workers' compensation	105	66.9
Other	3	1.9
Work status		
Working at start of treatment	76	43.5
Not working	60	38.2
Not reported	21	13.4

HMO Health Maintenance Organization, *PPO* Preferred Provider Organization

Predictive Validity

Table 5 depicts significant differences at Visit 3 between the mean scores of participants working and not working at the time of telephone follow-up for the DASH ($p = 0.002$), NDI ($p = 0.027$), ODI ($p = 0.018$), SF-12 PCS ($p = 0.007$), VAS ($p = 0.004$), and MTAP ($p = 0.026$). Non-significant differences were found for the LEFS and the SF-12 MCS.

Discussion

This study used a convenience sample of 157 adults to analyze self-report data from three community-based outpatient rehabilitation clinics providing services to patients

Table 3 Global scores for patient reported outcome measures at baseline, 2-day follow-up, and 30-day follow-up time points

Variable	Baseline Mean ± SD n	2-day Mean ± SD n	30-day Mean ± SD n
MTAP (0–200)[a]	127.0 ± 45.9 156	135.7 ± 51.3 110	148.1 ± 45.5 99
Pain intensity (0–10 cm)[b]	4.5 ± 2.4 156	4.0 ± 2.4 110	3.2 ± 2.3 99
SF-12 physical (0–100)[a]	35.7 ± 9.9 117	NA	39.4 ± 10.3 103
SF-12 mental (0–100)[a]	50.5 ± 10.6 117	NA	49.5 ± 11.8 103
DASH (0–100)[b]	37.4 ± 23.2 50	NA	31.3 ± 24.9 47
LEFS (0–80)[a]	37.4 ± 17.0 40	NA	48.5 ± 18.8 35
NDI (0–100)[b]	38.2 ± 16.9 26	NA	34.5 ± 20.9 22
ODI (0–100)[b]	30.3 ± 17.0 46	NA	23.9 ± 19.1 40

SD standard deviation, *NA* not assessed, *MTAP* Multidimensional Task Ability Profile, *Pain intensity* pain intensity assessed on a 10-cm visual analog scale, *SF-12 physical* Short Form 12 physical component score, *SF-12 mental* Short Form 12 mental component score, *DASH* Disabilities of the Arm, Shoulder, and Hand, *LEFS* Lower Extremity Function Scale, *NDI* Neck Disability Index, *ODI* Oswestry Disability Index

[a]Higher score equals better outcome—i.e. higher physical function (MTAP, SF-12 physical, LEFS) or higher mental function (SF-12 mental)

[b]Higher score equals worse outcome, i.e. higher pain intensity (pain intensity) or higher disability (DASH, NDI, ODI)

Table 4 Pearson correlation coefficient (r) between the Multidimensional Task Ability Profile (MTAP) and other patient-reported outcome measures: raw global score at baseline and 30-day follow-up time points

Variable	Baseline Pearson r* n	30 days Pearson r* n
DASH (0–100)	−0.875 50	−0.798 46
LEFS (0–80)	0.735 40	0.707 35
NDI (0–100)	−0.745 26	−0.697 22
ODI (0–100)	−0.776 46	−0.849 38
SF-12 physical (0–100)	0.545 117	0.531 99
OAHMQ (0–24)	−0.465 156	NA NA
Pain intensity (0–10 cm)	−0.333 156	−0.52 99
SF-12 mental (0–100)	0.361 117	0.464 99

NA not assessed, *n* sample size for comparison of MTAP with other patient reported outcome measure, *OAHMQ* Older Adult Health and Mood Questionnaire, *Pain intensity* pain intensity assessed on a 10-cm visual analog scale, *SF-12 physical* Short Form 12 physical component score, *SF-12 mental* Short Form 12 mental component score, *DASH* Disabilities of the Arm, Shoulder, and Hand, *LEFS* Lower Extremity Function Scale, *NDI* Neck Disability Index, *ODI* Oswestry Disability Index

*$p < 0.001$ for all comparisons

with musculoskeletal injuries. The focus of the study was on the concurrent validity of the Multidimensional Task Ability Profile (MTAP) compared with other self-report measures. Selected for comparison with the MTAP were well-established patient-report outcome measures that are used frequently with this patient population—the ODI, NDI, DASH, LEFS, and both the mental and physical components of the SF-12. The design of the study used follow-up measures 30 days after the initial test in order to examine reproducibility of concurrent validity comparisons as well as to measure the effect size for each instrument.

