ORIGINAL REPORT

CROSS-REGIONAL VALIDITY OF THE ASSESSMENT OF MOTOR AND PROCESS SKILLS FOR USE IN MIDDLE EUROPE

Brigitte E. Gantschnig, MScOT[1], Julie Page, PhD[1] and Anne G. Fisher, ScD[2]

From the [1]Zurich University of Applied Sciences, School of Health Professions, Institute of Occupational Therapy, Zurich, Switzerland and [2]Umeå University, Faculty of Medicine, Department of Community Medicine and Rehabilitation, Division of Occupational Therapy, Umeå, Sweden

Objective: To evaluate cross-regional validity of the Assessment of Motor and Process Skills (AMPS) with a specific focus on valid use with Middle Europeans.

Design: Descriptive cross-regional validation study.

Participants: A total of 1346 participants from Middle Europe and 144,143 participants from North America, UK/Ireland, the Nordic Countries, other Europe, Australia/New Zealand and Asia, between the ages of 3 and 103 years, in good health and with a variety of diagnoses, were selected from the AMPS database.

Methods: Many-facet Rasch analysis was used to analyse participant raw data, and effect sizes were used to evaluate for differential item functioning. Evaluation for differential test functioning was also implemented.

Results: None of the 20 activity of daily living process items, and only one of the activity of daily living motor items demonstrated differential item functioning. The activity of daily living motor item *Aligns* exceeded the significant effect size criterion of ±0.55 logit, but the significant differential item functioning did not lead to differential test functioning (i.e. all measures fell within the 95% confidence bands).

Conclusion: This study provides further evidence of validity of the AMPS when used to evaluate quality of activity of daily living tasks performance across world regions. The AMPS measures can be used as objective indices of activity of daily living ability in rehabilitation settings and in international collaborative research related to activity of daily living task performance.

Key words: activities of daily living; rehabilitation; differential item functioning; differential test functioning; occupational therapy; Rasch analysis.

INTRODUCTION

Rehabilitation professionals aim to optimize activity, social participation and quality of life of clients with acute and/or chronic health conditions (1–3). In order to evaluate the effectiveness of rehabilitation and to meet the increasing need for evidence-based practice, it is important to use reliable, valid and sensitive outcome measures (1–5). Such measures should also be compatible with the International Classification of Functioning, Disability and Health (ICF) (6). The components of activity and participation in the ICF define aspects of functioning and disability (6) that are the main focus of occupational therapy (7). Thus, as part of the multi-professional rehabilitation team, occupational therapists enable and evaluate clients' abilities to perform activity of daily living (ADL) tasks (1, 2, 7).

While contemporary practice demands the use of standardized outcome measures (5), there is a critical lack of ADL instruments that have been validated for use in Middle European countries (8). This situation has occurred for several reasons. First, while Middle European countries share common values, beliefs and an evocative history, they also use a variety of languages and English is not ranked among them (9). The result is that assessments that have been developed in Anglo-American countries are often not available in languages spoken in Middle Europe, nor are they validated for use in this region. Secondly, only a few assessments have been developed within Middle Europe and many of them were developed without establishing their validity and reliability (10, 11). Thirdly, the majority of assessments used in Middle European rehabilitation settings were designed to evaluate body functions, not activity or participation (11, 12). Thus, there is a need for activity- or participation-based outcome measures that are validated for use in Middle European countries.

The specific focus of this study was the Assessment of Motor and Process Skills (AMPS) (13, 14). The AMPS was chosen because it is an internationally standardized observational, performance-based assessment designed to be used by occupational therapists to measure the quality of a person's performance of ADL tasks in naturalistic settings. Thus, the AMPS is activity-based. The AMPS items have been linked to, and shown to be compatible with, the concepts of activity and participation in the ICF (7). Currently, there are 116 standardized ADL tasks in the AMPS (14) that are hierarchically ordered according to their task challenge (see Fig. 1). While some are generally viewed as being world-region-specific (e.g. *peanut butter and jelly sandwich* for use in North America,

