

Using Video Animation to Assess Mobility in Older Adults

Anthony P. Marsh,¹ Edward H. Ip,^{2,3} Ryan T. Barnard,² Yue-Ling Wong,⁴ and W. Jack Rejeski⁵

¹Department of Health and Exercise Science, ²Department of Biostatistical Sciences, ³Department of Social Sciences and Health Policy, Public Health Sciences, ⁴Department of Computer Science, and ⁵Department of Health and Exercise Science, Wake Forest University, Winston-Salem, North Carolina.

Address correspondence to Anthony P. Marsh, PhD, Department of Health and Exercise Science, Wake Forest University, P.O. Box 7868, Winston-Salem, NC 27109. Email: marshap@wfu.edu

Background. Although the importance of the context of task performance in the assessment of mobility in older adults is generally understood, there is little empirical evidence that demonstrates how sensitive older adults are to subtle changes in task demands. Thus, we developed a novel approach to examine this issue.

Methods. We collected item response data to 81 animated video clips, where various mobility-related tasks were modified in a systematic fashion to manipulate task difficulty.

Results. The participants ($N = 234$), 166 women and 68 men, had an average age of 81.9 years and a variety of comorbidities. Histograms of item responses revealed dramatic and systematic effects on older adults' self-reported ability when varying walking speed, use of a handrail during ascent and descent of stairs, walking at different speeds outdoors over uneven terrain, and carrying an object. For example, there was almost a threefold increase in reporting the inability to walk at the fast speed compared with the slow speed for a minute or less, and twice as many participants reported the inability to walk at the fast speed outdoors over uneven terrain compared with indoors.

Conclusions. The data provide clear evidence that varying the contextual features and demands of a simple task such as stair climbing has a significant impact on older adults' self-reporting of ability related to mobility. More work is needed on the psychometric properties of such assessments and to determine if this methodology has conceptual and clinical relevance in studying mobility disability.

Key Words: Mobility—Aging—Disability—Physical function.

Received February 2, 2010; Accepted October 22, 2010

Decision Editor: Luigi Ferrucci, MD, PhD

THE loss of functional independence with aging has profound effects on older adults' quality of life and results in substantial social and economic burden to society at large (1). As a result, the prevention of disability has emerged as an important topic for gerontology and geriatric medicine. Central to these efforts is the topic of assessment. Although self-report and performance measures of physical disability offer complementary information (2–7), self-perception of abilities is at the heart of disability assessment (8).

Mobility is a critical factor for independence as people age (9,10) and is thus central to quality of life (11). In previous work, we have shown that mobility is related to, but has substantial unique variance from, activities of daily living (ADLs) and instrumental activities of daily living (12). This finding is consistent with the seminal work by Katz (13) and the structure of the International Classification of Functioning, Disability, and Health system (14). In the International Classification of Functioning, Disability, and Health system,

mobility, as used in this manuscript, would be classified as a subdomain under “activities,” that is, the performance of discrete tasks that are important to the performance of social roles and living independently (15).

Disability prevention research has been hindered by measurement-related issues. For example, most assessment tools target difficulty with a very basic set of tasks rather than “ability” across a broad range of tasks that place varying demands on the individual. This leads to ceiling effects when evaluating higher functioning older adults (16,17). The focus on “inability” rather than “ability” has meant that we know very little about the large segment of the older population that is “well functioning,” that is, those who do not report difficulties with basic functioning. An innovative approach that has been used to address this limitation in existing instruments is the use of item response theory (IRT) alone or in conjunction with computer adaptive testing (16,18–21). Nevertheless, most currently used self-report

measures of mobility require participants to make complex judgments about the implicit meaning of tasks. For example, what exactly are the demands of climbing a set of stairs or walking a quarter of a mile? Does not this depend on the physical demands that the stairs may require or the terrain that may be involved in ambulation? These ambiguities add to measurement error. For example, different individuals may respond differently (or similarly) to the same question because of the way they are visualizing themselves doing the task, for example, climbing stairs with versus without a handrail.

To assist in alleviating this form of measurement error, we have developed animated videos as stimuli for assessing mobility and other related ADLs. This methodological approach uses a computer-based program to display video clips constructed from computer animations. After viewing each video clip of the animated task, participants are asked a question about their ability to do the task. Animation serves three purposes: First, compared with video with actual performers, it removes potential biases in judgments that may arise from characteristics, such as the sex, race, age, or body type of the actor. A frail-looking older female actor performing a task may elicit different responses from male respondents than an older able-bodied male actor. Second, it standardizes item interpretation. Respondents view the actual demands of the task and are no longer required to make implicit judgments regarding item content. For example, when asking about climbing a flight of stairs, we present the task standardizing the speed, number of steps, and the presence or absence of handrails. And third, animation enables us to create a broad range of tasks in controlled settings so that we are able to capture the entire range of perceived ability of mobility from walking with a cane to jogging and walking over uneven terrain.

The purpose of the current article is to provide descriptive information about this new assessment method, the mobility assessment tool (MAT), and demonstrate the importance of contextual factors in the self-report assessment of mobility disability. The specific aims are as follows: (a) to illustrate that participants are sensitive to the contextual information provided in video animations of mobility-related tasks, (b) to demonstrate that the nature of the task has a significant effect on ratings of perceived ability, and (c) to document that these methods have the potential to assess older adults across a wide range of function.

METHODS

Participants

The sample for the initial scale development included 234 men and women who lived in independent residences in North and South Carolina. Participants had a mean age of 81.9 years and women represented just over two thirds of the sample. The most common comorbidity was high blood

Table 1. Participant Demographics/Descriptive Information

Characteristic	<i>N</i> (<i>M</i> ± <i>SD</i> or %)
Age (y)	234 (81.9 ± 5.3)
Race	
Black	19 (8.1%)
White	215 (91.9%)
Sex	
Male	68 (29.1%)
Female	166 (70.9%)
Education	
Elementary	26 (11.1%)
High school	52 (22.6%)
College	104 (44.4%)
Postgraduate	44 (18.8%)
Other	7 (3.0%)
Comorbidities	
High blood pressure	153 (65.4%)
Heart disease	76 (32.5%)
Arthritis	85 (36.3%)
Diabetes	28 (12.0%)

pressure followed by arthritis and heart disease (Table 1). The physical health status of the sample has been reported in Rejeski and colleagues (22). The sample had evidence of physical disability, yet exhibited considerable variability in physical function. This is most evident in the 400-m walk test where the average gait speed was 1.04 m/s, yet performance ranged from 0.45 to 1.76 m/s. Also, the range in the Short Physical Performance Battery (4), a test of lower extremity function, was from 2 to 12.

Mobility Task Item Pool

A total of 81 items were used in an IRT-based calibration procedure (23). The polytomous items were calibrated using the graded response model (24), and the binary items were calibrated using the two-parameter logistic model. Marginal maximum likelihood methods were used to estimate the item parameters. A chi-squared statistic was used to evaluate the discrepancy between the observed and estimated probabilities for each item. Multiple testing was controlled for using the Bonferroni procedure with nominal significance level set at 0.05. Overall fit statistics were reported for the following commonly used metrics: Comparative Fit Index (CFI), Tucker–Lewis index (TLI), root mean square error approximation (RMSEA), and standardized root mean square residual (SRMR). Averaged variance explained across the items was also reported. Comparative Fit Index and Tucker–Lewis index values of above 0.9 and RMSEA and SRMR of under 0.08 are considered acceptable fit (25). IRT-based scores were estimated based on the maximum a posteriori. All the IRT-based calibration was conducted using MULTILOG version 7.0 (Scientific Software International, Lincolnwood, IL) (26).

