Cross-Cultural Assessment of Process Skills

Anne G. Fisher, Yihfen Liu, Craig A. Velozo, Ay Woan Pan

Key Words: activities of daily living • activities of daily living evaluation

A standardized activities of daily living evaluation that has acceptable psychometric qualities, can relate discrete component skills to functional performance, includes culture-relevant test items, is standardized on culture-specific samples, and is free of cultural bias is needed to evaluate diverse cultural populations. The Assessment of Motor and Process Skills (AMPS) (Fisher, 1990a) offers a unique solution. The AMPS consists of 35 motor and process skill items assumed to represent two universal taxonomies that are free of cultural bias. The study described in this paper focused on the 20 process skill items of the AMPS process skills scale.

To test the hypothesis that the AMPS process skills scale is suitable for cross-cultural applications, a translation of the AMPS was calibrated on a group of 20 Taiwanese subjects. The validity and reliability of the AMPS process skills scale were examined when applied to this sample. Examination of reliability included the extent to which rater scoring remained stable over time.

The results revealed that the AMPS process skills scale has high intrarater reliability and is valid when applied to young nondisabled Taiwanese subjects. The results suggested that the AMPS could be applied to Taiwanese samples. However, further investigation is needed to determine whether Taiwanese activities can be calibrated onto the same scale as North American activities to make a single cross-cultural AMPS.

The purpose of this study was to examine the use of the process skills scale of a new functional assessment, the Assessment of Motor and Process Skills (AMPS) (Fisher, 1990a), for cross-cultural evaluation. To this end, the process skills scale of the AMPS was administered to a sample of Taiwanese persons, members of a cultural group judged to differ markedly from the North American samples to which the AMPS had previously been applied. The validity and reliability of the AMPS, when used with this distinctly different cultural group, were then examined.

Assessment is a process of collecting, organizing, and interpreting the relevant information that is necessary to plan and implement a meaningful, effective program of treatment. Occupational therapists in the United States and Canada can choose from many standardized and nonstandardized functional evaluation tools. Although some clinicians and researchers consider discrete or global tests of motor, social, physical health, and mental status to be functional evaluations, evaluations of the level of independence in basic and complex activities of daily living (ADLs) are most commonly considered under the rubric of functional evaluation (Lawton, 1987). The number of evaluation tools available for use in physical rehabilitation, gerontology, and psychiatry is reflected in several recent reviews of functional evaluations (Eakin,
Although many ADL evaluations are available, they are consuming, and there is extensive evidence that they come from stroke patients were home-grown evaluations without known validity or reliability. Poorly conceptualized outcome criteria, lack of standardization, disagreement about methods, influence of setting on performance, and cultural bias have also impeded the development of effective functional ADL evaluations (Keith, 1984).

Occupational therapists often supplement their more global evaluations of ADLs by administering a variety of discrete tests, including tests of range of motion, strength, perception, and memory. Such tests assess the discrete component skills (constituents) assumed to be necessary for ADL performance. However, comprehensive evaluations of discrete component skills can be time consuming, and there is extensive evidence that they reveal little information about ADL performance (Fisher, 1990b). That is, the therapist must infer a relationship between discrete evaluations and ADL performance. Moreover, research suggests that the relationship between discrete component skills and ADL performance is too low to allow therapists to make valid inferences about the effect of discrete skill performance deficits on ADL performance (Skurla, Rogers, & Sunderland, 1986). This situation becomes more complicated for occupational therapists interested in evaluating persons from diverse cultural groups. For example, in selecting a Taiwanese sample for use in this investigation, we recognized that few instruments have been standardized on Taiwanese populations. Evaluations that have been adopted by occupational therapists for use in Taiwan are those that are easily translated and presumed to be free of cultural bias (e.g., range of motion, manual muscle testing, sensory testing) (Daniels, Williams, & Worthingham, 1980; Scott & Trombly, 1989; Werner & Omer, 1970). All of these are used by occupational therapists to evaluate discrete component skills.

When Taiwanese occupational therapists assess more complex occupational behavior, including ADLs, they must rely even more heavily on nonstandardized, home-grown evaluations and subjective informal observations, because most existing standardized ADL evaluations were developed for and normative data collected on North American populations. That borrowing ADL evaluations from the United States and Canada is problematic becomes clearer when we consider not only the critical need to translate the instrument from English to Chinese but also the need to restandardize it with Taiwanese samples and tasks relevant to Taiwanese culture.

Mosey (1986) described culture as a set of understandings, shared by members of a group, about how things should be done and what is desirable and good. Thus, culture leads a person to a characteristic way of perceiving and acting in the world, so that the person has a particular life-style (Barris, Kielhofner, Levine, & Neville, 1985). Areas of cross-cultural difference include the inherent nature of the person, language, work, leisure, food, and historical heritage (Mosey, 1986).