eating an Asian meal with chopsticks for use in Asian countries), the majority of the AMPS tasks are among the most commonly performed personal and instrumental ADL tasks internationally (e.g. *upper and lower body dressing, cleaning a bathroom*). Standard administration procedures specify that the person is to choose 2 ADL tasks from among a subset of the 116 ADL tasks included in the AMPS manual based on the following criteria: (*i*) they are meaningful for and relevant to the person's daily life; (*ii*) they are ADL tasks that currently are presenting a challenge; and (*iii*) the person has prioritized them for further assessment and intervention (13, 14). No matter which 2 tasks the person is observed performing, they are scored on the same 16 ADL motor and 20 ADL process items, once for each task performed (see Fig. 1). *"The ADL motor [items] are occupational performance skills observed as the person interacts with and moves task objects, and moves oneself around the task environment. The ADL process [items] are occupational performance skills observed as a person selects, interact with and uses task tools and material, carries out individual actions and steps and modifies performance when problems are encountered"* (15). Thus, each ADL task can be thought of as a unique form or version of the AMPS, whereas the ADL motor and ADL process items are common to each form. For example, if a person is observed vacuuming, he or she is scored on how effective he or she was as he or she *lifted* the vacuum and/or lightweight furniture, and *moved* the vacuum back and forth across the floor. If the person is then observed showering, he or she is scored, for example, on how effective he or she was as he or she *lifted* the towel or other task objects, and *moved* the towel to dry his or her body. As the ADL tasks become more difficult, each of the ADL items (e.g. *Lifts* and *Moves*) also become proportionally more difficult. In other words, the AMPS can be thought of as being comprised of two item banks, one for ADL motor scale and one for ADL process scale of the AMPS. Within the current item bank, each ADL item is represented 116 times, once for each ADL task included in the AMPS manual, and each person is scored on 32 ADL motor items and 40 ADL process items (14–16).

Because the AMPS ADL items have been developed to be universal (i.e. observable during any ADL task), and the person performs only those tasks that are familiar, culturally-relevant and chosen, the AMPS measures should remain free of cross-regional bias when used in Middle Europe. This assertion has been supported by studies evaluating for the item difficulty stability among North America, Scandinavia and the UK (17) and among 6 world regions, where Middle Europe was combined with all other countries in continental Europe to form a single region: other Europe (18). The AMPS has also been shown to be free of cross-cultural bias between Black and White Americans (19), between Cuban Americans and European Americans (20) and among Mexican Americans (21).

The present study was based on the premise that a cost-effective method for developing standardized tools for use in a specific world region is to adapt, as needed, and validate existing tools. Rasch analysis methods are a family of methodological approaches that are commonly used in rehabilitation, not only for developing new measures, but also for validating existing ones. Specific to Rasch analysis is that the person's raw item scores are converted into a linear measure, expressed in logit (log-odds probability units) (22). The specific Rasch model used to develop the AMPS was a many-faceted Rasch (MFR) model, where person ability, rater severity, task challenge and item difficulty are each calibrated along a common logit scale. The MFR model of the AMPS has been described in more detail elsewhere (14, 16).

Rasch analysis methods can also be used to evaluate whether the items in an instrument behave in the same way across groups of persons from different world regions (i.e. if item difficulties remain stable across world regions, without evidence of differential item functioning (DIF)). By comparing the item difficulties based on a Middle European sample to the item difficulties based on samples from other world regions, the presence or absence of DIF in Middle Europe can be detected. If an assessment is free of DIF among world regions, the item hierarchies will be the same (i.e. stable, independent of region where the instrument is used). In contrast, if differences in item difficulty calibrations between regions arise, for example, if an item is calibrated as easier for persons in one group than for persons in other groups, DIF is detected. DIF can mean that persons tested from one world region may be at an unfair disadvantage compared with persons from other world regions. Therefore, it is necessary to determine whether DIF leads to differential test functioning (DTF) (23, 24). DTF is commonly analysed by plotting measures for all persons based on the item difficulty calibrations from one group against measures for all persons based on the item difficulty calibrations from another group. The location of the paired measures should fall within 95% confidence bands based on the standard errors (SEs) of the estimated person measures. If more than 5% of the paired measures are located outside the 95% confidence bands, DTF is detected, signalling that test bias is present (22, 25).

Continental Europe is comprised of 4 sub-regions (i.e. southern, central or middle, western and eastern). The aim of this study was to evaluate for cross-regional DIF of the AMPS items when the AMPS is used in Middle Europe. A secondary aim was to ensure that any detected DIF does not lead to DTF when the AMPS is used to evaluate Middle Europeans. More specifically, our research questions were: (*i*) Do MFR DIF analyses of the ADL motor and ADL process items of the AMPS reveal significant differences in item difficulty calibration values between (*a*) Middle Europe, and (*b*) North America, UK/Ireland, the Nordic Countries, other Europe (continental Europe, not including Middle Europe), Australia/New Zealand or Asia? (*ii*) If DIF is detected, does it impact the final ADL ability measures of the AMPS (i.e. is there test bias in the form of DTF)?