Three items were not used because of their outlying parameter values and content that overlapped with other items. The remaining 78 items comprised seven functional clusters: Cluster A (10 items)—ambulating with a cane as

an assistive device on level ground and up an inclined ramp with and without the use of a handrail; Cluster B (4 items)—walking/jogging on a flat surface at different speeds; Cluster C (5 items)—walking up an inclined ramp at different speeds with and without the use of a handrail; Cluster D (2)—walking while stepping over hurdles at two speeds; Cluster E (6)—walking outdoors on level ground, uneven terrain, and uphill on uneven terrain at different speeds; Cluster F (25 items)—ascending and descending a flight(s) of stairs with different numbers of steps (3, 6, and 9 steps), with and without handrails, at different speeds; and Cluster G (26 items)—ascending a flight of stairs with different numbers of steps while carrying a light or heavy bag in one arm, with and without the use of a handrail or carrying a light or heavy bag in both arms.

For each of the 78 items, we examined the histogram of the responses. There were three types of responses across the item pool. For seven items, participants were asked how long they could do the task shown in the video using the stem, “For how many minutes could you . . . ?” Item responses were none, 1, 5, 10, 15, 20, 25, 30, 40, 50, and 60 minutes. For the second type of response (16 items), participants were asked how many times they could do the task shown in the video using the stem, “How many times, . . . ?” Item responses were none, 1, 2, 3, or 4 times. For the third type of response (58 items), participants were asked could they do the task shown in the video using the stem, “Can you . . . ?” Item responses were yes or no. After examining the distribution of the three types of responses, we collapsed the first two response types into an ordinal scale of four categories. For the first response type, the four categories were none and 1, 5–15, 20–30, and >30 minutes. For the second response type, the four categories were none, 1, 2, and ≥ 3 times. The third response type remained dichotomous, either yes or no.

We also examined the Item Characteristic Curves (ICCs) and the information curves obtained from the MULTILOG procedure. In general terms, an ICC is a graphical representation of the probability of answering an item with a given response at a particular level of ability on the construct being measured. For our purposes, ability represents mobility. The ICC curve gives you a picture of the item’s difficulty and discriminatory power. Difficulty, on an ICC, is defined by the point at which the curve indicates a chance probability of 0.5 (a 50–50 chance) for responding positively to the item. The higher the level of ability needed to obtain a .5 probability (curve shifted to the right), the more difficult the item.

For dichotomized items (yes/no), the ICC depicts the likelihood of a “yes” response to the item across the range of functional abilities (x-axis), in other words, the probability that an individual with a particular ability will report that they can perform the task. Only one ICC curve is shown because the curve for the alternate response, “no,” is one minus ICC. For items that were converted to the four category ordinal

scale (0, 1, 2, or 3), each curve represents the likelihood of an individual’s response to each one of the categories of functional ability for that item. A steeper ICC generally suggests higher discriminating power of the item or item category at the location where the curve has its steepest slope.

The information curve denoted by the dotted line indicates the amount of information contained in each item. Higher information suggests more accurate estimates of ability for a particular item or category. The y range of the information curves reflects the precision of the estimate for the ability score. If only the single item had been administered, an information value of 3 would be equivalent to a standard error of 0.58 (reciprocal of the square root of 3). Information is additive. Therefore, if the cumulative value of information is 30 for a participant responding to 10 items, then the standard error of the ability estimate is 0.18.

In reviewing the histograms, ICCs, and information curves of the 78 items, we have elected to present data focused on several key contextual factors. These include the speed of performance, characteristics of the terrain, number of steps in climbing stairs, the use of a handrail when climbing stairs, and carrying objects upstairs. The complete report of histograms, ICCs, and information curves is available from the corresponding author.

RESULTS

IRT Model

At the item level, the goodness of fit was assessed in terms of the chi-square statistic. No item was found to be significant, suggesting that the IRT model fitted the data well at the item level. The overall model fit statistics for the unidimensional model were as follows: Comparative Fit Index = 0.94, Tucker–Lewis index = 0.96, RMSEA = 0.18, and SRMR = 0.16. Averaged variance explained across items is equal 0.77. We noticed that although the Comparative Fit Index and Tucker–Lewis index both suggested reasonably good fit, the other two indices—the RMSEA and SRMR—did not appear to support the unidimensionality assumption. The poor fit indexes in RMSEA and SRMR could be an artifact of the relatively large number of items present in the item bank. The discord between fit indices deserves further study because it is plausible that a multidimensional model may provide a better fit for the items. However, based on our conceptual framework and patterns observed in the current data set, we suspect that continued research with the item pool will support the existence of mobility as the core clinically relevant dimension and one or more minor dimensions that are “nuisance” factors. Further investigation into the psychometric properties (eg, bifactor analysis and differential item functioning), as suggested in Reeve and colleagues (27), is planned for a separate report.

