Objective. The purpose of this study was to examine whether the Assessment of Motor and Process Skills (AMPS), an assessment of personal and domestic activities of daily living (ADL) performance, can be used as a valid, nonbiased tool when assessing black Americans.

Method. The participants were 466 blacks and 466 whites drawn from the entire sample of blacks and whites contained in the AMPS database who met the following criteria: (a) were 16 years of age and older; (b) had a notable history of a neurological, musculoskeletal, medical, developmental, cognitive, or psychiatric disorders or were healthy older persons; and (c) resided in North America. The participants were matched according to functional level, gender, diagnosis, and age. Examination for bias included between-group comparison of (a) item difficulty and task challenge hierarchies of the AMPS, (b) goodness-of-fit of the participants to the many-faceted Rasch (MFR) model, and (c) mean ADL motor and ADL process abilities.

Results. Both the item difficulty and the task challenge hierarchies remained stable between the two groups. On the ADL Motor scale, 95.3% of the black participants and 92.4% of the white participants demonstrated acceptable goodness-of-fit (MS ≤ 1.4, z < 2) to the MFR model. On the ADL Process scale, 91.2% of the black participants and 90.1% of the white participants demonstrated acceptable goodness-of-fit. A significant difference, t(2, 930) = 3.56, p < .01, between the two groups was found in mean ADL process ability, but no significant difference, t(2, 930) = .69, p = .49, was found in mean ADL motor ability.

Conclusion. The results of this study support the validity of the AMPS when applied to black Americans.

The Assessment of Motor and Process Skills (AMPS) (Fisher, 1999) is a tool used by occupational therapists to assess and measure the effectiveness of a person's occupational performance in a culturally relevant context. In the AMPS theoretical framework, occupational performance is viewed as “a meaningful sequence of actions in which the person enacts and completes a specified task that is relevant to his or her culture or daily life roles” (Fisher, 1999, p. 24).

The AMPS contains 76 different standardized personal and domestic activities of daily living (ADL), ranging from simple to complex. The variety of different ADL tasks provides a number of options from which persons from different cultural and ethnic backgrounds can choose. Moreover, to allow for widespread applicability of AMPS tasks to persons from diverse backgrounds, even within more global cultural groups (e.g., Western, North American, black), the AMPS tasks have been standardized in a manner that allows individuals to perform tasks in their usual manner, perhaps reflecting microcultures (e.g., ethnic subgroups, black subcultures, family units). The standardization of the AMPS tasks for use cross-culturally has been discussed in detail elsewhere (Bernspång & Fisher, 1995b; Dickerson & Fisher, 1995; Fisher, 1997, 1999; Goldman & Fisher, 1997; Goto, Fisher, & Mayberry, 1996).

When a trained and calibrated occupational therapist observes the client performing a task of his or her choice, the therapist scores the client on 16 ADL motor and 20 ADL process skill items. Each AMPS skill item is a verb naming the universal, learned, and goal-directed motor and process actions that comprise all ADL task performances, regardless of the task performed and cultural or ethnic group to which the person belongs (e.g., grasping, lifting, transporting task objects; searching, locating, and gathering task objects). When scoring each ADL motor and ADL process skill item, the therapist considers the client’s degree of physical effort, efficiency, independence, and safety as each goal-directed action is enacted during the client’s ADL task performance.

The ADL motor and ADL process skill items included in the AMPS are asserted to possess the same meaning across cultural or ethnic subgroups, which may ultimately affect estimations of ADL motor or ADL process ability measures.

In modern test theory, a test is judged to be free of bias when the set of items or tasks that comprise the test does not place any given subgroup at an unfair disadvantage because an item or task is less difficult for members of that subgroup than for members of another subgroup (Crocker & Algina, 1986; Embretson & Hershberger, 1999). Inclusion of items or tasks in a test that remain stable (in terms of difficulty) across ethnic subgroups are preferred over those that are biased (Embretson & Hershberger, 1999).

It is not sufficient, therefore, to merely assert that the AMPS motor skill items, process skill items, and ADL tasks are free of ethnic bias. It is important that the ADL motor and ADL process test items and tasks be evaluated for the presence of differential item response between ethnic subgroups, which may ultimately affect estimations of ADL motor or ADL process ability measures.