METHODS

Participants

The participants in this study included all available persons from the international AMPS database, Ft Collins, Colorado, USA as of June

Easier ADL tasks	Easier ADL motor items	Easier ADL process items
Eating a snack with a utensil	Lifts	Uses
Brushing teeth	Moves	Sequences
Folding a basket of laundry	Transports	Searches/locates
Feeding a cat: dry cat food and water	Grips	Gathers
Showering	Bends	Terminates
Hot cereal & beverage	Coordinates	Restores
Vacuuming the inside of an automobile	Calibrates	Notices/responds
Vegetable soup, vegetables sautéed	Positions	Accommodates
Harder ADL tasks	**Harder ADL motor items**	**Harder ADL process items**

Fig. 1. Selected activities of daily living (ADL) tasks and ADL motor and ADL process items included in the Assessment of Motor and Process Skills (AMPS); adapted from Fisher & Griswold (15). 1) Each person is to choose 2 ADL tasks from the 116 ADL tasks included in the AMPS that are meaningful, are presenting a challenge and are prioritized for intervention. 2) The 16 ADL motor and 20 ADL process items are scored for each of the 2 selected ADL tasks, based on the person's observed quality of ADL task performance (degree of clumsiness or physical effort, efficiency, safety and/or need for assistance).

2010 who had been scored by raters in a valid manner (i.e. free of rater scoring error). For this study 145,489 persons, 3 years of age and above, were selected, of whom 1,346 were from Middle Europe, both healthy persons and those with a broad variety of ages and diagnoses (e.g. orthopaedic/musculoskeletal, neurological) so as to reflect the variety of persons evaluated by occupational therapists in Middle Europe; and 144,143 were persons from North America, UK/Ireland, the Nordic Countries, other Europe, Australia/New Zealand and Asia. Demographic characteristics of the participants are presented in Tables

I and II. Data from approximately 5% of the total AMPS database were excluded because of invalid data due to rater scoring error.

The data for the Middle European sample was submitted to the AMPS database by 117 occupational therapists who had attended AMPS courses in Austria, Germany, Slovenia and Switzerland, and who calibrated as reliable and valid AMPS raters. To become calibrated as a valid and reliable rater, each rater participates in a 5-day training course, during which he or she co-scores a minimum of 8 videotaped and live calibration cases, and then independently scores an additional

Table I. *Gender and diagnostic characteristics of participants as a proportion of the sample*

	Region							
	NA	UK/Ireland	Nordic	OEurop	ANZ	Asia	ME	Total
Characteristic	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
Gender								
Male	42.1 (9,866)	46.1 (13,674)	42.7 (20,246)	42.0 (6,317)	47.1 (6,278)	45.3 (6,879)	44.6 (600)	43.9 (63,860)
Female	57.8 (13,559)	53.9 (15,985)	57.3 (27,182)	58.0 (8,737)	52.9 (7,046)	54.7 (8,321)	55.4 (746)	56.1 (81,576)
Unknown	0.1 (16)	0.0 (11)	0.0 (0)	0.0 (3)	0.0 (5)	0.0 (1)	0.0 (0)	0.0 (53)
Diagnoses								
Well	15.1 (3,541)	7.9 (2,330)	7.1 (3,387)	4.6 (691)	7.7 (1,025)	10.4 (1,579)	4.8 (65)	8.7 (12,618)
OldRiskFrail	0.8 (196)	0.8 (251)	0.5 (250)	0.4 (55)	0.8 (100)	1.4 (212)	0.0 (0)	0.7 (1,064)
Mild	0.9 (208)	0.6 (192)	1.0 (475)	0.7 (104)	0.5 (72)	1.1 (162)	6.8 (91)	0.9 (1,304)
DevelNeur	1.7 (387)	2.4 (716)	2.5 (1,182)	2.4 (364)	2.3 (303)	5.8 (875)	2.1 (28)	2.6 (3,855)
OtherNeur	11.3 (2,649)	9.4 (2,797)	16.1 (7,634)	15.8 (2,385)	11.7 (1,562)	11.6 (1,765)	23.8 (321)	13.1 (19,113)
MR	1.2 (285)	5.7 (1,693)	1.3 (627)	0.7 (98)	1.1 (146)	2.6 (392)	0.5 (7)	2.2 (3,248)
CVA	6.0 (1,407)	6.0 (1,767)	16.8 (7,976)	20.7 (3,121)	6.1 (814)	22.4 (3,403)	18.5 (249)	12.9 (18,737)
Mskl	10.3 (2,414)	6.3 (1,861)	12.5 (5,934)	10.7 (1,604)	8.9 (1,190)	11.0 (1,677)	8.8 (118)	10.2 (14,798)
MedSens	5.8 (1,370)	3.7 (1,102)	3.0 (1,429)	2.7 (409)	4.1 (547)	2.3 (345)	1.3 (17)	3.6 (5,219)
OtherPsych	6.1 (1,419)	9.5 (2,825)	6.1 (2,880)	2.7 (411)	10.4 (1,391)	2.1 (323)	2.2 (29)	6.4 (9,278)
MultUnknown	31.3 (7,344)	32.7 (9,705)	25.0 (11,838)	34.7 (5,222)	31.0 (4,138)	18.1 (2,752)	28.4 (382)	28.4 (41,381)
OtherMem	0.4 (102)	0.6 (183)	0.3 (163)	0.1 (21)	0.5 (60)	0.3 (44)	0.1 (1)	0.4 (574)
Dem	4.5 (1,055)	5.4 (1,600)	3.4 (1,598)	2.3 (340)	2.1 (283)	2.9 (448)	0.9 (12)	3.7 (5,336)
SchizThought	4.5 (1,064)	8.9 (2,648)	4.4 (2,072)	1.5 (232)	12.7 (1,698)	8.1 (1,224)	1.9 (26)	6.2 (8,964)
Total	100.0 (23,441)	100.0 (29,670)	100.0 (47,445)	100.0 (15,057)	100.0 (13,329)	100.0 (15,201)	100.0 (1,346)	100.0 (145,489)

