# 14. CONCEPTUALIZING AND DEVELOPING THE AMPS WITHIN A FRAMEWORK OF MODERN OBJECTIVE MEASUREMENT

by

Anne G. Fisher

Brenda K. Merritt

## 14.1 Conceptualizing and Developing the AMPS

The AMPS was developed in response to a critical need for the development of a valid and reliable performance-based assessment that (a) was client-centered and allowed persons to choose familiar and relevant ADL tasks to perform for the assessment, (b) generated linear measures of the quality of a person's ADL task performance, and (c) provided information about the quality of a person's ADL task performance that could be used to guide the process of planning intervention. Such goals required the use of a measurement model that allowed us to calibrate and equate ADL tasks for their challenge, simultaneously discover the hierarchical order of ADL skill item difficulty and ADL task challenge, and verify that the ADL skill items and tasks worked together to define unidimensional constructs (Fisher, 1993, 1994, 1997; Fisher, Bryze, Granger, Haley, Hamilton, Heinemann, Puderbaugh, Linacre, Ludlow, Mc Cabe, & Wright, 1994).

During the initial development of an objective assessment tool like the AMPS, great care was taken to first develop a theoretical model of the construct to be measured — ADL ability or, more specifically, the ***quality*** of a person's ADL task performance (Bernspång & Fisher 1995b; Bond & Fox, 2007; Fisher, 1993; W. P. Fisher, 2002; Wright & Stone, 1979). To do this, we needed to conceptualize a line that represented a continuum of ADL ability, and determine what it would look like to have more or less of this trait. We began this process by conceptualizing a unidimensional continuum that represented the quality of a person's ADL ability. First, we envisioned well adults who had very high ADL ability, and would be able to ***competently*** complete all desired self-care and home maintenance tasks, including complex meal preparation and heavy house or yard work. We then envisioned people with very little ADL ability, including very young children and persons with severe disabilities. These were people who were just learning (or re-learning) how to perform ***very simple*** ADL tasks (e.g., brushing teeth, eating, pour a glass of milk). These persons may even require extra time, supervision, and/or assistance when performing very simple ADL tasks.

Once we had envisioned the two extremes of our theoretical continuum, we quickly realized that we were dealing with a wide variation in ADL ability, ranging from a person who is just beginning or re-learning how to perform very basic self-care tasks, all the way up to a healthy, well adult who is competently performing all simple and complex ADL tasks that are necessary for community dwelling.

Our next step was to identify a set of ADL tasks that would fall along the continuum of relative challenge. We theorized that if very able persons performed very easy tasks (i.e., tasks that were too easy), or if persons who were severely disabled performed very challenging tasks (i.e., tasks that were too hard), the resulting ADL motor and ADL process ability measures would have limited utility because floor or ceiling effects would likely result. We concluded, therefore, that we needed to develop ADL tasks that varied in challenge, ranging from very basic to very complex, such that they offered a challenge to persons with progressively greater ADL ability. In addition, we developed the following criteria for each ADL task included in the AMPS:

- ***Each task included in the AMPS must be an ADL task*** — the construct we aimed to measure was ADL performance, not work, play, schoolwork, or other occupational domains. If we tried to measure more than one construct, the AMPS would become multidimensional, no longer unidimensional. Sound measurement models require that each scale be unidimensional. The AMPS now includes two unidimensional scales, one for ADL motor ability and one for ADL process ability.

- ***Each task included in the AMPS must be able to be performed within a relatively short period of time*** — commonly, time constraints within occupational therapy practice severely limit the time most occupational therapists have available for assessment. If the AMPS is to be practical, it must be able to be administered within a reasonable time frame. Thus, each ADL task performance should not take more than 10 to 15 minutes, ideally even less.

Over time, occupational therapists have requested that we add some ADL tasks that may take longer to perform. We continue, however, to avoid including ADL tasks that would take so long to perform that occupational therapists would not be able to use them in the context of everyday practice.

- ***Each task included in the AMPS must be a task that is culturally relevant and commonly performed, ideally by people living in more than one major world region*** — when a task is relevant to many persons, it will more likely be useful to occupational therapists in many different world regions. While this would enable us to develop an international evaluation of ADL ability, ***we recognized that it was not possible for all tasks to be relevant to all world regions or cultures***.