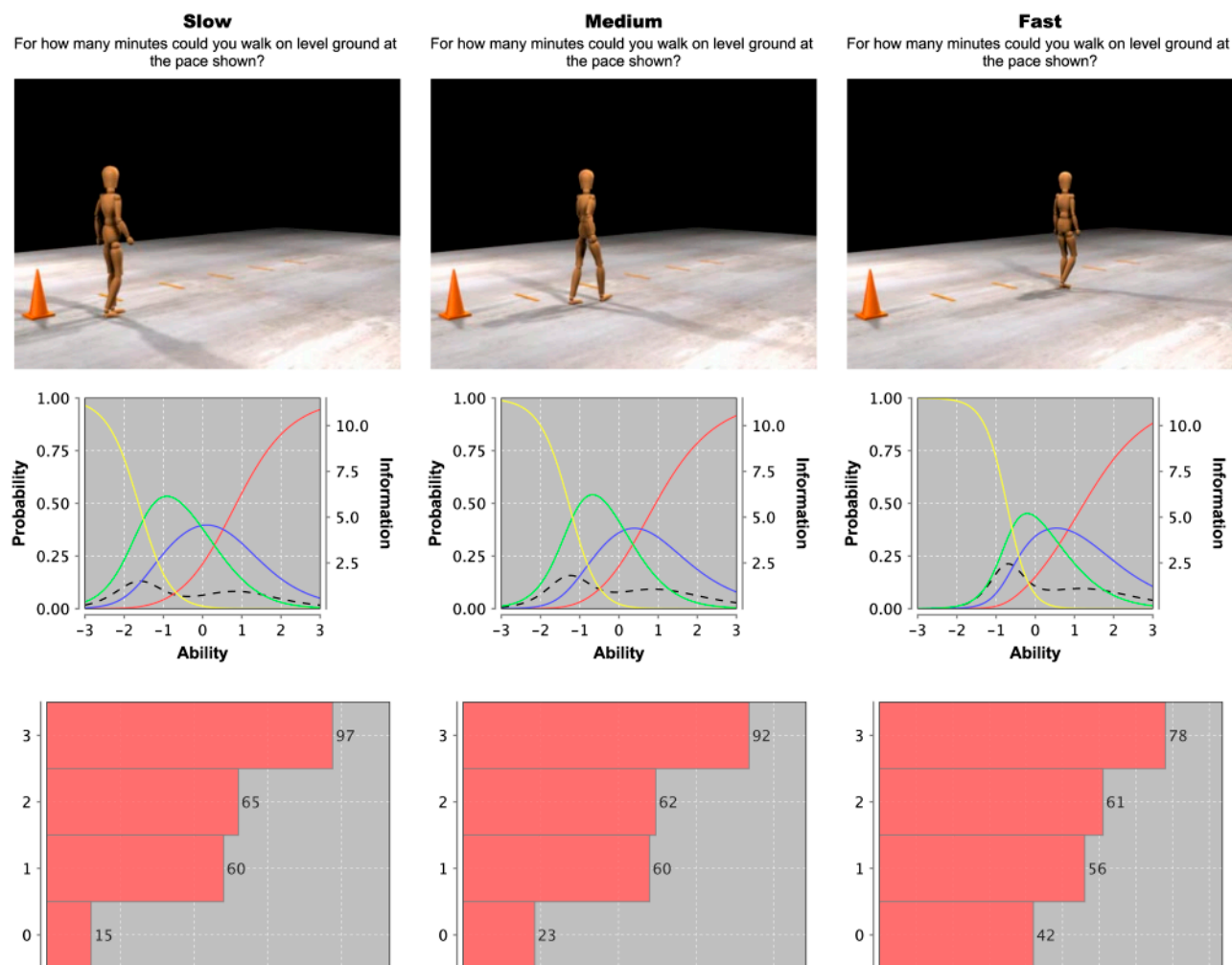


Figure 1. Screenshot, Item Characteristic Curve, information curve, and histogram of responses for the item asking “For how many minutes could you walk on level ground at the pace shown?” Responses to the four categories (0, 1, 2, and 3, respectively, in the histogram) are none and 1 (yellow line), 5–15 (green line), 20–30 (blue line), and >30 minutes (red line). The four lines represent the probability of each response across the range of function and the y-axis is on the left. The dotted line represents the information curve, and the y-axis on the right provides the scale. The image from the video shows the position of the mannequin at the same instant in time, reflecting the differences in walking speed (slow, medium, and fast).

MAT Testing and Sample Characteristics

Participants required between 20 and 30 minutes to complete the entire battery of 81 items. We collected cognitive function data on a subsample of 112 participants using the Mini-Mental State Examination. The participants had an average Mini-Mental State Examination score of 27.9 ± 2.3 (range: 21–30). The gait speed of the participants during a 400-m walk test was 1.04 ± 0.27 m/s, and they had an average Short Physical Performance Battery score of 8.61 ± 2.70 . The items were well received by all participants. All participants found the test to be interesting, the touch screen computer technology to be user-friendly, and remarked in focus groups that the animations were extremely helpful in communicating the physical demands of each task. Anecdotally, some participants commented that they projected themselves onto the figure suggesting that the neutrality of the mannequin was beneficial. A number of participants expressed concern about response burden and possible

dulling effects as the same task is often repeated multiple times over different speeds and modes. They expressed great interest in a version that is more adaptive so they would not have to respond to the entire item pool. In response to this concern, we have recently developed a short-form version of the MAT (MAT-sf) (22).

Speed of Walking and Terrain

Figure 1 presents a screen shot, ICC, and histogram of responses for walking on level ground at three speeds (0.6, 1.0, and 1.3 m/s). For these three items, the participant was asked “For how many minutes could you walk on level ground at the pace shown?” Recall that for this item responses was collapsed to a 4-point (0, 1, 2, and 3) ordinal scale (none and 1, 5–15, 20–30, and >30 minutes, respectively). The histogram indicates that, as speed increases, there was a shift in responses; specifically, participants indicated

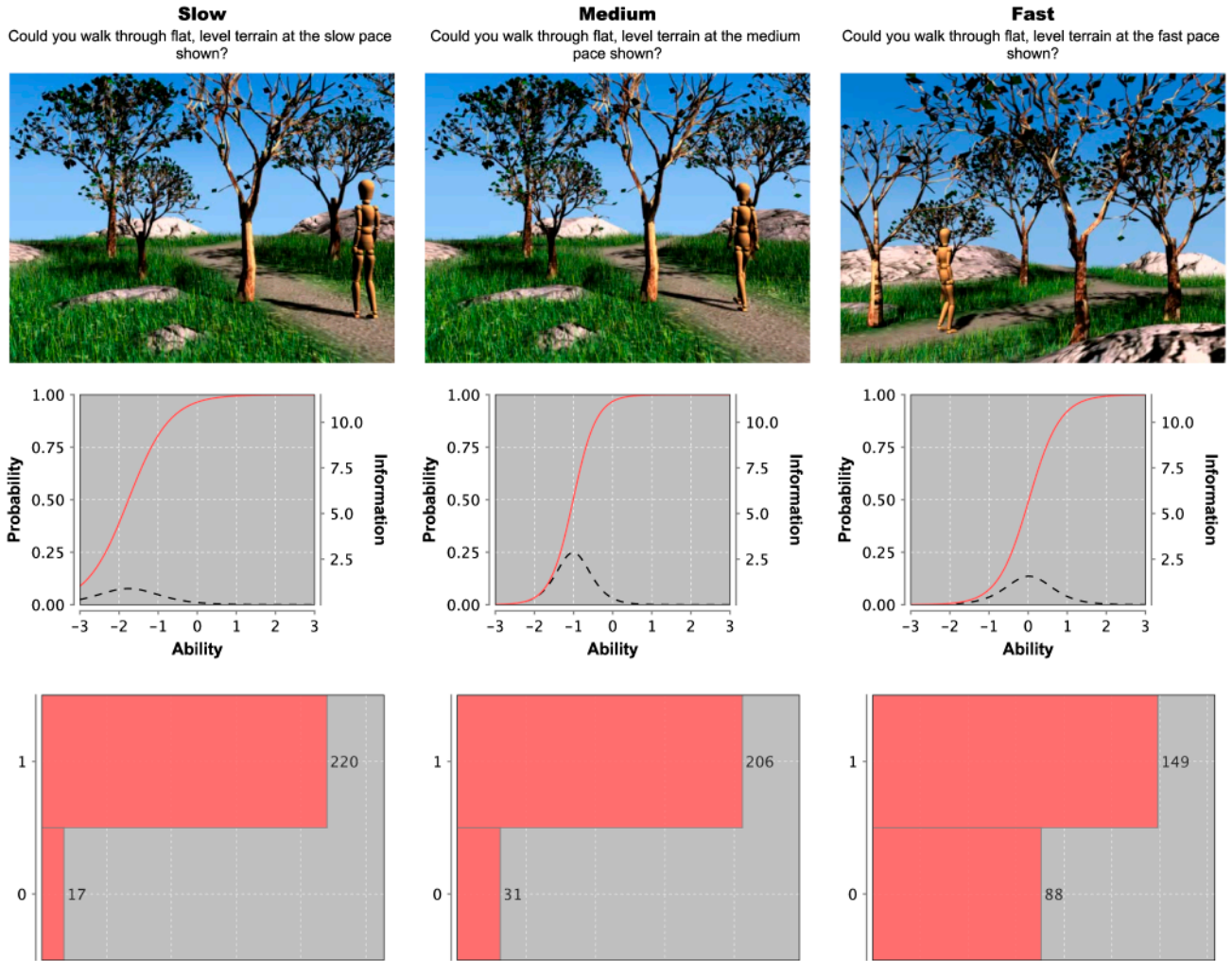


Figure 2. Screenshot, item characteristic curve (ICC), information curve, and histogram of responses for the item asking “Could you walk through flat, level terrain at the slow/medium/fast pace shown in the movie?” Responses are yes = 1 and no = 0. For the ICC, the red line represents the probability of being able to do the task across the range of function and the y-axis is on the left. The dotted line represents the information curve, and the y-axis on the right provides the scale.

that with increasing speed, the duration of the time they could walk decreased.