NA: North America; UK: UK and Republic of Ireland; Nordic: Nordic Countries; OEurop: Western, Southern, Eastern Europe; ANZ: Australia and New Zealand; Asia: Asia; ME: Middle Europe. Well: well persons; OldRiskFrail: older adults 60 years of age and older, who are frail or at risk for functional decline, but without known medical problems; Mild: persons at risk for or who have been diagnosed with mild disability; DevelNeur: persons with neurological developmental disorders; OtherNeur: persons with other types of neurological disorders, e.g. traumatic brain injury; MS; MR: persons with mental retardation; CVA: Persons with right- or left-sided cerebral vascular accident; Mskl: persons with musculoskeletal disorder; MedSens: persons with medical conditions (e.g. cardiovascular, respiratory, burns, AIDS/HIV) or sensory disorders (e.g. visual, auditory, vestibular); Other Psych: persons with other psychiatric disorders or disorder on the autism spectrum; MultUnknown: persons with two or more diagnoses from different categories or whose diagnoses were unknown; OtherMem: persons with memory disorders not associated with dementia; Dem: persons with dementia; SchizThought: persons with schizophrenia or other type of thought disorder.

Table II. *Demographic and activities of daily living (ADL) motor and ADL process item abilities (in logits) of participants per world region*

Characteristic	Region							
	NA Mean (SD) [range]	UK/Ireland Mean (SD) [range]	Nordic Mean (SD) [range]	OEurop Mean (SD) [range]	ANZ Mean (SD) [range]	Asia Mean (SD) [range]	ME Mean (SD) [range]	Total Mean (SD) [range]
Age, years	53.28 (25.82) [3−100]	53.66 (24.01) [3−100]	55.02 (23.72) [3−103]	57.80 (22.93) [3−100]	50.15 (24.48) [3−100]	53.12 (25.86) [3−103]	50.27 (24.66) [3−96]	54.06 (24.43) [3−103]
ADL motor ability	1.21 (0.96) [−3.00−3.90]	1.21 (0.96) [−2.90−3.88]	1.14 (0.93) [−2.96−4.06]	0.97 (0.91) [−2.81−3.90]	1.22 (0.96) [−2.96−3.82]	0.90 (0.95) [−2.89−3.83]	0.84 (0.89) [−2.60−3.57]	1.13 (0.95) [−3.00−4.06]
ADL process ability	0.78 (0.72) [−2.00−3.01]	0.69 (0.70) [−2.02−2.88]	0.84 (0.69) [−2.01−3.02]	0.69 (0.68) [−1.97−2.97]	0.75 (0.68) [−2.00−2.81]	0.70 (0.66) [−2.01−2.84]	0.59 (0.64) [−1.92−2.55]	0.76 (0.69) [−2.02−3.02]

NA: North America; UK: UK and Republic of Ireland; Nordic: Nordic Countries; OEurop: Western, Southern, Eastern Europe; ANZ: Australia and New Zealand; Asia: Asia; ME: Middle Europe.

10 persons (live) after the course. All data, co-scored and independent, are then subjected to MFR analyses and the results are evaluated in terms of overall rater severity and goodness of fit of the rater to the MFR model of the AMPS. Each rater's person data are also subjected to a detailed analysis to determine if the person ADL motor and ADL process ability measures are valid. Consistent with all persons whose data are included in the AMPS database, the persons from Middle Europe had been evaluated using the AMPS in naturalistic settings, both clinical- and community-based (e.g. fully equipped kitchens).

Sample size selection for this study was based on the premise that a minimum of 200 persons are required in each of the regions (i.e. Middle Europe and each of the other world regions), but that it is desirable to use the largest possible sample sizes when performing DIF and DTF analyses (26). For the purpose of this study, we included data for persons from Austria, Germany, Slovenia, Lichtenstein and Switzerland in the Middle European sample as AMPS data were only available for those Middle European countries. We are aware, however, that Middle Europe can be considered to extend beyond the borders of those 5 countries.