The purpose of this study was to use many-faceted Rasch (MFR) analysis (Linacre, 1993) to examine whether the AMPS can be used as a valid, unbiased tool when assessing blacks. One way to examine for ethnic bias is to use MFR analysis to examine the skill item and task challenge calibration hierarchies, verifying their stability among ethnic groups. If differential item response on either tasks or items is present, and if there is a significant difference in the mean ADL motor or ADL process ability measures of the two groups, then bias may be present within the assessment. If the skill item difficulty and task challenge calibration hierarchies are stable between the groups, then neither subgroup is being placed at an unfair disadvantage because of differential item response when being assessed with the AMPS (Crocker & Algina, 1986; Embretson & Hershberger, 1999).

An additional method for examining bias is to evaluate whether the participants from a given ethnic group demonstrate response patterns across test items that conform to the assertions of the MFR model (Fisher, 1993, 1994a, 1997, 1999; Linacre, 1993; Wright & Stone, 1979). When a person demonstrates unexpected response patterns, he or she will not demonstrate acceptable goodness-of-fit to the MFR model. When several members of one (e.g., minority) ethnic group unexpectedly misfit the MFR model, there may be some reason to suspect ethnic bias, especially when the percentage of persons in a minority subgroup who do not demonstrate acceptable goodness-of-fit to the MFR model of the AMPS is greater than that found in the majority reference group.
The AMPS has been examined internationally for bias across groups in North America, Sweden, and the United Kingdom (Goldman & Fisher, 1997; Magalhães et al., 1996). The findings of these studies suggest that the AMPS is an unbiased tool for use in these three regions because the hierarchical order of the AMPS skill item difficulty and task challenge calibrations remained stable between the three regions. There also is some evidence that it is free of bias when used across black, Mexican-born, and Japanese subcultures within the United States (Bennett 1995; Clawson, 1995; Goto et al., 1996). These studies, however, used small sample sizes (n < 200), so the presence of differential item response could not be investigated.

A need remains, therefore, to examine the stability of the ADL motor and ADL process item difficulty and task challenge calibration hierarchies between whites and blacks using the larger sample sizes currently available in the AMPS database. The following main research questions were examined:

1. Do meaningful differences exist between the ADL motor and ADL process item difficulty or task challenge calibration hierarchies for blacks versus whites?
2. Do blacks and whites demonstrate comparable rates of acceptable goodness-of-fit of the MFR model of the AMPS?

We also addressed the following secondary question: Do the mean ADL motor or ADL process ability measures differ significantly between black and white persons?

Method

Participants

The participants in this study were drawn from the sample of all black and white individuals contained in the AMPS database as of December 1997 who met the following criteria: (a) were 16 years of age and older; (b) were representative of those persons with whom the AMPS is used in that they had a notable history of a neurological, musculoskeletal, medical, developmental, cognitive, or psychiatric disorders or were healthy older persons; and (c) resided in North America. The total number of available participants who met our inclusion criteria was 2,974 whites and 522 blacks. All participants with unexpectedly high AMPS ability measures (> 4.05 on the AMPS Motor scale and > 3.05 on the AMPS Process scale) were eliminated from this sample. Such unexpectedly high ratings can be attributed to rater scoring error that can bias the results.

Eliminated participants were 125 (4%) of the total available white participants and 33 (6%) of the total available black participants. In addition, 44 (1%) participants were co-rated by multiple raters as a part of the rater calibration process were eliminated from the sample because multiple ratings also could potentially bias the results. Finally, because of missing demographic data relevant for matching to a white sample, 20 (4%) of the 522 black participants were eliminated. This left a final sample of 466 black participants.

An equal number of white participants from the AMPS database were matched to the black sample, with first-level matching done for gender and functional level (i.e., able to live independently and safely in the community, requires minimal assistance or supervision to living in the community, requires moderate to maximal assistance to live in the community) and second-level matching for diagnostic grouping and age. The matching was done by a person who was unaware of the ADL motor or ADL process ability measures of the participants or the purpose of this study. Detailed demographic data for the sample are shown in Table 1.

The mean age of the black sample was 54.7 years (SD = 18.2, range = 17 to 91 years). The mean age of the white sample was 54.8 years (SD = 18.3, range = 16 to 91 years). A t test verified that the two groups were matched for age, t (2, 930) = .10; p = .92.

Instrument

The AMPS is an occupational therapy–specific ADL assessment that has been fully standardized on more than 25,000 persons internationally and cross-culturally. The AMPS is administered by trained and calibrated occupational therapists according to the standardized procedures described in the AMPS manual (Fisher, 1999).