Tripp-Reimer, Brink, and Saunders (1984) emphasized how the importance of understanding cultural variables helps the health practitioner to better understand the client's behavior. For example, a standardized evaluation may not be valid when it is used by a therapist to evaluate persons from a cultural group other than one on which it was standardized (Teresi, Cross, & Golden, 1989). Therefore, when researchers standardize ADL instruments on different cultural groups, they should consider whether the language, activity choices, tools and materials used, and cultural standards for performing those activities are relevant to the specific subgroup to be tested. Teresi et al. (1989) summarized reasons why cross-national bias might exist in measurement. Even when countries share a common language (e.g., the United States and Great Britain), differences in the meaning of language can lead to misinterpretation of the rating instructions. Thus, when an assessment is standardized for use with different cultural groups, word-by-word translation is not enough. According to Bettelheim (1983), translators should be very sensitive not only to what is written but also to what is implied. Their task very definitely includes an obligation to try to transmit not just the words forming a sentence but also the meanings to which these words allude. The translators must be responsive to the author's efforts to speak also to the reader's subconscious, to arouse an emotional response as well as an intellectual one. In short, they must also translate the author's attempts to convey covert meanings. (p. 31)

Another reason an ADL instrument may not be valid across cultures is that examiners from different cultural backgrounds use different social and cultural norms and expectations concerning functional independence when assessing whether a person's behavior is appropriate, expected, or excusable (Hickey, 1980). Beall and Eckert (1986) stressed the necessity of understanding the person's particular cultural context, especially when one is trying to measure functional status. Therefore, examiners with the same cultural background as the group to be studied should be included.

Finally, an ADL evaluation may not be valid for cross-cultural comparison if the items relevant to one culture are not relevant to the other culture. Many ADL evaluations standardized on North American samples are not appropriate for Taiwanese samples because the content

The Need for a Culture-Free ADL Evaluation

Although many ADL evaluations are available to occupational therapists in the United States and Canada, most were developed to assess a single disability group (Keith, 1984). Moreover, Jongbloed (1986) found that two thirds of the ADL evaluation tools used to assess functional outcomes of stroke patients were home-grown evaluations without known validity or reliability. Poorly conceptualized outcome criteria, lack of standardization, disagreement about methods, influence of setting on performance, and cultural bias have also impeded the development of effective functional ADL evaluations (Keith, 1984).

Another reason an ADL instrument may not be valid across cultures is that examiners from different cultural backgrounds use different social and cultural norms and expectations concerning functional independence when assessing whether a person's behavior is appropriate, expected, or excusable (Hickey, 1980). Beall and Eckert (1986) stressed the necessity of understanding the person's particular cultural context, especially when one is trying to measure functional status. Therefore, examiners with the same cultural background as the group to be studied should be included.

Finally, an ADL evaluation may not be valid for across-cultural comparison if the items relevant to one culture are not relevant to the other culture. Many ADL evaluations standardized on North American samples are not appropriate for Taiwanese samples because the content

The American Journal of Occupational Therapy
of activities is not relevant to Taiwanese culture. For example, Taiwanese and North American meal preparation practices require different utensils and methods of cooking. The same conditions may exist when evaluators compare cultural groups within the same country (e.g., white, black, Hispanic). Thus, functional capacity evaluations that are truly anthropologically sensitive, with test items free from or controlled for cross-cultural bias, must be developed.

Considered together, these issues point to the need to develop a standardized culture-free ADL evaluation that (a) has acceptable psychometric qualities, (b) can relate directly discrete component skills to functional performance, (c) includes culture-relevant tasks, and (d) is standardized with culture-specific samples.

The Assessment of Motor and Process Skills (AMPS) (Fisher, 1990a) offers a unique solution. The AMPS is an observational assessment that permits the simultaneous evaluation of motor and process skills as a person performs two or three complex or instrumental activities of daily living (IADL) tasks (e.g., meal preparation, home maintenance, laundry) of his or her choice. Motor skills enable the person to cause or impart motion to the body or to objects. They pertain to underlying postural, mobility, coordination, strength, and endurance capacities. Process skills, which were the focus of this study, allow the organized performance of a series of actions enroute to task completion (see Appendix A). They pertain to underlying attentional, conceptual, organizational, and adaptive capacities of the person. By evaluating 15 motor and 20 process skills in the context of a client's performing a daily life task, the therapist can identify specific motor or process skill deficits that directly affect performance of daily living tasks and that should be targeted for intervention.

Each process skill item is rated on a 4-point scale, from deficit skill (1) to competent skill (4) (see Appendix B). The process skill items of the AMPS are hypothesized to represent a universal taxonomy of component skills required for all task performances. They are thought to be free of cross-cultural bias, provided the person is observed performing a task that is (a) relevant to his or her culture and (b) familiar and meaningful to the person within the context of his or her daily life (Fisher, 1990a). However, the hypothesis that the skill items can be applied to cross-cultural contexts has not previously been tested.