The test–retest reliability with this version of MTAP is consistent with the temporal stability of earlier versions of MTAP [3]. In the current study, both the MTAP and the VAS demonstrated test retest correlations in the "good to excellent" range. These temporal stability findings are consistent with prior studies of each of the other self-report measures examined in this study [11–15].

The responsiveness and sensitivity to change of the instruments in the study over the 30-day interval of treatment was established for the MTAP, VAS, SF-12 PCS, SF-12 MCS, and LEFS, with each instrument providing a score that indicated improvement. The DASH, NDI, and ODI were not found to be sensitive to change over the same interval. For the NDI, this may have been due to the relatively small sample size (n = 22), but the DASH and ODI sample sizes (n = 47 and 40, respectively) were greater than the sample size of the LEFS (n = 35), which was sensitive to change over this interval.

Concurrent validity comparisons of the MTAP and each of the other self-report measures were fair to excellent and in the expected directions at each time of testing. Convergent validity was demonstrated with the high correlations between the MTAP and the physical-task-focused measures. Discriminant validity can be inferred from the relative differences among the correlations between the MTAP and the other measures; those that are physical-task-focused have much higher correlations than those that focus on other constructs, such as mood (OAMHQ) or pain (Pain VAS) or mental function (SF-12). This suggests

Table 5 Post-treatment work status prediction based on self-report at Visit 3

Measure	Working?	n	Mean	SD	SS	df	MS	F	Sig.
DASH	Yes	25	20.22	21.18	5504.32	1	5504.32	10.53	0.002*
	No	18	43.15	25.05	21434.06	41	522.78		
	Total	43	29.82	25.33	26938.38	42			
LEFS	Yes	13	51.77	20.39	183.40	1	183.40	0.50	0.485
	No	19	46.89	18.31	11024.10	30	367.47		
	Total	32	48.88	19.01	11207.50	31			
NDI	Yes	10	24.20	15.04	2129.66	1	2129.66	5.78	0.027*
	No	11	44.36	22.28	7000.15	19	368.43		
	Total	21	34.76	21.37	9129.81	20			
ODI	Yes	18	17.11	15.48	2040.03	1	2040.03	6.17	0.018*
	No	18	32.17	20.53	11238.28	34	330.54		
	Total	36	24.64	19.48	13278.31	35			
SF-12 physical	Yes	48	42.18	10.18	799.74	1	799.74	7.70	0.007*
	No	45	36.31	10.20	9449.52	91	103.84		
	Total	93	39.34	10.55	10249.26	92			
SF-12 mental	Yes	48	51.64	10.43	126.41	1	126.41	1.00	0.321
	No	45	49.31	12.08	11538.49	91	126.80		
	Total	93	50.51	11.26	11664.90	92			
VAS	Yes	48	2.45	2.33	46.00	1	46.00	8.89	0.004*
	No	42	3.88	2.21	455.52	88	5.18		
	Total	90	3.12	2.37	501.53	89			
MTAP	Yes	48	159.10	39.03	10089.52	1	10089.52	5.14	0.026*
	No	42	137.88	49.64	172646.88	88	1961.90		
	Total	90	149.20	45.31	182736.40	89			

*$p < 0.05$ for all comparisons

that the measures can provide supplemental information that is likely to be consistent. This may be useful, based on other clinical and administrative issues. For example, in a clinic that provides services for patients with a wide range of impairment types, the MTAP may be a useful omnibus measure to supplement specialty instruments such as the DASH, NDI, or LEFS. In clinics that provide specialty care, such as a focus on upper extremity injuries, the DASH would provide more focus on specific tasks that are pertinent. Adding the MTAP provides a potentially valuable cross-reference to the Physical Demand Characteristics rating system for occupations in the Dictionary of Occupational Titles [23] on which the MTAP has been cross-validated [3].