Administration and scoring procedures for the Assessment of Motor and Process Skills

The AMPS was administered to all participants included in this study by AMPS trained occupational therapists (raters) who scored 16 ADL motor and 20 ADL process items, for each of two different ADL tasks performed by each person, based on the observed quality of person's performance of each ADL task (see Fig. 1). The ADL motor and ADL process items of the AMPS comprise goal-directed actions carried out when performing a personal or instrumental ADL task (e.g. *Reaching* for, *Grasping, Choosing* and *Lifting* a glass; and then *Initiating* filling the glass with water). When these actions are linked together, they result in a chain of actions that are the observed ADL task performance (7, 14, 27). The ADL motor and ADL process items (i.e. the smallest observable actions of occupational performance, performance skills) are each rated in terms of any observed increase in physical effort or clumsiness, decrease in efficiency, and decrease in safety and/or frequency of assistance provided in relation to that action. It is important to stress that the ADL motor and ADL process items represent the smallest observable units of occupational performance, not underlying body functions, e.g. musculoskeletal, neurological, cognitive (7, 13, 14, 27). Afterwards, the raw item scores for each observed ADL task are entered into the rater's personal copy of the AMPS computer-scoring software (AMPS, Fort Collins, USA) (28), which is used to (*i*) convert the person's raw scores into linear ADL motor and ADL process ability measures expressed in logits, taking into account the rater's severity, the challenge of the two tasks performed and the difficulties of the ADL motor and ADL process items, and (*ii*) generate AMPS graphic and summary reports (14). Both English and German versions of the manual were used to test persons from Middle Europe (13, 14, 29). The AMPS, however, is an observational, performance-based tool, and in all cases, the AMPS was administered using the German or Slovenian language. Only the AMPS rater reads

the AMPS manual, and as long as the rater is fluent in English, there should be no impact on the results.

Data analysis

This study was a descriptive cross-regional validation study. When the data were analysed, all AMPS item difficulty calibration values were generated using FACETS, an MFR computer software program, (FACETS, Chicago, USA) described elsewhere in more detail (16, 30). Two analyses were performed (one for the ADL motor and one for the ADL process items), and in each of these analyses, the task challenges and rater severities were anchored at pre-established values based on the current AMPS computer-scoring program (28). Within FACETS, it is possible to request DIF analyses which enable comparison of the item difficulty calibrations for each world region directly by calculating the logit differences between world regions. In total, we compared Middle Europe with 6 other world regions as well as with the total sample, resulting in 7 comparison pairs.

Because of the risk that our large sample sizes would result in too much power and over-identifying significant differences based on p-values alone (31), we used effect sizes to evaluate for significant DIF (31, 32). More specifically, we set our criteria for the presence of significant DIF based on a logit difference of at least ± 0 .55 logit between Middle Europe and each of the other 6 world regions and the total combined sample. This criterion was based on Tristán (32), who found that when standard errors (SEs) are normalized, the minimum possible SE is 0.20 logit. With SE values of 0.20 logit, a difference in item difficulty calibration values of ±0.55 logit is required for statistical significance. The rationale for the use of Tristan's criterion is discussed in more detail by Munkholm et al. (33), who implemented a similar cross-regional validation study of the School Version of the AMPS.

In the second phase, DTF was evaluated by plotting the ADL measures for all persons when estimated based on the item difficulty calibrations for Middle Europe against the ADL measures for all persons when based on the item difficulty calibrations for each of the other world regions or the total combined sample. We then evaluated whether the paired ADL measures fell within the 95% confidence bands, indicating no evidence of DTF (25, 33). More specifically, two different sets of item difficulty calibration values are expected to yield invariant person measures, such that the plotted measures fall within the 95% confidence bands that are based on the SEs for each item pair (22, 25).

The study was ethically approved by the Regional Ethical Review Board, Faculty of Medicine, Umeå University, Sweden (Dnr03-509). Furthermore, the Ethics Committee of Canton Zurich confirmed that the secondary analysis of anonymous medical data does not need to be submitted to the Ethics Committee in Switzerland.

RESULTS

Comparison of each ADL motor and ADL process item difficulty calibration value for the Middle European sample with

Table III. Activities of daily living (*ADL*) Motor and Process Item difficulty calibration values (in logits)