Doble (1985) made the first attempt to develop a reliable and quantifiable evaluation of process skills for the psychiatrically disabled adult population. She used a Likert-type rating scale to record subjects' performance after observing them performing daily living activities of their choice. Six areas were assessed: (a) seeking information, materials, and resources; (b) sequencing in time; (c) paying attention; (d) temporal pace; (e) monitoring progress; and (f) evaluating performance. Low to moder-
wok versus frying it in a pan). Moreover, our experience with the AMPS has revealed that Taiwanese therapists differed from North American therapists in the way they scored North American subjects on the AMPS. This difference may have been due not only to different cultural values, but also to Taiwanese therapists' lack of familiarity with the North American tasks observed. Therefore, to test the hypothesis that the AMPS process skills scale is suitable for cross-cultural applications, we had to test Taiwanese subjects performing tasks relevant to Taiwanese culture and to examine the validity and reliability of the AMPS when applied to the Taiwanese sample.

**Method**

**Subjects**

The subjects for this study were 20 nondisabled community-living Taiwanese adults who had resided in the United States for less than 3 years. All subjects were known to the second author and volunteered to participate. Fifteen of the subjects were graduate students at universities in the Chicago area. The remaining five subjects were members of a Taiwanese church in the Chicago area. The mean age of this sample was 28.5 years (SD = 3.32) and their ages ranged from 24 to 35 years. Fifty-five percent (n = 11) of the subjects were men and 45% (n = 9) were women.

**Instrumentation**

The process skills scale of the AMPS was administered to all subjects by the second author in accordance with the standardized administration procedures described in the test manual (Fisher, 1990a). However, two deviations were made from standard AMPS administration and scoring procedures. First, the AMPS manual was translated into Chinese, and the Chinese version of the manual was used for scoring. Second, three new tasks were developed and added to the Taiwanese version of the manual to reflect procedures and materials relevant to Taiwanese culture. In the Chinese version of the manual, the English skill item definitions and scoring examples were translated into Chinese, and the scoring examples were revised to reflect the newly developed tasks. To validate the translations, a back translation method was applied: The second author translated the English version into Chinese, then a bilingual Taiwanese person who was unfamiliar with the AMPS translated the Chinese version back into English. We then examined this back-translated version of the AMPS to ensure that the originally intended meanings were conveyed in the Chinese version. Revisions were made as needed.

The task choice criteria were consistent with those in the English version of the manual (Fisher, 1990a). The tasks were chosen not only for their relevance to Taiwanese culture, but also for their familiarity and potential relevance to a wide range of Taiwanese clients. The tasks were also chosen to be purposeful, functional, or useful and to take the average nondisabled adult 15 to 30 min to complete.

After potential tasks were pilot tested on several nondisabled young adults, five different tasks were chosen for calibration: (a) washing 10 to 15 dishes; (b) preparing toast, fried eggs, and a beverage; (c) cooking Chinese noodle soup; (d) cooking Chinese dumplings; and (e) mopping the kitchen floor. The first two tasks were taken directly from the English version of the AMPS manual, whereas the remaining three were new tasks relevant to Taiwanese culture. All Taiwanese tasks were performed with tools and materials commonly employed in Taiwanese culture. Provided that the process skill items were found to be free of cross-cultural bias, the retention of two tasks that were relevant to both Taiwanese and American culture provided the basis for future linkage of the two versions of the AMPS. Such calibration of tasks relevant to both cultures on the same linear scale would make cross-cultural investigations possible.

**Procedure**

Before the test, the subjects were informed that the project was designed to study the usefulness of a newly developed assessment for evaluating Taiwanese clients. All procedures were fully explained to the subjects, and any questions were answered by the examiner. One to 2 days before the observation, the subject was interviewed to determine which of the five available task choices were familiar to the subject and which of those familiar tasks the subject chose to perform. This interview allowed the examiner to obtain any needed tools and materials and ensured that the subjects were familiar with the task choices. Just before initiating the observation, the examiner verified that the subject understood the task criteria. For example, if the task of cooking noodle soup was chosen, the examiner verified that the subject understood that this task required the use of vegetables, meat, dried noodles, and seasonings and that the subject was to prepare the noodle soup and serve it in a bowl. Finally, the examiner verified what type of meat, vegetables, and seasonings the subject intended to use.

In accordance with the procedures delineated in the manual (Fisher, 1990a), each subject chose and performed tasks with which he or she was familiar. Each subject was tested in an environment that was distraction free, familiar, and relevant to the task being performed. If the environment was unfamiliar to the subject, the subject was fully oriented to it before observations were initiated. All needed tools and materials were readily available and the subject knew where to find them. All subjects performed at least 2 tasks; 46 task observations were scored by one or two trained Taiwanese raters.
All subjects were videotaped as they performed the tasks, and all scoring was based on the videotaped performances. Although the AMPS is designed to be scored without the use of videotaping, the subjects were videotaped so that intrarater reliability could be assessed and the task performances could be scored later by other trained Taiwanese raters. Intrarater reliability was examined by comparing ratings of 20 randomly selected videotaped observations scored 2 weeks apart.