When scoring an AMPS observation, the calibrated rater scores 16 ADL motor items and 20 ADL process items with a 4-point rating scale (4 = competent, 3 = questionable, 2 = ineffective, 1 = unacceptable). As noted earlier, when rating each skill item, the therapist considers whether the participant experiences increased effort, decreased efficiency, decreased safety, or need for assistance on the basis of detailed scoring criteria included in the AMPS manual (Fisher, 1999).

### Table 1

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Male White</th>
<th>Female White</th>
<th>Male Black</th>
<th>Female Black</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthy older adult</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Developmental disability</td>
<td>4</td>
<td>5</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Memory impaired or dementia</td>
<td>5</td>
<td>3</td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td>Musculoskeletal disorder</td>
<td>16</td>
<td>14</td>
<td>38</td>
<td>40</td>
</tr>
<tr>
<td>Hemispheric stroke</td>
<td>45</td>
<td>43</td>
<td>51</td>
<td>54</td>
</tr>
<tr>
<td>Other neurological (e.g., Parkinson’s disease, brainstem stroke)</td>
<td>27</td>
<td>26</td>
<td>33</td>
<td>24</td>
</tr>
<tr>
<td>Medical diagnoses (e.g., cardiac, cancer)</td>
<td>16</td>
<td>25</td>
<td>23</td>
<td>22</td>
</tr>
<tr>
<td>Multiple diagnoses</td>
<td>50</td>
<td>50</td>
<td>62</td>
<td>63</td>
</tr>
<tr>
<td>Functional level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Able to live independently in the community</td>
<td>37</td>
<td>39</td>
<td>46</td>
<td>45</td>
</tr>
<tr>
<td>Requires minimal assistance or supervision</td>
<td>73</td>
<td>76</td>
<td>120</td>
<td>111</td>
</tr>
<tr>
<td>Requires moderate to maximal assistance</td>
<td>68</td>
<td>70</td>
<td>123</td>
<td>125</td>
</tr>
</tbody>
</table>
An MFR computer program is used to convert the raw, ordinal data into equal interval data in the form of ADL ability measures (Fisher, 1993, 1994a). The MFR model for the AMPS takes into account the task challenge, rater severity, and item difficulty when estimating the final participant ADL motor and ADL process ability measures. The MFR model has been discussed in detail in other studies (Bernspång & Fisher, 1995a, 1995b; Doble, Fisk, Fisher, Rivvo, & Murray, 1994; Fisher, 1993, 1994a, 1997, 1999; Magalhães et al., 1996).


Procedure and Data Analysis

Before we began, this research project was approved by the Human Subjects Committee at Colorado State University. The MFR computer program, FACETS (Linacre, 1993), was used to analyze the raw motor and process skill item scores and calculate an ADL motor and an ADL process ability measure for each participant. A person with a higher ADL ability measure is more able than a person with a lower ADL ability measure.

FACETS was used to calculate separate item difficulty calibrations and task challenge calibrations for the two ethnic groups. Differences between the calibration values for each group were then calculated to evaluate whether meaningful differences existed between groups. We set our criteria for a meaningful group difference, between-item difficulty calibration at greater than .43 logit, a value equivalent to a 95% confidence interval when the standard error for the calibration values is .15 logit. The rationale for setting our criteria at greater than .43 logit has been discussed elsewhere (Bernspång & Fisher, 1995b; Duran & Fisher, 1996; Magalhães et al., 1996; Silverstein, Fisher, Kilgore, Harley, & Harvey, 1992).

To investigate whether the black participants and white participants demonstrated comparable rates of acceptable goodness-of-fit to the MFR model, the mean square (MS) infit and outfit values and their associated z statistics generated by FACETS for each participant were evaluated. We considered a participant to misfit if his or her fit statistics exceeded the criteria MS > 1.4 and z ≥ 2. We expected a comparable percentage of misfitting black participants and white participants in our sample.

Finally, two t tests were performed on the data to determine whether significant differences existed in the mean ADL ability measures between the two ethnic groups. The level of significance was set at p < .05.

Results

Item Difficulty Calibrations and Task Challenge Calibrations

Comparison of the ADL motor and ADL process item difficulty calibrations for the white and black groups revealed that none of the ADL motor or ADL process items differed by more than .43 logit; the maximum group difference in calibrations was .33 logit on the ADL Motor scale and .35 logit on the ADL Process scale. We concluded, therefore, that the item difficulty calibrations remained stable between the two ethnic groups.