If we were to have a very sensitive tool, we also needed to develop items that represented smaller units of ADL task performance. We recognized that many of the common ADL assessment tools are utilized to gather global information related only to how much assistance a person needs when performing ADL tasks. In contrast, we wanted to develop a measure of the quality of ADL task performance that enabled occupational therapists to evaluate the physical effort, efficiency, safety, and/or need for assistance a person demonstrated when performing ADL tasks. This required the development of a set of items that reflected skilled actions that, when performed, would make up an ADL task.[9] We developed the following criteria for each skilled action (ADL skill item) included in the AMPS. Each ADL skill item must:

- Represent an observable action of ADL task performance, ***not*** an underlying body function

---

[9] Another way to make a test of ADL more sensitive is to break each task down into small parts, and allow those "parts" to become items. For example, putting on one's socks and shoes could be broken down into (a) putting the sock over the toes, (b) pulling the sock over the heel, (c) pulling the sock all the way up, (d) putting the foot into the shoe, (e) pulling the shoe laces tight, and (f) tying the shoe laces. Obviously, this set of items is not universal — none can be observed in ADL tasks not related to putting on shoes and socks. Such an approach would also require that each potential task be broken down into similar steps, and the result would be the demand to create scoring criteria and then standardize hundreds of items. Our approach enabled us to develop only 36 sets of scoring criteria (actually only 35, as Paces is included on both the AMPS motor scale and AMPS process scale), while at the same time, enabling us to create an item bank that now includes 1856 ADL motor items and 2320 ADL process items (see Section 14.1.2).

- Be observable in virtually any ADL task performance (i.e., it must be "universal")
- Be "neutral" or free of bias associated with gender, culture, or world region
- Be able to be conceptualized as being performed more or less skillfully (i.e., some people perform the action in a manner that is qualitatively better than others) and scored using a rating scale

After developing guidelines for selecting ADL tasks and ADL skill items to include in the AMPS, we needed to choose a statistical measurement model that would measure the quality with which a person performed each ADL skill item within the performance of a chosen and familiar ADL task. We chose to use the many-faceted Rasch measurement model. This model could generate objective measurement of ADL ability by converting ordinal raw data scores into linear measures. The many-faceted Rasch model does this by applying specific assertions to the data, these assertions are based on expected responses.

When conceptualizing and developing the AMPS, we sought to develop test items which adhered to the basic assertions of a many-faceted Rasch measurement model. More specifically, it was important to develop test items (i.e., ADL tasks, ADL skill items) which not only added to the measurement of the construct of ADL ability, but also generated expected responses from all persons. When ADL skill items, ADL tasks, AMPS raters, and a person's ADL performance demonstrate response patterns across AMPS observations that conform to these assertions, they demonstrate acceptable goodness-of-fit to the many-faceted Rasch model, supporting the possibility that valid and reliable linear measures can be generated (Fisher, 1993, 1994, 1995a, 1997; Linacre, 1993). See Section 14.2 for a detailed explanation of Rasch measurement models, measurement assertions, and definitions of terms related to Rasch measurement methods.

### 14.1.1 Developing the AMPS Skill Items

The choice of items to include in most ADL tests has been driven by "item content specification" derived from expert opinion and common practice. The problem with this approach is that it does not ensure that the test can be used to evaluate or measure the intended construct (cf. Fisher, 1993). In developing the AMPS, we chose a different approach to ADL skill item development and selection. We began with a conceptual model and allowed that model to drive the design and content of the AMPS.

We used a variety of procedures to select the final ADL skill items. First, we implemented an extensive review of the literature to identify goal-directed actions of performance that could be derived from existing theories of practice. These theories included sensory integration (Ayres, 1972, 1979, 1986; Fisher, Murray, & Bundy, 1991) as well as rehabilitation, motor control, and information processing theories (cf. Bernstein, 1967; Brooks, 1986; Kelso, 1982; Keogh & Sugden, 1985; MacKay, 1987; Schmidt, 1988; Trombly, 1989, 1995). It is important to stress that while the models we reviewed focused primarily on body functions, our intent was to generalize from these models to observable, goal-directed actions that comprise ADL task performance (see Chapter 1, Section 1.1).

For example, the ADL process skill items Initiates and Continues were generalized from sensory integration theory (Fisher et al., 1991). When an occupational therapist implements a standardized or nonstandardized evaluation of bilateral integration and sequencing praxis (a body function), one observes for hesitations before and during the performance of sequential motor actions (e.g., jumping in a series of circles on the floor). These are symptoms of the underlying motor impairment, bilateral integration and sequencing praxis. With the AMPS, we observe people as they perform ADL tasks, and if we see pauses before they start an action, we score them down on the ADL process skill Initiates, and if they pause during an action, we score them down on the ADL process skill Continues. When scoring the AMPS, we carefully avoid consideration of what underlying person factors, body functions, or environmental factors might be the "cause" of such pauses.