Data Analysis

Most of the analyses were based on the Rasch measurement model (Wright & Masters, 1982; Wright & Stone, 1979). The simple Rasch measurement model is a one-parameter latent-trait model that can be used to develop criterion-referenced hierarchical scales (Hambleton, 1989). In the case of the AMPS, it was necessary to use the many-faceted Rasch model (Linacre, 1988, 1989). The many-faceted Rasch analysis permitted the calibration of item difficulties and task challenges on the same linear scale (Fisher, in press).

One advantage of the many-faceted Rasch model is that measurement is sample-free for the items and tasks and test-free for the persons. That is, the measuring function of the AMPS scales is not affected by the sample used to calibrate the items and tasks (i.e., sample-free), and it is possible to evaluate a person's ability regardless of which items and tasks the person actually performed (i.e., test-free) (Fisher, in press; Linacre, 1989; Wright & Stone, 1979).

As it is applied to the AMPS, the many-faceted Rasch model specifies the following expectations: (a) persons of higher ability are more likely to obtain higher scores than persons of lower ability, (b) a person has a higher probability of obtaining a higher score on an easy item than on a hard item, (c) easier skill items are more likely to be easier for all persons than are hard items, and (d) persons are more likely to obtain higher total scores on less challenging tasks than on more challenging tasks (Andrich, 1988; Linz & Stahl, 1990; Wright & Masters, 1982). When the data conform to these expectations, they will fit the measurement model (Fisher, in press).

The data (i.e., raw skill item scores) are modeled according to these specifications until the expected (estimated) responses predicted by the model are as close as possible to the observed responses (Linz & Stahl, 1990). With the AMPS, the skill item easiness calibration is the estimated location of that skill item on the continuum of increasing IADL process ability. The task challenge calibration is the estimated location of that task on the same continuum of increasing IADL process ability. Finally, the person measure is the estimated location of that person on the continuum of increasing ability that has been defined by the easiness of skill items and the challenges of the tasks, after the ability measure is adjusted for the difficulty of the tasks performed. These calibrations and measures are expressed in equal interval units of measurement based on the logarithm of the odds (log-odds probability units, or logits) of obtaining a given skill item score when a person of a given ability is observed performing a given task (Andrich, 1988; Fisher, in press; Linz & Stahl, 1990; Wright & Masters, 1982).

The detailed fit statistics that are computed by the FACETS computer program (Linacre, 1988) are then examined to verify that a valid measurement system is being constructed. The mean square residuals, which show differences between observed and expected scores, provide a measure of the degree to which the skill items and tasks fit the expectations of the Rasch model. The skill item and task mean square fit statistics verify the internal validity of the AMPS process scale (Fisher, in press).

Scale Validity

Because both tasks and skill items define the AMPS scales, a similar set of research questions was applied to both the process skill items and the five tasks. The first question addressed whether the AMPS process skill items and tasks, when applied to a Taiwanese sample, fit the measurement model. We determined the answer by examining item and task fit statistics. Mean square standardized residuals deviate from the expected value of 1.0 as the observed value deviates from the value expected by the model. For purposes of this study, infit mean square values ≥ 1.3 and ≤ 0.7 were targeted for further examination (Fisher, in press).

The second question was whether the order of the item difficulty and task challenge calibrations were logical or valid. Task challenge calibrations were compared to a priori judgments of rank order task challenges. Washing dishes was judged to be the least challenging task; mopping the kitchen floor and making toast, eggs, and a beverage were judged to be tasks of an intermediate challenge; and preparing Chinese dumplings and noodle soup were judged to be the most challenging tasks.

Person Response Validity and Reliability

The second aspect of validity examined was whether or not the Taiwanese subjects fit the measurement model. That is, were the ability measures derived for each person valid? In this case, we examined infit mean square for the subjects (Fisher, in press).

Finally, we examined whether rater scores remained stable over time (intrarater reliability). Rater infit mean square statistics can be used to evaluate intrarater and interrater consistency for ratings assigned within a single rating period (Fisher, in press). However, to assess the stability of ratings between two rating periods 2 weeks apart, we computed traditional Pearson product-moment correlation coefficients.
correlations between two sets of ratings for 20 randomly selected task observations.

Results

Scale Validity

The results of the many-faceted Rasch analysis for the AMPS process skill items are summarized in Table 1. All items except Accommodates, Heeds, and Chooses had infit mean square values that indicated acceptable fit to the model, suggesting internal validity of the AMPS process skills scale (Fisher, in press).