Scale	Region							
	NA	UK/Ireland	Nordic	OEurop	ANZ	Asia	ME	Total
ADL motor								
Endures	0.54	0.63	0.54	0.73	0.50	1.03	0.89	0.56
Lifts	0.45	0.40	0.36	0.44	0.44	0.35	0.55	0.39
Aligns	0.26	0.27	0.50	0.56	0.33	0.37	0.91	0.27
Moves	0.34	0.42	0.33	0.33	0.39	0.33	0.38	0.34
Transports	0.25	0.24	0.14	0.06	0.31	0.12	0.07	0.16
Flows	0.19	0.12	0.10	0.14	0.14	0.11	0.11	0.08
Grips	0.05	0.08	0.16	0.19	0.09	0.09	0.10	0.05
Reaches	0.09	0.12	0.03	0.06	0.13	0.03	0.08	0.13
Bends	−0.06	−0.07	−0.07	−0.11	−0.03	−0.06	−0.02	−0.04
Manipulates	−0.03	−0.19	−0.09	−0.02	−0.02	−0.29	−0.40	−0.04
Walks	−0.11	0.01	−0.19	−0.32	0.03	−0.11	−0.15	−0.20
Stabilizes	−0.24	−0.15	−0.15	−0.23	−0.12	−0.07	−0.11	−0.13
Coordinates	−0.07	−0.22	−0.16	−0.16	−0.07	−0.34	−0.60	−0.07
Paces	−0.34	−0.49	−0.40	−0.32	−0.55	−0.30	−0.53	−0.39
Calibrates	−0.49	−0.55	−0.40	−0.33	−0.64	−0.41	−0.48	−0.41
Positions	−0.82	−0.60	−0.70	−1.12	−0.95	−0.84	−0.97	−0.70
ADL process								
Uses	1.23	1.42	1.34	1.28	1.24	1.36	1.36	1.33
Chooses	0.57	0.75	0.57	0.57	0.67	1.12	0.84	0.67
Sequences	0.58	0.55	0.51	0.63	0.56	0.53	0.58	0.55
Searches/locates	0.50	0.60	0.49	0.38	0.64	0.74	0.55	0.55
Attends	0.33	0.34	0.58	0.72	0.28	0.55	0.45	0.47
Inquires	0.27	0.31	0.30	0.42	0.33	0.35	0.35	0.31
Gathers	0.24	0.27	0.30	0.35	0.33	0.30	0.39	0.30
Heeds	0.24	0.30	0.23	0.27	0.28	0.41	0.29	0.28
Terminates	0.10	0.08	0.07	0.29	0.03	−0.15	0.06	0.07
Navigates	0.04	0.09	−0.11	−0.18	0.07	−0.10	−0.03	−0.04
Handles	0.04	0.02	−0.02	−0.06	0.04	−0.34	−0.24	−0.04
Adjusts	−0.09	−0.18	−0.02	−0.15	−0.09	−0.02	0.23	−0.08
Continues	−0.14	−0.11	−0.19	−0.04	−0.16	−0.17	−0.10	−0.14
Restores	−0.25	−0.31	−0.09	−0.14	−0.21	−0.15	−0.02	−0.18
Initiates	−0.17	−0.33	−0.19	−0.23	−0.24	0.04	−0.22	−0.19
Organizes	−0.16	−0.20	−0.24	−0.31	−0.21	−0.36	−0.40	−0.24
Paces	−0.24	−0.31	−0.38	−0.24	−0.36	−0.38	−0.53	−0.33
Notices/responds	−0.59	−0.55	−0.55	−0.59	−0.61	−0.62	−0.67	−0.58
Benefits	−1.08	−1.18	−1.08	−1.20	−1.14	−1.33	−1.31	−1.15
Accommodates	−1.46	−1.53	−1.50	−1.74	−1.49	−1.81	−1.72	−1.55

NA: North America; UK: UK and Republic of Ireland; Nordic: Nordic Countries; OEurop: Western, Southern, Eastern Europe; ANZ: Australia and New Zealand; Asia: Asia; ME: Middle Europe. Item difficulty calibration values are listed in hierarchical order based on the combined values. The higher the value, the easier the item.

those for each of the other regional groups and with those for the total combined sample revealed that one ADL motor item and none of the ADL process items demonstrated DIF. More specifically, across the 112 comparisons (7 comparison pairs × 16 ADL motor items) for the ADL motor items, 3 differed by at least ± 0.55 logit (2.68%) all 3 of which were for the ADL motor item *Aligns*. Table III shows the ADL motor and ADL process item difficulty calibration values by world region.

When we investigated further in an attempt to identify possible sources of DIF for the ADL motor item *Aligns* (i.e. whether DIF was related to specific raters, age, gender or version of the AMPS manual (English vs German translation) (29)), we found that 8 of 117 Middle European raters scored *Aligns* unexpectedly high compared with raters from the other world regions. Unexpectedly high ratings are very unusual for the ADL motor item *Aligns* as it is one of the easiest ADL

motor items (Table III). Usually, when rater error occurs in relation to scoring *Aligns*, unexpectedly low, not high, ratings are observed. Further investigation revealed no misfit of raters or any other evidence of rater scoring error among these 8 raters. Interestingly, all 8 raters who scored *Aligns* relatively high (easy) were calibrated as stricter than average raters when scoring the items on the ADL motor scale. No other systematic patterns associated with participant age, gender or version of the AMPS manual could be identified, which could explain the DIF.

While the presence of DIF for only one item probably results in little to no risk to the measurement system, we proceeded, as planned, to test for DTF. A total of 6 comparisons for ADL motor ability measures and 6 comparisons for ADL process ability measures were made, and none revealed any evidence of DTF. That is, all paired ADL ability measures of both ADL

scales fell within the 95% confidence bands, and none of the paired measures differed by more than 0.09 logit.