To aid in the interpretation of the ICC graphs, we provide the following explanation for the slow walking speed and the accompanying ICC shown in Figure 1. First, each ICC curve represents the probability of a response to each of the four categories, that is, 0 (yellow line), 1 (green line), 2 (blue line), and 3 (red line), across the ability level continuum, -3 to 3. Where curves intersect, it is equally likely for a participant to respond to either of the two levels under consideration. For example, an older adult with an ability level near -1.6 is equally likely to answer 0 or 1 on this item. An older adult with an ability level near 0.6 is equally likely to report a score of 2 or 3. An older adult with an ability level of about -0.7 is equally likely to report a score of 0 or 3 (the intersection of the yellow and red lines), but they most likely will report a score of 1 (green line) because, at this ability level, that response has the highest probability of

occurring (probability of ~0.55). Finally, consider the two extremes. An older adult with very low ability is almost certain to report a score of 0, that is, that they cannot walk at the slow speed for more than 1 minute. Conversely, a person with very high ability has a high probability of reporting a score of 4, that is, that they can walk at the pace shown for more than 30 minutes. Because a person can only give one response to the item and the four options are mutually exclusive, the sum of the three probabilities at each level of ability is one.

Figure 2 illustrates the influence of outdoor terrain on the self-perceived ability to walk at slow, medium, and fast speeds. For these three items, the participant was asked “Could you walk on level terrain at the slow/medium/fast pace shown in the video clip?” The participant gave a yes/no response. As speed increased (0.6, 1.0, and 1.3 m/s), the number of participants who reported an inability to do the task increased fivefold. Of course, the interpretation of

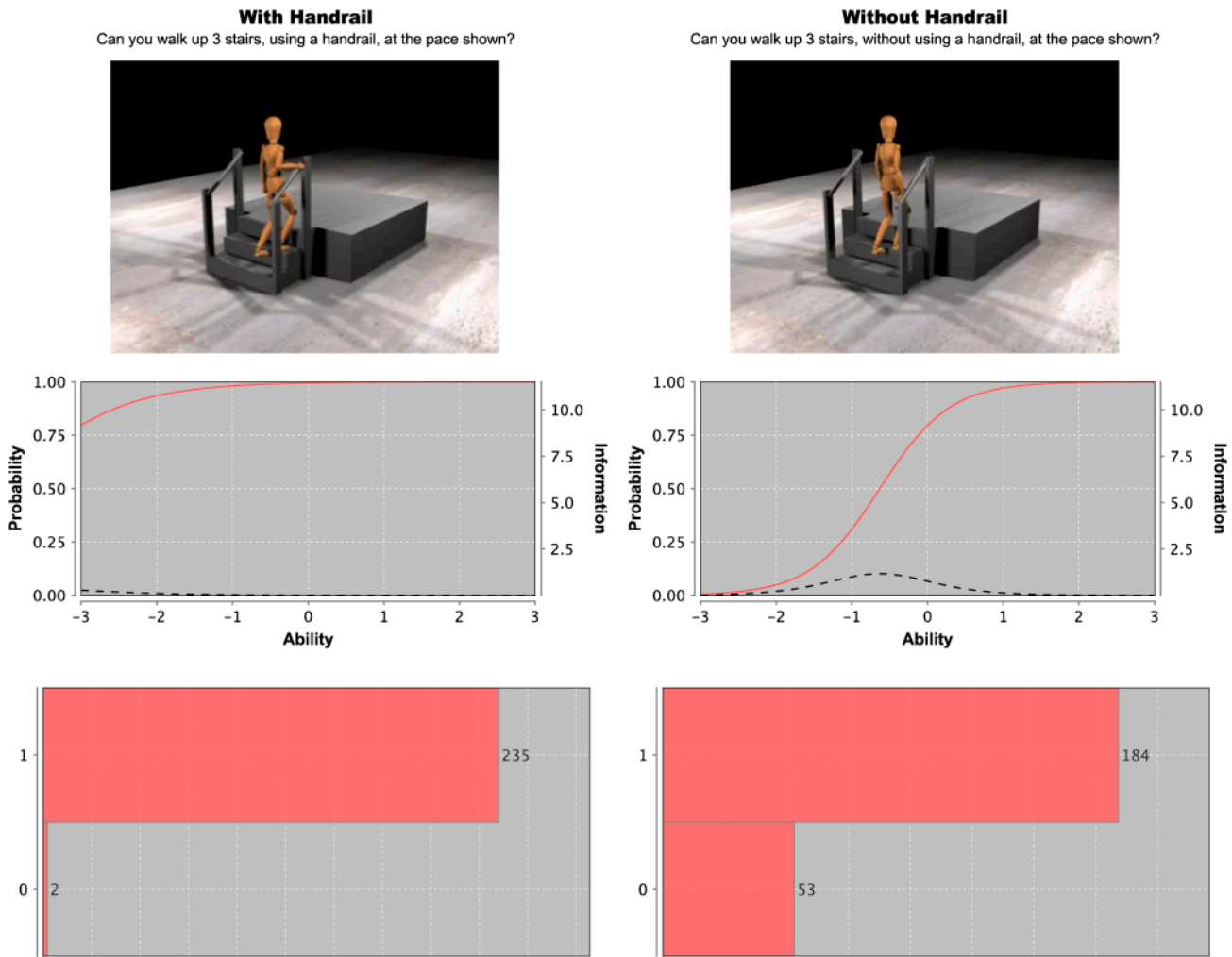


Figure 3. Screenshot, Item Characteristic Curve (ICC), and histogram of responses for the item asking “Can you walk up 3 stairs, using/without using a handrail, at the pace shown?” Responses are yes = 1 and no = 0. For the ICC, the red line represents the probability of being able to do the task across the range of function and the y-axis is on the left. The dotted line represents the information curve, and the y-axis on the right provides the scale.

the ICC graph is simplified for those items with a yes/no response. At the slow walking speed, the ICC indicates that individuals in the top third of ability (>ability 1) almost certainly will respond that they can do the task shown in the video.