Task Validity

The internal validity of the AMPS process skills scale when applied to Taiwanese adults was further supported when all five tasks were found to fit the measurement model (infit mean square close to 1) and the calibrated order of task challenge met expectations (see Table 2). Preparing Chinese dumplings was the most difficult task; washing dishes was the easiest task; and preparing toast, eggs, and a beverage; preparing noodle soup; and mopping the floor were tasks of intermediate challenge.

Person Response Validity and Intrarater Reliability

Evaluation of the person fit statistics shown in Table 3 revealed that when the common scale, defined by item difficulties and task challenges, was used to measure

Table 1
Results of Many-Faceted Rasch Analysis for Process Skill Item Easiness Facet

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Skill Item</th>
<th>Score</th>
<th>Count</th>
<th>Calibra-</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>tion (logs)</td>
<td>Infit</td>
</tr>
<tr>
<td>Easy</td>
<td>Inquires</td>
<td>274</td>
<td>94.0</td>
<td>1.13</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Searches</td>
<td>274</td>
<td>94.0</td>
<td>1.13</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Handles</td>
<td>277</td>
<td>94.0</td>
<td>0.65</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>Gathered</td>
<td>267</td>
<td>94.0</td>
<td>0.55</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>Organizes</td>
<td>259</td>
<td>94.0</td>
<td>0.32</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>Sequences</td>
<td>257</td>
<td>94.0</td>
<td>0.26</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>Continues</td>
<td>255</td>
<td>94.0</td>
<td>0.40</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>Accommodates</td>
<td>255</td>
<td>94.0</td>
<td>0.20</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>Uses</td>
<td>248</td>
<td>94.0</td>
<td>0.02</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Benefits</td>
<td>248</td>
<td>94.0</td>
<td>0.02</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Navigates</td>
<td>252</td>
<td>94.0</td>
<td>-0.31</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Addresses</td>
<td>230</td>
<td>94.0</td>
<td>-0.34</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Heeds</td>
<td>250</td>
<td>94.0</td>
<td>-0.34</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Terminates</td>
<td>228</td>
<td>94.0</td>
<td>-0.38</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Adjusts</td>
<td>227</td>
<td>94.0</td>
<td>-0.59</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Initiates</td>
<td>226</td>
<td>94.0</td>
<td>-0.41</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Chooses</td>
<td>223</td>
<td>94.0</td>
<td>-0.46</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Places</td>
<td>219</td>
<td>94.0</td>
<td>-0.52</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Notices</td>
<td>216</td>
<td>94.0</td>
<td>-0.56</td>
<td>0.12</td>
</tr>
<tr>
<td>Hard</td>
<td>Restores</td>
<td>198</td>
<td>92.0</td>
<td>-0.76</td>
<td>0.12</td>
</tr>
<tr>
<td>M</td>
<td></td>
<td>241</td>
<td>93.9</td>
<td>0.00</td>
<td>0.17</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>21</td>
<td>0.4</td>
<td>0.53</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Table 2
Results of Many-Faceted Rasch Analysis for Task Challenge Facet

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Task</th>
<th>Score</th>
<th>Count</th>
<th>Calibra-</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>tion (logs)</td>
<td>Infit</td>
</tr>
<tr>
<td>Easy</td>
<td>Wash dishes</td>
<td>1080</td>
<td>380</td>
<td>0.68</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Mapping</td>
<td>475</td>
<td>180</td>
<td>0.05</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Noodle soup</td>
<td>780</td>
<td>29.9</td>
<td>-0.11</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Toast/drink</td>
<td>815</td>
<td>520</td>
<td>-0.15</td>
<td>0.08</td>
</tr>
<tr>
<td>Hard</td>
<td>Dumplings</td>
<td>1682</td>
<td>699</td>
<td>-0.40</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Taiwanese subjects, all of the subjects except Subject 16 fit the measurement model. This suggests that valid person response patterns are elicited when the AMPS process scale is used to test Taiwanese adults. The Pearson product-moment correlation between the two sets of 20 ratings by the same examiner was \( r = .93 (p < .001) \). These results suggest that scores of trained raters remain stable over time.

Discussion

The findings suggested that most process skill items and all tasks fit the measurement model and define a variable—ADL process skill—of increasing intensity. The unusually high infit statistic values for Heeds and Chooses indicated that the subjects' responses on these two items were erratic or more variable than expected. Further in-