Comparison of the ADL motor and ADL process task challenge calibrations for the 35 tasks performed by a minimum of five participants in each group revealed that none of the task challenge calibrations differed more than .43 logit; the maximum difference in calibrations was −27 on the ADL Motor scale and .24 logit on the ADL Process scale. The comparisons were limited to those tasks performed by at least five participants in each group to ensure reasonably stable task calibration values. We concluded that the task challenge calibrations remained stable between the two ethnic groups.

Goodness-of-Fit

On the ADL Motor scale, 444 (95.3%) of the 466 black participants and 431 (92.4%) of the 466 white participants demonstrated acceptable goodness-of-fit to the MFR model. On the ADL Process scale, 425 (91.2%) of the black participants and 431 (92.4%) of the white participants demonstrated acceptable goodness-of-fit to the MFR model. The black participants had a slightly higher overall rate of acceptable goodness-of-fit than did the white participants. We concluded, therefore, that the black minority sample did not demonstrate a greater rate of misfit than the white majority sample. Moreover, the overall rate of misfit between the two groups was comparable.

Comparison of Mean Ability

The mean ADL motor and ADL process abilities of each group are shown in Table 2. The t tests revealed no significant difference in mean ADL motor ability between groups, but a significant difference was found between...
groups in the mean ADL process ability. The white participants had a mean ADL process ability measure greater than their black counterparts. We concluded, therefore, that the black participants demonstrated significantly lower mean ADL process ability than did the white participants.

This was an unexpected finding, so we completed a secondary analysis in an attempt to discover the source. Included in our sample were 75 participants who had completed only one AMPS task. To ensure the highest possible rate of participant fit to the MFR model, however, the standardized testing procedures for the AMPS recommend that persons perform at least two AMPS tasks for reliable estimation of ADL ability measures (Fisher, 1999); performance of only one AMPS task can reduce the precision of the person's final estimation of ADL ability (Kirkley & Fisher, 1999). Because of the potential confound of data from the 75 participants who completed only one task on comparisons of mean ability between the two participant groups, we eliminated these 75 participants from the sample in this calculation. We also eliminated an additional 17 participants from the original sample to equate the numbers of participants in the new sample in each of the first-level matching criteria groups (functional level, gender). These 17 participants were chosen at random from all available participants who were members of functional level and gender groups with surplus representation. The person doing the elimination was, again, blind to each participant's ADL motor and ADL process ability. The new sample contained 420 participants in each ethnic group.

To investigate whether the participants who only completed one task confounded the comparison of mean ability, we recalculated ADL motor and ADL process ability means and performed associated t tests on the new sample. The results of this secondary analysis were similar to those of the first analysis. The mean ADL motor ability of the black participants was 1.4 logit (SD = 1.3, range = –3.0 to 3.9 logit) and the mean ADL motor ability of the white participants was 1.4 logit (SD = 1.2, range = –3.0 to 3.9 logit). A t test verified that the two groups did not differ significantly on the ADL Motor scale, t(2, 838) = –.88, p = .38. The mean ADL process ability of the black participants was .7 logit (SD = .7, range = –3.0 to 2.8 logit) and the mean ADL process ability of the white participants was .9 logit (SD = .9, range = –3.0 to 2.9 logit). A t test verified that the two groups again differed significantly on the ADL Process scale, t(2, 838) = 3.09, p < .01.

### Discussion

The overall results of this study support the findings of other studies that the AMPS can be used with varied and diverse cultural or ethnic groups (Dickerson & Fisher, 1995; Goldman & Fisher, 1997; Goto et al., 1996). That is, we found that none of the AMPS motor or process items and none of the AMPS tasks demonstrated differential item response that would indicate an internal test bias toward one of the participant groups. In addition, the finding that the black participants did not have a higher percentage of misfit than the white participants also supports the validity of the AMPS for use with both groups.

At issue, however, is the secondary finding that despite (a) the lack of differential item response and (b) the comparable rate of misfit between the two groups, there was a significant difference in the mean ADL process ability measure between the two groups. The mean difference in ADL process ability between the groups was .2 logit, with the black participants having lower overall ADL process ability. Kirkley and Fisher (1999) found, however, that a person's ability measure can vary as much as .5 logit between two different paired task performances, indicating that a clinically meaningful difference may not occur unless the variation is greater than .5 logit. It is likely, therefore, that our observed difference of .2 logit between blacks and whites on the ADL Process scale does not indicate a clinically meaningful difference between the groups.

Nevertheless, we believe it is important to consider other possible explanations for the difference. If the issue was internal test bias in item or task content, we would have found either (a) a differential item or task response between the two groups or (b) a higher rate of misfit in the black sample. As this was not the case, we reasoned that the source of the difference was external and may have stemmed from two possible causes.