Predictive validity for return to work was examined for each of the instruments based on a telephone follow-up. A fair to moderate level of predictive validity with regard to post-treatment work status was established with all of the measures, except the LEFS and the SF-12 MCS. The former finding was not unexpected because there are no published studies that confirm the predictive validity of the LEFS with regard to return to work. However, the SF-12 MCS has demonstrated good predictive validity for return to work after treatment for disabling low back pain [34]. It may be that the current study's exclusion of participants based on elevated scores on the OADHQ, indicating more depressive symptoms diminished the variance of the SF-12 MCS, thus limiting its ability to identify participants who would be less likely to return to work.

Although the MTAP was designed to correspond to the physical demands scale in the Dictionary of Occupational Titles [19], the significant predictive validity of the MTAP for post-treatment work status is a new finding has not been studied previously. The positive findings for predicting post-treatment work status by the other self-report measures were not unexpected. The DASH [35], NDI [36], ODI [36–38], SF12 PCS [39], and VAS [39, 40] all have had at least one study that has demonstrated significant predictive validity for return to work.

This study leads to recommendations for use of the MTAP in clinical practice. The MTAP may also be used to help guide treatment approaches and the intensity of the plan of care, aimed at improving general physical function. The MTAP is a comprehensive PRO that provides a large amount of information about patient function, health status and work capacity. This information, combined with the patient's history, diagnoses and physical findings, enhances communications, and helps direct plan of care and treatment algorithms.

Clinicians could address deficits in specific activities of daily living and work that are uncovered in the MTAP items, particularly during the shared decision-making process and to identify patient preferences. Furthermore, since the MTAP is linked to metabolic equivalents (METS) of common activities of daily living, it can be used to help guide the intensity and volume of a plan of care.

One aspect of MTAP that is also available in measures that are developed with the item response theory approach, but not in the other patient reported outcome measures used in the current study is the availability of two scores indicating inconsistency of responding. When either the INFIT or the OUTFIT indices exceed a value of 1.50, clinicians should inquire about the reasons for inconsistency, such as comorbid conditions that have not yet been identified or problems with motivation leading to less than full effort responding. Another important use of the MTAP can be inferred from the design of the study; excluding potential participants who were illiterate in both Spanish and English. With the exception of the pain VAS, all of the other patient reported outcome measures are solely text-based, thereby requiring literacy. Combination of pictures and text allows the MTAP to be used with illiterate patients.

Limitations of study

The study enrolled a sample of convenience in a pragmatic clinical setting across specific time periods. Thus, at baseline, patients were enrolled at various stages of care and the intervention period was only 30 days. Thus, all patients were not new patients in this clinical setting, and dose-responsiveness, sensitivity, and effect of treatment could not be assessed.

This was a pragmatic pilot study with a non-experimental, 1-arm observational cohort design. Thus, it was not adequately powered for complete analyses. For some of the musculoskeletal impairment groups, the sample size was insufficient.

Exclusion of participants based on reports of clinically-significant depressive symptoms was a reasonable feature of the current study's design, but the high number of participants excluded on this basis was unexpected. The exclusion of such a large proportion of participants was unlikely to not affect some of the findings, especially those that were related to mental status. This is certainly likely with the SF-12 MCS, which was found not to have predictive validity for return to work.

Exclusion of participants based on inconsistent MTAP findings may have provided bias in this study. Although only 6 of 157 participants were excluded on this basis, this advantage was not provided to any of the other self-report measures. It is reasonable that participants with inconsistent MTAP findings were either less cooperative or less cognitively capable of participating, leading to diminished likelihood of success in rehabilitation.

Work status questions were administered up to several months following the 30-day follow-up visit, thus imposing a recall bias that may have differentially affected the predictive validity of the instruments. Work status was determined at one time point via one phone call and, therefore, changing work situations were unable to be considered. Twenty-one participants were unable to be contacted via the follow-up phone call. Therefore, work status was missing from these 21 participants and these data were not included in the predictive validity calculations. Also, the Physical Demand Characteristics levels of the jobs [23] to which patients returned were not collected during the telephone interview, which missed an opportunity to further examine predictive validity because the MTAP items are cross-calibrated on this system [1, 2].

Future Research

In addition to securing impairment subgroup sample sizes that provide adequate statistical power to examine the sensitivity of the impairment-specific patient reported outcome measures, future research should collect baseline data at the onset of treatment, at a set time during treatment, and at the conclusion of treatment, along with follow-up at a set interval after treatment for each participant.