Table 3
Results of Many-Faceted Rasch Analysis for Person Ability Facet

<table>
<thead>
<tr>
<th>Ability</th>
<th>No.</th>
<th>Age (Years)</th>
<th>Sex</th>
<th>Score</th>
<th>Count</th>
<th>Calibra-</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>tion (logs)</td>
<td>Infit</td>
</tr>
<tr>
<td>More Able</td>
<td>11</td>
<td>29</td>
<td>F</td>
<td>341</td>
<td>120</td>
<td>2.60</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>25</td>
<td>F</td>
<td>230</td>
<td>80</td>
<td>2.55</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>30</td>
<td>M</td>
<td>478</td>
<td>190</td>
<td>2.12</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>35</td>
<td>M</td>
<td>104</td>
<td>40</td>
<td>2.11</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>30</td>
<td>M</td>
<td>511</td>
<td>120</td>
<td>2.07</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>25</td>
<td>M</td>
<td>103</td>
<td>40</td>
<td>2.06</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>25</td>
<td>F</td>
<td>100</td>
<td>40</td>
<td>1.92</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>34</td>
<td>M</td>
<td>521</td>
<td>190</td>
<td>1.89</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>30</td>
<td>M</td>
<td>210</td>
<td>80</td>
<td>1.89</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>29</td>
<td>F</td>
<td>199</td>
<td>80</td>
<td>1.88</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>29</td>
<td>M</td>
<td>325</td>
<td>120</td>
<td>1.98</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>24</td>
<td>F</td>
<td>99</td>
<td>40</td>
<td>1.87</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>27</td>
<td>F</td>
<td>197</td>
<td>80</td>
<td>1.85</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>30</td>
<td>F</td>
<td>253</td>
<td>90</td>
<td>1.85</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>24</td>
<td>F</td>
<td>98</td>
<td>40</td>
<td>1.85</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>25</td>
<td>F</td>
<td>286</td>
<td>120</td>
<td>1.71</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>30</td>
<td>M</td>
<td>456</td>
<td>180</td>
<td>1.68</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>27</td>
<td>M</td>
<td>159</td>
<td>60</td>
<td>1.65</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>34</td>
<td>M</td>
<td>296</td>
<td>120</td>
<td>1.57</td>
<td>0.12</td>
</tr>
<tr>
<td>Less Able</td>
<td>19</td>
<td>29</td>
<td>M</td>
<td>85</td>
<td>40</td>
<td>1.36</td>
<td>0.18</td>
</tr>
<tr>
<td>M</td>
<td></td>
<td>241</td>
<td>94</td>
<td>1.91</td>
<td>0.17</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>131</td>
<td>49</td>
<td>0.26</td>
<td>0.05</td>
<td>0.2</td>
<td></td>
</tr>
</tbody>
</table>

Note: M = male, F = female.
vestigation of the possible reason for the high positive fit statistics on Heeds revealed that this was the most commonly misfitting item. Five of the subjects forgot a step (e.g., swept but did not mop the floor, prepared eggs and toast but no beverage). Similar analysis of the responses of the subjects on the skill item Chooses revealed that the high positive fit statistics seemed to result from the performance of the same 5 subjects who forgot to obtain the predetermined steps and the related tools and materials. Upon completion of the task observation, all 5 of these subjects indicated that they were aware that they had omitted the predetermined steps and the related tools and materials.

The subjects did not specify why they made errors and then did not correct their errors before completing the task. Possible reasons may have included the effects of videotaping the subjects or their mistaken perception that they should not correct their errors. In either case, this occurrence may be culture specific, because we have found that it is unusual for North American subjects to omit a major step of a task in the absence of obvious cognitive deficits (e.g., dementia, mental retardation).

The infit mean square statistics for Accommodates suggested that scores on this item were too predictable or too invariant. This condition occurs when an item measures overlapping aspects of abilities that are also measured by other items or when subjects all obtain similar scores. Because none of the other skill items had similarly low infit mean square values, which would indicate redundancy in the items, it seems that the low mean square for Accommodates resulted because the majority of the subjects obtained scores of 4 (see Fisher [in press] for more details on the process of analyzing fit statistics).

The findings indicated that all the tasks fit a line of increasing task challenge, and the calibrated challenge of these tasks was similar to what was expected, on the basis of a priori judgment. Comparison of the difficulty calibrations of the five tasks, expressed in logits (see Table 2), revealed that a large gap occurred between the dish washing and the floor mopping tasks and between the Chinese dumpling and eggs, toast, and beverage tasks. Thus, consideration should be given to adding tasks that would offer challenges to fill these voids. More important, however, is the need to define more challenging tasks if the AMPS is to be used to test young nondisabled adults, especially for research purposes.

The results suggested that when we used the AMPS process skills scale, defined by the items and tasks, only Subject 16 failed to fit the measurement model. This subject omitted steps on two of the tasks that she performed, and this performance contributed to the misfit of the items Heeds and Chooses with the model. These results should be considered carefully because of the homogeneity and the high ability levels of this young, nondisabled adult sample. Further investigation should include more difficult tasks and further examination of their usefulness for assessing more heterogeneous samples of young and older subjects. Finally, intrarater reliability was high for 20 task observations for 9 subjects. This finding provided preliminary evidence that raters do remain consistent over time. Further investigation should include more raters to examine the generalizability of these results.