First, existing evidence indicated that, overall, blacks have a higher prevalence of disability, including higher rates of chronic disease and mental health problems, than whites (Fischer et al., 1996; Fitzpatrick & Tran, 1997; Mendes de Leon et al., 1995; Mendes de Leon et al., 1997). Even though we matched our samples on the basis of a global rating of functional level, this may not have been enough to control for level of disability. That is, the functional level rating used to match our participants is based on the judgment of the occupational therapist and is an overall, global estimate of a client's ability to function and live safely and independently in the community (i.e., 3 = able to live independently and safely in the community, 2 = requires mini-

<table>
<thead>
<tr>
<th>Measure</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>1.4</td>
<td>1.3</td>
<td>–3.0 to 3.9</td>
<td>0.69</td>
<td>2, 930</td>
<td>.49</td>
</tr>
<tr>
<td>White</td>
<td>1.4</td>
<td>1.2</td>
<td>–3.0 to 4.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Process</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>0.7</td>
<td>0.8</td>
<td>–3.0 to 2.8</td>
<td>3.56</td>
<td>2, 930</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>White</td>
<td>0.9</td>
<td>0.9</td>
<td>–3.0 to 2.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. ADL = activities of daily living.
need to be cognizant of these cultural or ethnic differences. A rater becomes unusually strict when rating a certain rater's performance, if AMPS raters are calibrated for their individual level of severity, if mean process ability between the two groups is that of maximal assistance and more whites who required moderate assistance or supervision to living in the community, 1 = requires moderate to maximal assistance to live in the community. The use of such a global rating scale may have resulted in imprecision in the matching for functional level such that our black sample had more disability, and, therefore, lower mean ADL process ability measures than the corresponding white sample. In the group of participants rated as requiring moderate to maximal assistance, for example, there may have been more blacks who required maximal assistance and more whites who required moderate assistance.

The second possibility to explain the difference in mean process ability between the two groups is that of rater bias. Goto et al. (1996) found that European AMPS raters became more severe when scoring Japanese persons on the ADL Motor and ADL Process scales than they were when scoring European persons. Because AMPS raters are calibrated for their individual level of severity, if a rater becomes unusually strict when rating a certain client, that client’s ability measure will be artificially low. It is possible that a percentage of the black participants were rated by AMPS raters who became somewhat more severe overall when scoring on the ADL Process scale. This increase in severity could have been related to the raters being unfamiliar with culturally appropriate ways of doing ADL tasks within the black subculture, thus scoring these participants inappropriately low on the AMPS Process scale.

Considered together, we concluded that a mean difference in ADL process ability of .2 logit may not reflect a meaningful clinical difference, but even if it does, it likely reflects real differences between our groups that can be associated with level of disability or rater scoring bias and not differential task or item response in the AMPS.

Implications for Practice

The results and conclusions of this study have important implications for occupational therapists working in North America. As the number of minority groups in North America continues to grow, occupational therapists will be treating an increasing number of persons from different cultural or ethnic backgrounds. Occupational therapists will need to be cognizant of these cultural or ethnic differences and consider them when evaluating and treating clients.

The AMPS provides a tool for occupational therapists to assess the effort, efficiency, independence, and safety of a client’s occupational performance in a culturally relevant context. The overall results of this study support the validity of the use of the AMPS to assess black and white populations, provided that the rater is familiar enough with the different ways that persons perform tasks so that he or she does not penalize persons for task performance that is different from the rater’s but appropriate within the person’s cultural and ethnic contexts.

Directions for Future Research

Although the overall results of this study support the use of the AMPS as a valid tool when used to evaluate blacks, there is a need for future research to investigate for the presence of fluctuations in rater severity when raters are scoring clients who are members of different cultural and ethnic groups than that of the rater.

Conclusion

This study provides evidence supporting the continued use of the AMPS with black Americans and white Americans. Our findings support the validity of the AMPS for such use because no differential item response exists in either the AMPS tasks or the AMPS items, and the rate of misfit is comparable between the two groups. Although differences in overall mean ADL process ability between the two groups may not be clinically significant, they may be due to differences in level of disability between the two groups in this study or to fluctuations in rater severity. Raters who are scoring clients of different cultural or ethnic subgroups need to be aware of differences in the ways in which persons belonging to these different subgroups perform ADL tasks and not penalize the person for doing tasks in a manner that is appropriate given his or her cultural or ethnic context. ▲

References