Use of a Handrail and Carrying a Bag

Figure 3 presents the influence of handrail use while ascending three steps at a slow pace one step at a time. For these three items, the participant was asked “Can you walk up 3 stairs, using/without using a handrail, at the pace shown?” Clearly handrail use had a profound effect on the responses. Figure 4 also shows the considerable influence that the use of a handrail has on self-reported ability to ascend three steps while carrying a bag.

Number of Steps and Ascending/Descending Stairs

Figures 3 and 5 allow a comparison of responses when the number of steps is increased from three to six as well as

when a handrail is available. The increased difficulty of the task is reflected in the responses and the ICC with more participants reporting that they were unable to ascend six steps; the ICC and information curves shifted to the right. Figure 6 underscores the added difficulty older adults have with stair descent even when using a handrail.

DISCUSSION

The specific aims of this study were (a) to demonstrate that older adults are sensitive to the contextual information provided in video of animated presentations of mobility-related tasks, (b) to demonstrate that the nature of the task has a significant effect on ratings of perceived ability, and (c) to document that these methods have the potential to assess older adults across a wide range of mobility function. We present a unique method for the assessment of mobility in older adults that uses short video clips of animated mobility tasks combined with questions that ask about

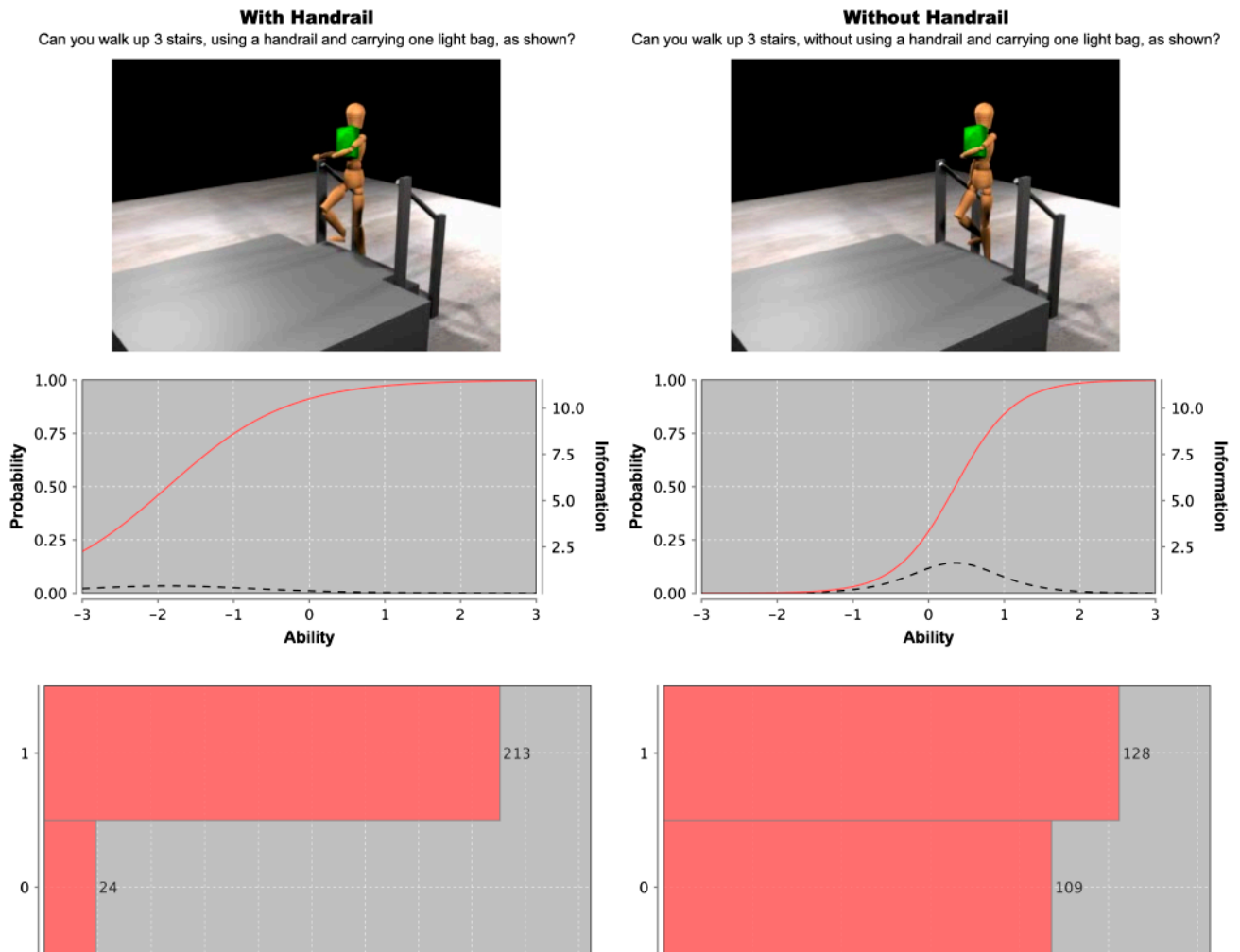


Figure 4. Screenshot, Item Characteristic Curve (ICC), and histogram of responses for the item asking “Can you walk up 3 stairs, using/without using a handrail and carrying one light bag, as shown?” Responses are yes = 1 and no = 0. For the ICC, the red line represents the probability of being able to do the task across the range of function and the y-axis is on the left. The dotted line represents the information curve, and the y-axis on the right provides the scale.

participants’ perceived ability to do the task. The data provide clear evidence that the contextual features and demands of a simple task such as stair climbing have a significant impact on older adults’ self-reporting of ability related to mobility.

Gait speed is becoming more recognized as a critically important capacity in older adults (28–32). However, at this time, gait speed is not characterized in any of the widely used self-report instruments of mobility. For example, the most commonly asked question is “Do you have difficulty walking a quarter of a mile or a city block?” Our data show that changes in gait speed from 0.6 to 1.0 and 1.3 m/s have a significant and systematic impact on older adult’s perception of their ability to walk for a given duration. There is an almost threefold increase in people reporting the inability to walk at the fast speed compared with the slow speed for a minute or less. Another interesting result is the shift of the ICC to the right and the increase in information provided at the upper end of the functional spectrum as the walking speed increases.

The influence of gait speed on task difficulty was magnified when the terrain was changed to an outdoor dirt path. As speed increased, there was a fivefold increase in the number of participants who reported an inability to do the task. As we observed for walking on level ground, the ICC shifted to the right, and the information curve showed that more information was provided for higher functioning participants as speed increased. Of particular note, twice as many participants reported the inability to walk at the fast speed over outdoor terrain (Figure 2) compared with level ground (Figure 1). In order to provide greater contextual specificity, we are currently developing animations that show walking in a residential community setting and an urban city setting.