Study Limitations and Clinical Implications

This study provides evidence that the process skills scale of the AMPS can be applied to the Taiwanese population, and a common scale of subject ability, task challenge, item difficulty, and rating scale steps can be developed. That is, the finding that items, tasks, and subjects generally fit the expectations of the measurement model suggests that the AMPS can be used within diverse cultural contexts.

Further study is needed to determine whether activities relevant to Taiwanese culture can be calibrated onto the same linear scale as is used for North Americans. However, although the estimated item and task calibrations were based on 94 scored observations between two raters, the number of subjects in this study was relatively small. Therefore, the results should not be generalized to the Taiwanese population without further study.

Summary and Recommendations

The purpose of this study was to examine the intrarater reliability and validity of the process skills scale of the AMPS when used to assess Taiwanese adults. The results demonstrated high reliability and validity when applied to young, nondisabled Taiwanese subjects as they performed tasks relevant to Taiwanese culture.

On the basis of our findings, we propose the following recommendations for further study:

- Add more difficult tasks to the AMPS and further examine its usefulness for assessing young nondisabled and mildly impaired subjects.
- Compare equivalent samples representative of different ethnic (e.g., black, Hispanic, Asian, white) and national (e.g., American, Taiwanese, British) origins and examine for the presence of item or rater bias. This information would enable occupational therapists to know to what extent it is possible to develop a cross-cultural AMPS or whether different versions of the AMPS should be developed for different cultural groups.

Acknowledgments

This project was supported in part by funding from the American Occupational Therapy Foundation through the Occupational Therapy Center of Research and Measurement at the University of Illinois at Chicago, and was completed in partial fulfillment of the second author’s requirements for a master of science degree. Portions of this paper were presented at the 7th Annual Conference of the American Occupational Therapy Association, Cincinnati, Ohio, June 1991.
Appendix A
Process Skill Item Definitions

Energy
- Maintains focused attention throughout the task sequence; implies the client can selectively focus attention on the task to be performed and appropriately allocate attentional resources to relevant aspects of the task and environment such that the client (a) is not distracted by extraneous auditory or visual stimuli, or (b) does not “over-attend” to stimuli or certain aspects of the task while disregarding others.

Using Knowledge
- Selects appropriate tools and materials; implies an understanding of what to choose to gather when specified.
- Uses goal-directed task performance that is focused on the completion of the specified task (i.e., the outcome originally agreed on), implies having a basic understanding of the goal or purpose of the task, and an absence of behavior driven by environmental cues. Pertains to the ability to manage one’s actions and behaviors in order to accomplish the specified task.
- Collects together needed or misplaced tools and materials; pertains to (a) the collection of located supplies into the workspace for the performance of the task, (b) the collection and replacement of materials that have been spilled or dispersed, and (c) the retrieval of misplaced or fallen supplies.
- Organizes logically positions or spatially arranges tools and materials in an orderly fashion in and between appropriate workspaces in order to facilitate ease of task performance.
- Maintains a rate or tempo of performance adequate to the initiation of the task, this also includes choosing and using the originally-agreed-on tools and materials.
- Supports, stabilizes, and holds tools and materials in an appropriate manner given the circumstances of the situation and abilities of the individual: pertains to recognizing the need, and knowing how, to hold, stabilize, and support objects. Includes providing support to protect tools and materials from damage or falling (dropping).
- Collects together needed or misplaced tools and materials; pertains to (a) the collection of located supplies into the workspace for the performance of the task, (b) the collection and replacement of materials that have been spilled or dispersed, and (c) the retrieval of misplaced or fallen supplies.

Space and Objects
- Looks for and locates tools and materials through the process of logical searching; pertains to the ability to investigate and look beyond the immediate environment in order to locate necessary or dispersed tools and materials (e.g., looking in, behind, on top of).
- Collects together needed or misplaced tools and materials; pertains to (a) the collection of located supplies into the workspace for the performance of the task, (b) the collection and replacement of materials that have been spilled or dispersed, and (c) the retrieval of misplaced or fallen supplies.
- Uses goal-directed task performance that is focused toward the completion of the specified task (i.e., the outcome originally agreed on), implies having a basic understanding of the goal or purpose of the task, and an absence of behavior driven by environmental cues. Pertains to the ability to manage one’s actions and behaviors in order to accomplish the specified task.
- Collects together needed or misplaced tools and materials; pertains to (a) the collection of located supplies into the workspace for the performance of the task, (b) the collection and replacement of materials that have been spilled or dispersed, and (c) the retrieval of misplaced or fallen supplies.
- Collects together needed or misplaced tools and materials; pertains to (a) the collection of located supplies into the workspace for the performance of the task, (b) the collection and replacement of materials that have been spilled or dispersed, and (c) the retrieval of misplaced or fallen supplies.
- Collects together needed or misplaced tools and materials; pertains to (a) the collection of located supplies into the workspace for the performance of the task, (b) the collection and replacement of materials that have been spilled or dispersed, and (c) the retrieval of misplaced or fallen supplies.