A measure of return to work status that includes the physical demand level of each participant's work that has been validated in a telephone follow-up paradigm should be used so that the long-term accuracy of the match between the self-report ability measures at conclusion of treatment can be studied. In addition, immediate, intermediate, and long-term follow-up should be used to examine the efficacy of rehabilitation treatment. In this population of people with musculoskeletal impairments, re-injury at work is a significant problem that may or may not be able to be addressed in rehabilitation. Unless the data are obtained to examine this circumstance, the ability of clinical professionals to improve services will be indeterminant.

Research that examines the predictive ability of the MTAP consistency measures should be undertaken. Concerns about full-effort participation are endemic in this aspect of rehabilitation, especially when a patient may have a possibility of secondary gain related to a poor rehabilitation outcome. As one behavioral indicator of noncooperation, elevated scores on the MTAP INFIT and OUTFIT scales should be examined for predictive validity of poor outcome.

Conclusions

This study examined the temporal stability, responsiveness and sensitivity to change, concurrent validity, and return to work predictive validity of the MTAP in comparison with well-established self-report measures in a musculoskeletal rehabilitation outpatient treatment program. Short-term temporal stability of the MTAP was demonstrated, along with significant responsiveness and sensitivity to change, as well as significant convergent and discriminant validity on a concurrent basis. Predictive validity for post-treatment work status of the MTAP and the DASH, NDI, ODI, SF-12 PCS and the VAS was also demonstrated.

Acknowledgements The authors thank the staff of Spine & Sport Inc. for assistance with coordination and data collection for this study. The authors also thank Dr. Bryan Kemp for assistance with the Older Adult Health Questionnaire and for serving as chair of the Foundation's Institutional Review Board.

Compliance with Ethical Standards

Conflicts of interest This study was funded by the Vert Mooney Research Foundation, a 501c(3) charitable organization that developed and currently owns the Multidimensional Task Ability Profile. Leonard Matheson, John Mayer, and Joe Verna are board members of the Vert Mooney Research Foundation. Leonard Matheson and Joe Verna own the intellectual property for the MTAP and benefit financially from MTAP sales and subscriptions. John Mayer received financial compensation for contributing to this project.

Ethical Approval All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000 [41].

Informed Consent Informed consent was obtained from all patients for being included in the study.