DISCUSSION

The aim of this study was to evaluate cross-regional validity of the AMPS for use in Middle European countries. Given our results, that only one ADL item, *Aligns*, demonstrated DIF, but did not result in DTF, our overall conclusion is that the AMPS measures can be said to be free of cross-regional bias when used in Middle Europe. From a validity perspective, DIF generally raises concern because its presence suggests that persons from different regions have differing probabilities of success on an assessment (23, 34). More specifically, our results indicated that Middle Europeans were overall advantaged when scored on the ADL motor item *Aligns* when compared with the other regional groups.

Because DIF can raise concerns, it becomes important to try to identify the source of the DIF and determine if it might be due to a factor that is resolvable. For example, if Middle Europeans actually perform better on the ADL motor item *Aligns* (i.e. demonstrate a decreased tendency persistently to prop on external objects when performing ADL tasks) (13), than do persons from any of the other world regions, the factor that led to DIF will probably not be "resolvable". If, on the other hand, it can be determined that Middle European raters tend not to assign persons lower scores on *Aligns* when they persistently propped, additional rater training to clarify the scoring criteria for *Aligns* would probably resolve the underlying "cause" of the identified DIF.

One reason Middle European raters scored *Aligns* relatively high compared with raters from other world regions could be due to translation of the AMPS manual into German. However, when we compared raters from AMPS courses trained in Middle Europe that used the original English manual compared with raters trained with the German translation of the manual (29) we did not find any differences in scoring. Another possibility we considered were actual differences in ADL task performance among persons from Middle Europe compared with other world regions. *Aligns*, however, pertains to performance of ADL tasks with the use of persistent need for propping (13), and it seems unlikely that persistent need for propping is related to cultural differences among world regions. Finally, we considered rater scoring error. The 8 Middle European raters who gave higher than expected ratings on *Aligns* had tested a somewhat *higher* proportion of persons who had neurological disorders compared with the other Middle European raters. These are person who have been shown to have *lower* item difficulty calibration values on *Aligns* (13). In addition, the Middle European sample overall had the lowest mean ADL motor ability among the 7 world regions (see Table II), which further suggests that *lower*, not *higher*, ratings on the ADL motor item *Aligns* would be expected. Thus, it appeared to be very likely that rater scoring error among these 8 Middle European raters in relation to the ADL motor item *Aligns* was contributing to the presence of DIF for this item. Removal of these 8 raters minimized the magnitude of the DIF, but did not totally resolve it.

In a study such as the present one, there are several limitations. In the current study, comparisons were made between global world regions, but differences may also exist between countries within the same regional group (e.g. between Germany and Switzerland). Consequently, further cross-country validation studies are suggested. Another limitation is that we only included data from Austria, Germany, Lichtenstein, Slovenia and Switzerland in the Middle European sample as there were no data from other Middle European countries (e.g. Slovakia, Hungary).

In conclusion, the results of this study revealed idiosyncratic DIF for only one ADL item and no evidence of disruption of the measurement system (no DTF) associated with world region. Our results, therefore, support the use of the AMPS to test persons from Middle Europe. Thus, the AMPS can be used in rehabilitation settings and in cross-regional, international collaborative research as an outcome measure for evaluation of effectiveness of rehabilitation services at the levels of activity and participation among clients with diverse diagnoses and disabilities.

ACKNOWLEDGEMENTS

REFERENCES

1. Gutenbrunner C, Ward AB, Chamerlain MA. White Book on Physical and Rehabilitation Medicine in Europe. J Rehabil Med 2007; 45: 6–47.
2. Gutenbrunner C, Ward AB, Chamerlain MA. White Book on Physical and Rehabilitation Medicine in Europe. Euro Medicophys 2006; 42: 292–332.
3. Neumann V, Gutenbrunner C, Fialka-Moser V, Christodoulou N, Varela E, Giustini A, et al. Interdisciplinary team working in physical and rehabilitation medicine. J Rehabil Med 2010; 42: 4–8.
4. Grimby G, Ekholm J, Fisher AG, Stibrant Sunnerhagen K. Editorial: measurement and evaluation of outcomes in rehabilitation. J Rehabil Med 2005; 37: 1–2.
5. Law M, Baum C, Dunn W. Measuring occupational performance: supporting best practice in occupational therapy. 2nd ed. Thorofare, NJ: Slack; 2005.
6. World Health Organization (WHO). ICF: International Classification of Functioning, Disability and Health. Geneva: WHO; 2001.
7. Fisher AG. Overview of performance skills and client factors. In: McHugh Pendleton H, Schultz-Krohn W, editors. Pedretti's occupational therapy: practice skills for physical dysfunction. 6th ed. St Louis: Mosby Elsevier; 2006, p. 372–402.
8. Röse K, Seitz C. Ergotherapeutische Betätigungsziele in der Pädiatrie – welche Faktoren beeinflussen deren Vereinbarung in Deutschland? Ergoscience 2008; 3: 47–56.