Temporal Organization
- Starts or begins doing an action or step without hesitation; implies an end to decision-making.
- Performs an action sequence of a step without unnecessary interruption and as an unbroken, smooth progression; pertains to the continuity of a series of actions such that, once an action sequence is initiated, the individual continues on until it is completed.
- Performs steps in an effective or logical order for efficient use of time and energy; implies an absence of randomness in the ordering, or the inappropriate repetition ("re-ordering"), of steps.
- Maintains a rate or tempo of performance across the entire task; implies the maintenance of a rate that permits the completion of the task within a reasonable amount of time (i.e., lack of hypo- or hyperactivity, slowing over time, or an uneven pace). Pertains to the use of an effective rate of performance throughout the steps of the specified task.
- Finishes or brings to completion single actions or steps without perseveration, inappropriate persistence, or premature cessation; implies stopping ongoing task performance in preparation for beginning the next action or step.

Adaptation
- Modifies one’s actions in anticipation of, or in response to, circumstances/problems that might arise in the course of action, or that require attention to avoid undesirable outcomes. The main focus of this behavior is that the individual change the method with which he/she is performing the action sequence, or the manner in which he/she interacts with or handles tools and materials already in the workspace.
- Modifies environmental conditions in anticipation of, or in response to, circumstances/problems that arise in the course of action or that require attention to avoid undesirable outcomes. The main focus of this behavior is that the individual make some change in the working environment by moving to a new workspace, bringing or removing tools and materials from the present workspace, or by changing an environmental condition (i.e., light, temperature, etc.).
- Modifies the movement pattern of the arm, body, or wheelchair to avoid or maneuver around existing obstacles that are encountered in the course of moving the arm, body, or wheelchair through space, and that require attention to prevent undesirable contact with obstacles (e.g., knocking over, stepping on, bumping into). Includes visually-guided arm movements and the ability to hold and maneuver an object around obstacles. The main focus of this behavior is that the individual make some change in the movement trajectory of the arm or hand when reaching, the body when walking, or the wheelchair when moving around the environment.
- Anticipates and prevents undesirable circumstances/problems from recurring or persisting. The main focus of this behavior is that the individual (a) recognize what actions already have been completed, what problems already have occurred, or the potential for recurrence or persistence of a circumstance problem; (b) learns from prior actions and adaptations, or from requested information; and (c) uses prior actions, adaptations, or information to alter the task progression accordingly.

The American Journal of Occupational Therapy
Appendix B

Scoring Examples for the Process Skill Item Searches

Searches = looks for and locates tools and materials through the process of logical searching, pertains to the ability to investigate and look beyond the immediate environment in order to locate necessary or dispersed tools and materials (e.g., looking behind, on top of). (Note: Score obtaining incorrect or inappropriate tools and materials under the process verb "chooses. Asking for the location of an object is also scored under the process verbs "inquires and accommodates.)

4 = readily and consistently searches for tools and materials through the process of logical searching
   - systematically opens one drawer after another to find a knife
   - systematically moves objects aside to find a broom in the back of a closet
   - systematically looks for misplaced plate

3 = questionable searching skill, but no apparent disruption of action item or task performance, or impact on other skill items
   - examiner questions the appropriateness of the need to search within a familiar environment
   - examiner questions the search strategy used

2 = ineffective searching skill impacts on action item or task performance, or results in inefficient use of time or energy
   - opens drawers in random order to search for a knife
   - has to search a closet a second time to find the broom
   - has difficulty finding a knife in the drawer in a familiar kitchen
   - in an unfamiliar room, asks for help in searching for a wall plug before attempting to look for one
   - has difficulty relocating a plate that previously had been searched for and located or gathered by the client during performance of the task
   - delay in searching for missing, but needed knife, slows task progression

1 = severity of searching skill deficit clearly impedes action item or task performance such that the results are unacceptable, or damage or danger is imminent
   - searches for, but marked delay in, locating a knife
   - random, ineffective search strategy results in marked delay before a broom is located
   - in a familiar kitchen, asks for help before attempting to search for a knife
   - unable or marked delay in relocating a plate that previously had been searched for and located or gathered by the client during performance of the task
   - therapist intervention required because severity of searching skill deficit results in unacceptable delay or task breakdown
   - does not search for missing, but needed, knife

References


Wright, B. D., & Stone, M. H. (1979). *Best test design.* Chicago: MESA.

**Coming in November:**

Special Issue on Feminism as an Inclusive Perspective

- Women and unpaid work in the home
- Gender differences in dementia management plans of spousal caregivers
- Women, HIV, and AIDS
- Weaving the fibers of motherhood, occupational therapy, and feminism

Turn to *AJOT* for the latest information on occupational therapy treatment modalities, aids and equipment, legal and social issues, education, and research.