The use of a handrail has a major impact on an older adult’s perception of their ability to climb stairs with more than 22% of the participants reporting that they did not believe they could ascend just three steps without the stability provided by a handrail. Comparing Figures 3 and 4, twice

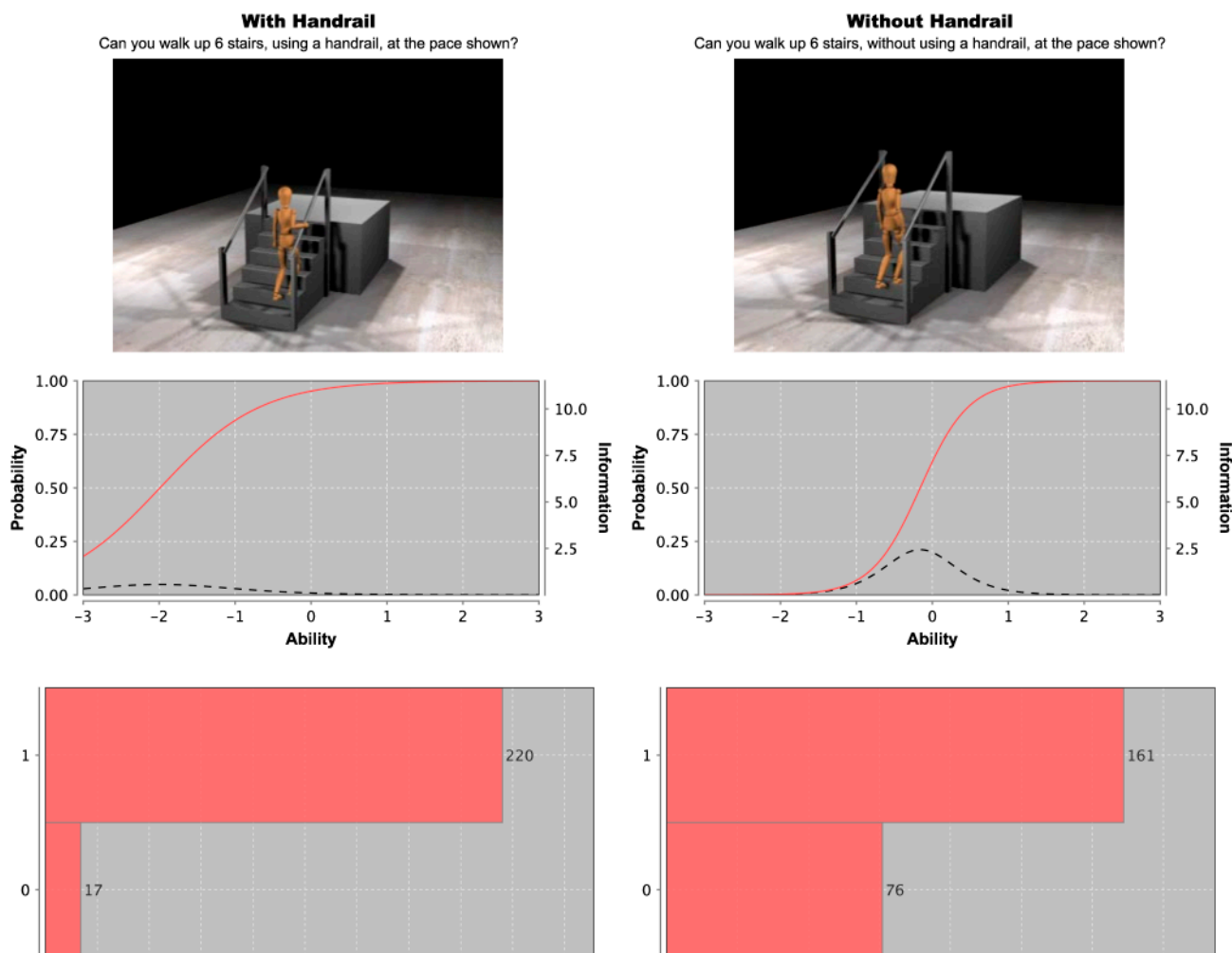


Figure 5. Screenshot, Item Characteristic Curve (ICC), and histogram of responses for the item asking “Can you walk up 6 stairs, using/without using a handrail, at the pace shown? Responses are yes = 1 and no = 0. For the ICC, the red line represents the probability of being able to do the task across the range of function and the y-axis is on the left. The dotted line represents the information curve, and the y-axis on the right provides the scale.

as many participants ($n = 53$ vs. 24) reported the inability to ascend the three stairs when handrail use was not allowed than the number who reported the ability to ascend the stairs carrying a light bag with handrail use. This result underscores the critical role that handrails have on self-perceived function and the high importance older adults place on maintaining stability in challenging situations, perhaps as a result of fear of falling (33,34).

The difficulty that older adults often report with stair descent is highlighted by the comparison of Figures 5 and 6. There was almost a fourfold increase in the number of participants who reported an inability to descend compared with ascend six steps even with using a handrail. Almost half the participants did not believe they could descend six steps without a handrail. The marked changes in the responses as a result of changes in handrail use and the number of steps in the flight of stairs reveal the limitations of a question that simply asks “Do you have difficulty climbing up/down stairs?” Interestingly, Verghese and colleagues

(35) recently reported that in 310 older men and women (mean age = 79.7 years, 62% women), 140 reported difficulties in climbing upstairs and 83 in climbing downstairs. They reported that self-reported difficulty climbing down stairs captured a wider spectrum of ADL limitations than climbing upstairs. However, individuals who reported difficulty in both ascending and descending the stairs were far more likely to report ADL limitations than those who reported no difficulty either ascending or descending stairs (odds ratio = 6.58; 95% confidence interval = 3.35–12.91). In discussing limitations of their work, they stated “The number of steps or flights of stairs was not specified in our questionnaire, because our intention was to get our subjects’ global impression of task difficulty. Also, these architectural elements may vary in our urban subjects with different housing arrangements.” Our data suggest that specifying the contextual factors of a task as simple as ascending/descending stairs is important in assessing the difficulty of the task.