9. Raunig F. Mitteleuropa und Identität: Zur Relevanz philosophischen Fragens. In: Flatscher M, Hornstein Tomić C, Periša A, Riefenthaler H, editors. Zur Relevanz mitteleuropäischer Identität. Frankfurt am Main: Peter Langer GmbH; 2010, p. 7–10.

10. Jobst U. Assessments in der neurologischen Rehabilitation. Bewegungstherapie und Gesundheitssport 2008; 24: 230–234.

11. Siebert J, Süess-Marz J, Henkes I, Köster E. Befundinstrumente der Ergotherapie in der Geriatrie. Karlbad: Deutscher Verband der Ergotherapeuten e.V. DVE. Fachkreis Geriatrie; 2007.

12. Schädler S, Kool J, Lüthi H, Marks D, Oesch P, Pfeffer A, et al. Assessments in der Rehabilitation. 2nd ed. Bern: Verlag Hans Huber; 2009.

13. Fisher AG. Assessment of motor and process skills. Vol. 2: user manual 7th ed. Fort Collins, CO: Three Star Press; 2010.

14. Fisher AG, Bray Jones K. Assessment of Motor and Process Skills. Vol. 1: development, standardization, and administration manual. 7th ed. Fort Collins, CO: Three Star Press; 2010.

15. Fisher AG, Griswold LA. Performance skills: implementing performance analyses to evaluate the quality of occupational performance. In: Schell BB, Gillen G, Scaffa M, Cohn E, editors. Willard & Spackman's occupational therapy. 12th ed. Philadelphia, US: Lippincott Williams & Wilkins (in press).

16. Fisher AG. The assessment of IADL motor skills: an application of many-faceted Rasch analysis. Am J Occup Ther 1993; 47: 319–329.

17. Goldman SI, Fisher AG. Cross-cultural validation of the Assessment of Motor Skills (AMPS). Br J Occupat Ther 1997; 60: 77–85.

18. Fisher AG, Merritt BK. Current standardization sample, item and task calibration values, and validity and reliability of the AMPS. In: Fisher AG, Jones KB. Assessment of Motor and Process Skills Vol 1: Development, standardization, and administration manual. Fort Collins, CO: Three Star Press; 2010, p. 15–1–82.

19. Stauffer LM, Fisher AG, Duran L. ADL Performance of Black Americans and White Americans on the Assessment of Motor and Process Skills. Am J Occup Ther 2000; 54: 607–613.

20. Dickerson AE, Fisher AG. Culture-relevant functional performance assessment of the Hispanic elderly. Occup Ther J Res 1995; 15: 50–68.

21. Clawson KJ. AMPS applied cross-culturally to a Mexican-born Hispanic population. Masters thesis. Fort Collins, CO: Colorado State University; 1995.

22. Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates; 2007.

23. Osterlind SJ, Everson HT. Differential item functioning. 2nd ed. Los Angeles: Sage Publications Inc.; 2009.

24. Pae T, Park GP. Examining the relationship between differential item functioning and differential test functioning. Language Testing 2006; 23: 475–496.

25. Wright B, Stone M. Best Test design: Rasch measurement. Chicago: MESA Press; 1979.

26. Zumbo BD. A Handbook on the theory and methods of differential item functioning (DIF): logistic regression modelling as a unitary framework for binary and Likert-type (ordinal) item scores. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defence; 1999.

27. Fisher AG. Occupational Therapy Intervention Process Model: a model for planning and implementing top-down, client-centered, and occupation-based interventions. Fort Collins, CO: Three Star Press; 2009.

28. AMPS Project International. AMPS Computer-Scoring Software (Version 9). Fort Collins: Three Star Press; 2010.

29. Fisher AG. Assessment of Motor and Process Skills. Band II – benutzerhandbuch [Licenced German translation]. 7th ed. Fort Collins: Three Star Press, Inc.; 2011.

30. Linacre JM. Facets: many-facet Rasch measurement computer program. Chicago: MESA Press; 1987–2008.

31. Wilson M. Constructing measures: an item response modeling approach. Mahwah: Lawrence Erlbaum Associates; 2005.

32. Tristán A. An adjustment for sample size in DIF analysis. Rasch Measurement Transactions 2006; 3: 1070–1071.

33. Munkholm M, Berg B, Löfgren B, Fisher AG. Cross-Regional Validation of the School Version of the Assessment of Motor and Process Skills. Am J Occup Ther 2010; 64: 786–775.

34. American Educational Research Association, American Psychological Association, National Counciul on Measurement in Education. Standards for education and psycholocical testing. Washington, DC: American Educational Research Association; 1999.