References

1. Rose M, et al. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). J Clin Epidemiol. 2008;61(1):17–33.
2. Matheson LN, Kaskutas VK, Mada D. Development and construct validation of the Hand Function Sort. J Occup Rehabil. 2001;11(2):75–86.
3. Mayer J, et al. The reliability and validity of a new computerized pictorial activity and task sort. J Occup Rehabil. 2005;15(2):185–195.
4. Kramer A, Holthaus D. Uniform patient assessment for post-acute care. Final report. Aurora, CO: Division of Health Care Policy and Research, University of Colorado at Denver and Health Sciences Center; 2006.
5. American Medical Association. Guides to the evaluation of permanent impairment. 6th ed. Chicago, IL: American Medical Association; 2008. p. 613.
6. Khorsan R, et al. Measures in chiropractic research: choosing patient-based outcome assessments. J Manipulative Physiol Ther. 2008;31(5):355–375.
7. Author. Official disability guidelines. Austin, TX: MCG Health; 2019.
8. Cella D, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. J Clin Epidemiol. 2010;63(11):1179–1194.
9. Centers for Medicare & Medicaid Services (CMS), HHS. Medicare Program; Merit-Based Incentive Payment System (MIPS) and Alternative Payment Model (APM) incentive under the physician fee schedule, and criteria for physician-focused payment models. Final rule with comment period. Fed Reg. 2016;81(214):77008–77831.
10. Quality ID #182 (NQF 2624): Functional Outcome Assessment 2019. https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/ClinicalQualityMeasures.html. Accessed 20 Dec 2018.
11. Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine. 2000;25(22):2940–2952.
12. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. J Manipulative Physiol Ther. 1991;14(7):409–415.
13. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH. The Upper Extremity Collaborative Group (UECG). Am J Ind Med. 1996;29(6):602–608.
14. Binkley JM, et al. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. Phys Therapy. 1999;79(4):371–383.
15. Ware JJ, Kosinski M, Keller S. A 12-item Short-Form Health Survey, construction of scales and preliminary tests of reliability and validity. Med Care. 1996;34(3):220–233.
16. Matheson L. History, design characteristics, and uses of the pictorial activity and task sorts. J Occup Rehabil. 2004;14(3):175–195.
17. Matheson L, Matheson M, Grant J. Development of a measure of perceived functional ability. J Occup Rehabil. 1993;3(1):15–30.
18. Verna JL, et al. Development and reliability testing of Spanish language and English language versions of the multidimensional task ability profile. J Occup Rehabil. 2013;23(2):220–227.
19. Matheson L, et al. A method to provide a more efficient and reliable measure of self-report physical work capacity for patients with spinal pain. J Occup Rehabil. 2008;18(1):46–57.
20. Mooney V, et al. Performance-integrated self-report measurement of physical ability. Spine J. 2010;10(1):433–440.
21. Andrich D. Rasch models of measurement. Quantitative applications in the social sciences. Newbury Park: Sage; 1988.
22. Rasch G. Probabilistic models for some intelligence and the attainment tests. Copenhagen: Danmarks Paedagogiske Institute; 1960.
23. U.S. Department of Labor. Dictionary of Occupational Titles, vol. 1. 4th ed. Washington, DC: United States Government Printing Office; 1991.
24. Kemp B, Adams B. The older adult health and mood questionnaire: a measure of geriatric depressive disorder. J Geriatr Psychiatry Neurol. 1995;8(July):162–167.
25. Matheson L. Multidimensional Task Ability Profile professional manual. St. Charles, MO: EpicRehab LLC.; 2011.
26. Bergbom S, et al. Relationship among pain catastrophizing, depressed mood, and outcomes across physical therapy treatments. Phys Therapy. 2011;91(5):754–764.
27. Moreno R, et al. Functional restoration for chronic low back pain: changes in depression, cognitive distortion, and disability. J Occup Rehabil. 1991;1(3):207–216.
28. Gatchel RJ, Mayer TG, Theodore BR. The pain disability questionnaire: relationship to one-year functional and psychosocial rehabilitation outcomes. J Occup Rehabil. 2006;16(1):72–91.

29. Adams H, et al. The relation between catastrophizing and occupational disability in individuals with major depression: concurrent and prospective associations. J Occup Rehabil. 2017;27(3):405–412.
30. Andrich D. A rating formulation for ordered response categories. Psychometrika. 1978;43(4):357–374.
31. Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. Pain. 1986;27(1):117–126.
32. Luo X, et al. Reliability, validity, and responsiveness of the short form 12-item survey (SF-12) in patients with back pain. Spine. 2003;28(15):1739–1745.
33. Portney L, Watkins M. Foundations of clinical research: applications to practice. 2nd ed. Upper Saddle River: Prentice-Hall, Inc; 2000. p. 742.
34. Baldwin ML, et al. Self-reported severity measures as predictors of return-to-work outcomes in occupational back pain. J Occup Rehabil. 2007;17(4):683–700.
35. Armijo-Olivo S, et al. Predictive value of the DASH tool for predicting return to work of injured workers with musculoskeletal disorders of the upper extremity. Occup Environ Med. 2016;73(12):807–815.
36. Marchand GH, et al. Change in pain, disability and influence of fear-avoidance in a work-focused intervention on neck and back pain: a randomized controlled trial. BMC Musculoskelet Disord. 2015;16(1):94–104.
37. McGirt MJ, et al. Prediction model for outcome after low-back surgery: individualized likelihood of complication, hospital readmission, return to work, and 12-month improvement in functional disability. Neurosurg Focus. 2015;39(6):E13.
38. Wind H, et al. Assessment of functional capacity of the musculoskeletal system in the context of work, daily living, and sport: a systematic review. J Occup Rehabil. 2005;15(2):253–272.
39. Gopinath B, et al. Prognostic indicators of social outcomes in persons who sustained an injury in a road traffic crash. Injury. 2015;46(5):909–917.
40. Tschernetzki-Neilson PJ, et al. Changing to an outcome-focused program improves return to work outcomes. J Occup Rehabil. 2007;17(3):473–486.
41. World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research involving human subjects. JAMA. 2013;310(20):2191–2194.