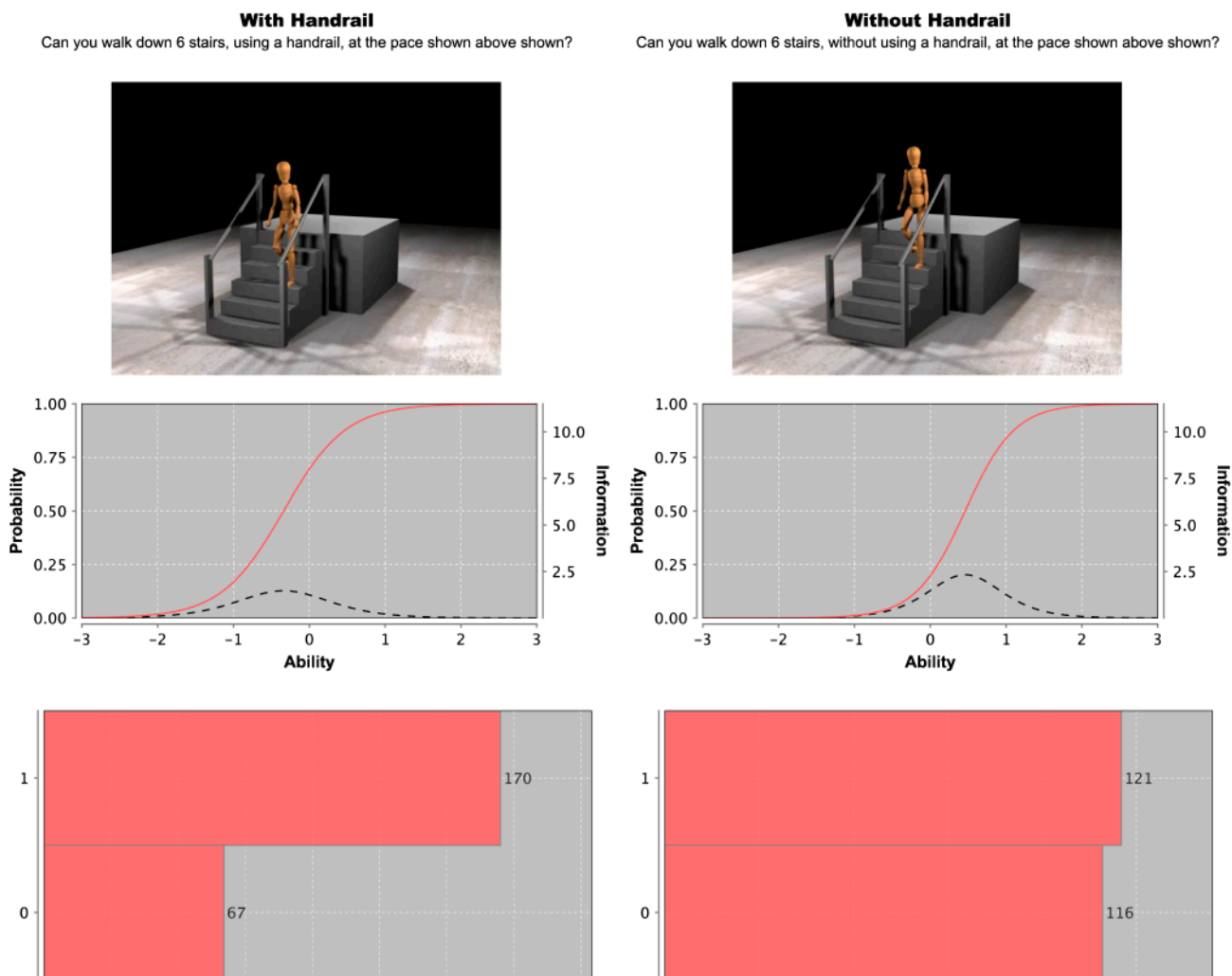


Figure 6. Screenshot, Item Characteristic Curve (ICC), and histogram of responses for the item asking “Can you walk down 6 stairs, using/without using a handrail, at the pace shown? Responses are yes = 1 and no = 0. For the ICC, the red line represents the probability of being able to do the task across the range of function and the y-axis is on the left. The dotted line represents the Information curve, and the y-axis on the right provides the scale.

The changes in the information curves as the difficulty of the tasks increases are important because they provide a way to ascertain the sensitivity of each question as ability levels change. A major limitation of traditional measures of self-reported physical function is that they were not designed to capture the entire range of function in older adults (7,36,37). In the current method of assessment, the full set of 78 items was specifically designed to assess as wide a spectrum of mobility function as possible. In fact, one of the advantages of this method is that it can be used to assess physical function in older adults with a variety of limitations in physical functioning. For example, there are animated clips that evaluate physical functioning in older adults who must use a cane to walk. The full set includes items that range from walking slowly with a cane to jogging, climbing three steps slowly one at a time to climbing nine steps quickly while carrying two bags, and walking slowly on a path to walking up a rocky inclined path. However, using all

78 items may present a burden, particularly in large clinical trials or observational studies where there are a large number of assessments. In these instances, a computer adaptive testing approach would be preferred. We are in the process of developing such an instrument.

Another approach that has been used to increase the sensitivity of self-report assessments of mobility is to modify items for existing self-reported mobility measures such that if older adults report no difficulty for specific items, then they are asked how easy it is to perform the task in question and to include more demanding levels of these tasks (7). This is a creative methodological approach, but it still has limitations. For example, walking ¼ mile may be easy for an older person when done at a slow pace, yet difficult if attempted at a brisk pace. Furthermore, the words “fast” and “slow” are relative concepts and, as function declines, surfaces and inclines become important contextual features of the ability to perform mobility-related tasks.

Multimedia and other interactive features have become a hallmark of modern assessment technology. In pilot testing, we found that touch screen technology dramatically decreased the time of test taking and obviated the need for a mouse, the use of which was a major challenge for some older adults. Animation technology allows great flexibility in altering the form, speed, and environmental parameters for mobility-related tasks and avoids confounds associated with using human actors. Although we have not done so here, it is a relatively simple process to alter the morphological characteristics of the mannequin to create changes in sex, race, and body mass index. This could provide an interesting method for examining how morphological characteristics bias cultural judgments about disability. Finally, animated clips are less expensive to produce than movie clips that use human actors.

In this study, we showed that the contextual features of the animations had a significant impact on older adults self-reporting of ability related to mobility. This is consistent with an earlier report by Haley and colleagues on the importance of contextual features in the assessment of physical disability in children (38) and the work of Shumway-Cook and colleagues on the importance of environmental factors that influence mobility disability in older adults (39–41). Assessments in research or clinical settings must strike a balance between being standardized and being relevant in a real-world environment. This presents some challenges for performance-based assessments of function because it is difficult to conduct assessments of walking over different terrain, upstairs and downstairs with varying numbers of steps, or walking within crowds of people. We believe the MAT offers great potential to make the assessment of physical function in older adults more contextually relevant.

The reliability, validity, and sensitivity to change of scores on the MAT require study. In this regard, a recent report of ours on a MAT-sf is encouraging (22). In this study, the association between the 78-item instrument and the MAT-sf was $r = .96, p < .001$. We conducted a 2-week test–retest reliability coefficient for the MAT-sf and found that the measure had high reliability, $ICC = 0.93, p < .0001$. To examine convergent validity, we computed bivariate correlations between the MAT-sf and a validated self-report measure of disability, the Pepper Assessment Tool for Disability (PAT-D). Our hypothesis was that of the three PAT-D subscales—ADL, mobility, and instrumental activity of daily living—the strongest relationship would exist between the PAT-D mobility subscale and the MAT-sf, whereas the weakest relationship would be found for the instrumental activity of daily living subscale. Supporting this hypothesis, the correlations of the MAT-sf with the mobility, ADL, and instrumental activity of daily living subscales of the PAT-D were $-.60, -.50, \text{ and } -.44$, respectively; all correlations were significant ($p < .001$). Furthermore, stepwise regression analyses showed that the MAT-sf contributed over and above the PAT-D in predicting scores on two tests

of lower extremity function, the Short Physical Performance Battery and the 400-m walk test. We caution the reader that these analyses were conducted on the 10-item MAT-sf; however, they do offer some preliminary support for the utility of the approach we have used to assess mobility in older adults.

In conclusion, these data provide compelling evidence that contextual factors have a significant impact of an older adult's perceived ability to perform a wide range of mobility-related tasks. More work needs to be done to determine if the use of animated videos in assessments of mobility and related activities leads to improved measurement precision and whether this methodology has utility in clinical research.

FUNDING

Support for this study was provided by (a) National Institutes for Aging P30 AG021332, (b) National Heart, Lung, and Blood Institute grant HL076441-01A1, (c) General Clinical Research Center grant, 5M01RR007122-18, and (d) the Wake Forest University Translational Science Institute.

REFERENCES

- Goulding MR, Rogers ME, Smith SM. Trends in aging—United States and worldwide. *MMWR Morb Mortal Wkly Rep.* 2003;52:101–104, 106.
- Myers AM, Holliday PJ, Harvey KA, Hutchinson KS. Functional performance measures: are they superior to self-assessments? *J Gerontol.* 1993;48:M196–M206.
- Reuben DB, Valle LA, Hays RD, Siu AL. Measuring physical function in community-dwelling older persons: a comparison of self-administered, interviewer-administered, and performance-based measures. *J Am Geriatr Soc.* 1995;43:17–23.
- Guralnik JM, Simonsick EM, Ferrucci L, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol.* 1994;49:M85–M94.
- Reuben DB, Siu AL, Kimpau S. The predictive validity of self-report and performance-based measures of function and health. *J Gerontol.* 1992;47:M106–M110.
- Reuben DB, Seeman TE, Keeler E, et al. Refining the categorization of physical functional status: the added value of combining self-reported and performance-based measures. *J Gerontol A Biol Sci Med Sci.* 2004;59:1056–1061.
- Simonsick EM, Newman AB, Nevitt MC, et al. Measuring higher level physical function in well-functioning older adults: expanding familiar approaches in the Health ABC study. *J Gerontol A Biol Sci Med Sci.* 2001;56:M644–M649.
- Iezzoni LI, McCarthy EP, Davis RB, Siebens H. Mobility problems and perceptions of disability by self-respondents and proxy respondents. *Med Care.* 2000;38:1051–1057.
- Guralnik JM, Ferrucci L, Balfour JL, Volpato S, Di IA. Progressive versus catastrophic loss of the ability to walk: implications for the prevention of mobility loss. *J Am Geriatr Soc.* 2001;49:1463–1470.
- Hadley EC. Testing interventions to preserve walking ability: progress against disability, one step at a time. *J Gerontol A Biol Sci Med Sci.* 2007;62:834–836.
- Rejeski WJ, Mihalko S. Physical activity and quality of life in older adults. *J Gerontol.* 2001;56A:23–35.
- Rejeski WJ, Ip EH, Marsh AP, Miller ME, Farmer DF. Measuring disability in older adults: the International Classification System of Functioning, Disability and Health (ICF) framework. *Geriatr Gerontol Int.* 2008;8:48–54.

13. Katz S. Assessing self-maintenance: activities of daily living, mobility, and instrumental activities of daily living. *J Am Geriatr Soc.* 1983;31:721–727.
14. World Health Organization. *International Classification of Functioning, Disability, and Health (ICF)*; Geneva, Switzerland: WHO; 2001.
15. Jette AM. Toward a common language for function, disability, and health. *Phys Ther.* 2006;86:726–734.
16. Jette AM, Haley SM. Contemporary measurement techniques for rehabilitation outcomes assessment. *J Rehabil Med.* 2005;37:339–345.
17. Ware JE Jr. Conceptualization and measurement of health-related quality of life: comments on an evolving field. *Arch Phys Med Rehabil.* 2003;84:S43–S51.
18. Fries JF, Bruce B, Bjorner J, Rose M. More relevant, precise, and efficient items for assessment of physical function and disability: moving beyond the classic instruments. *Ann Rheum Dis.* 2006;65(suppl 3):iii16–iii21.
19. Bruce B, Fries JF, Ambrosini D, et al. Better assessment of physical function: item improvement is neglected but essential. *Arthritis Res Ther.* 2009;11:R191.
20. Haley SM, Ni P, Hambleton RK, Slavin MD, Jette AM. Computer adaptive testing improved accuracy and precision of scores over random item selection in a physical functioning item bank. *J Clin Epidemiol.* 2006;59:1174–1182.
21. Jette AM, Haley SM, Ni P, Olarsch S, Moed R. Creating a computer adaptive test version of the late-life function and disability instrument. *J Gerontol A Biol Sci Med Sci.* 2008;63:1246–1256.
22. Rejeski WJ, Ip EH, Marsh AP, Barnard RT. Development and validation of a video-animated tool for assessing mobility. *J Gerontol A Biol Sci Med Sci.* 2010;65:664–671.
23. Lord FM. *Applications of Item Response Theory to Practical Testing Problems.* Mahwah, NJ: Lawrence Erlbaum Associates; 1980.
24. Samejima F. *Estimation of Ability Using a Response Pattern of Graded Scores.* Psychometrika Monograph No. 17, 1969.
25. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling.* 1993;6:1–55.
26. Thissen D, Chen W-H, Bock RD. *Multilog (v. 7).* Lincolnwood, IL: Scientific Software International, 2003.
27. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care.* 2007;45:S22–S31.
28. Guralnik JM, Ferrucci L, Pieper CF, et al. Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. *J Gerontol A Med Sci.* 2000;55:M221–M231.
29. Studenski S. Bradypedia: is gait speed ready for clinical use? *J Nutr Health Aging.* 2009;13:878–880.
30. Studenski S, Perera S, Wallace D, et al. Physical performance measures in the clinical setting. *J Am Geriatr Soc.* 2003;51:314–322.
31. Kuo HK, Leveille SG, Yu YH, Milberg WP. Cognitive function, habitual gait speed, and late-life disability in the National Health and Nutrition Examination Survey (NHANES) 1999–2002. *Gerontology.* 2007;53:102–110.
32. The LIFE Study Investigators. Effects of a physical activity intervention on measures of physical performance: results of the Lifestyle Interventions and Independence for Elders Pilot (LIFE-P) Study. *J Gerontol A Biol Sci Med Sci.* 2006;61:1157–1165.
33. Maki BE, Holliday PJ, Topper AK. Fear of falling and postural performance in the elderly. *J Gerontol.* 1991;46:M123–M131.
34. Startzell JK, Owens DA, Mulfinger LM, Cavanagh PR. Stair negotiation in older people: a review. *J Am Geriatr Soc.* 2000;48:567–580.
35. Verghese J, Wang C, Xue X, Holtzer R. Self-reported difficulty in climbing up or down stairs in nondisabled elderly. *Arch Phys Med Rehabil.* 2008;89:100–104.
36. Guralnik JM, Simonsick EM. Physical disability in older Americans. *J Gerontol.* 1993;48(Spec No):3–10.
37. Simonsick EM, Montgomery PS, Newman AB, Bauer DC, Harris T. Measuring fitness in healthy older adults: the Health ABC long distance corridor walk. *J Am Geriatr Soc.* 2001;49:1544–1548.
38. Haley SM, Coster WJ, Binda-Sundberg K. Measuring physical disablement: the contextual challenge. *Phys Ther.* 1994;74:443–451.
39. Shumway-Cook A, Patla A, Stewart A, Ferrucci L, Ciol MA, Guralnik JM. Environmental components of mobility disability in community-living older persons. *J Am Geriatr Soc.* 2003;51:393–398.
40. Shumway-Cook A, Patla AE, Stewart A, Ferrucci L, Ciol MA, Guralnik JM. Environmental demands associated with community mobility in older adults with and without mobility disabilities. *Phys Ther.* 2002;82:670–681.
41. Shumway-Cook A, Patla A, Stewart AL, Ferrucci L, Ciol MA, Guralnik JM. Assessing environmentally determined mobility disability: self-report versus observed community mobility. *J Am Geriatr Soc.* 2005;53:700–704.