In order to ensure that we focused on occupational performance, not body functions or impairments (see Figure 11-1), we videotaped and observed a wide range of ADL task performances of persons with a variety of diagnoses — developmental, cognitive, neurologic, orthopedic, psychiatric, and medical. Through this process, we identified observable, goal-directed actions that we could observe across all of the ADL task performances. Those actions that were not universal across all ADL tasks (e.g., Sits/Stands — the action of sitting from a standing position or coming to a standing position from sitting), or were redundant or not routinely observed during ADL task performances, were eliminated in the early phases of the development of the AMPS.

Finally, we verified that our data conformed to the assertions of the many-faceted Rasch model of the AMPS, and that the ADL motor and ADL process skill items measured two unidimensional constructs. When we identified ADL skill items that did not fit the assertions of the many-faceted Rasch model of the AMPS, we either revised them or eliminated them from the AMPS. At the end of this process, we had 16 universal ADL motor and 20 universal ADL process skill items.

In developing the AMPS, we have continually asserted that there is no theoretical reason for assuming that the goal-directed operations (ADL motor and ADL process skills) represent different constructs for different subgroups (e.g., gender, racial/ethnic, nationality), a basic assertion of item bias examination (Teresi, Cross, & Golden, 1989). See Chapter 15 for examples of how researchers have verified that the AMPS is a person-free measurement model, and does not differentially function by gender, race, world region, etc. While we expected some degree of differential item functioning associated with diagnosis, we have also ensured that any differential item functioning that might be present did not threaten the AMPS measurement system, and that the AMPS measures were free of test bias. For more details about our process of developing the ADL skill items, see Table 14-1 for a summary of the history of the development of the AMPS.

### 14.1.2 Developing Standardized ADL Tasks for the AMPS

Through the work of occupational therapists and researchers worldwide, standardized ADL tasks were developed, based on clinical and cultural needs. ADL tasks were developed with standardized, yet flexible criteria that allowed persons to perform the tasks in their usual manner in a familiar environment. Additionally, as AMPS tasks were developed, we sought to create tasks that encompassed multiple life areas (e.g., personal care, home maintenance, meal preparation); were relevant to persons from different world regions; and offered varying challenge (e.g., from easy to hard). The resulting list of tasks allowed persons being evaluated with the AMPS to choose to perform tasks which were not only relevant and familiar, but also of an appropriate challenge. We anticipated that by ensuring that a person performed appropriately challenging tasks, the likelihood of encountering ceiling or floor effects would be minimized.

Each AMPS observation is based on observing a person perform at least two familiar and sufficiently challenging ADL tasks. All of the tasks that a person performs

**Table 14-1 Overview of the History of the Development of the AMPS**

| Year and version/ edition | Key developments | Number of tasks, standardization sample, and number of calibrated raters |
|---|---|---|
| 1985 | • Susan Doble developed an assessment of decision-making/executive functions which she called the Process Skills Assessment (PSA)<br>• The idea of assessing persons as they performed familiar and relevant tasks that they chose to perform was a critical element of the PSA that was retained in the AMPS<br>• Five tasks were offered as choices: prepare sandwich and a bowl of soup, iron two shirts, repot two plants, polish shoes, and sew three buttons onto shirts | • 5 tasks |
| 1987: PSA-3 | • The original PSA was completely redesigned to become an assessment of occupational performance, not cognitive functioning; it was named PSA-3<br>• The items on the new PSA-3 became verbs (observable actions) that represented 15 universal ADL process skill items [current AMPS skill item name, if later revised and retained in AMPS]: Sequences, Initiates, Terminates, Paces, Calibrates, Locates [Searches/ Locates], Collects [Gathers], Organizes, Returns [Restores], Maneuvers [Positions], Manipulates [Handles], Seeks [Inquires], Attends [Heeds], Adapts [Accommodates and Adjusts as one item], and Utilizes [Notices/Responds]<br>• Scoring of the first PSA-3 was based on a 3-catgegory rating scale (unsatisfactory, ineffective, effective)<br>• Potential new tasks included housekeeping, cooking, potting plants, crafts, and puzzles, but these tasks were not yet developed and thus not included in the PSA-3. | |

(continued)

**Table 14-1 (continued)**

| | | |
|---|---|---|
| 1988: Revised PSA-3 (informally considered Research Version 1) | • The ADL process skill items were revised three times, and the total number increased to 19 [current AMPS skill item name, if later revised and retained in AMPS]: Inquires, Investigates [Notices/Responds], Heeds, Uses [Chooses and Uses as one item], Attends, Sequences, Initiates, Continues, Terminates, Paces, Locates [Searches/Locates], Gathers, Organizes, Restores, Handles, Maneuvers [Positions], Calibrates, Accommodates, and Adjusts<br>• 4- and 5-category rating scales were piloted. A list of 12 potential tasks to include in the PSA-3 was generated: peanut butter and jelly sandwich, green salad, macaroni and cheese from a package mix, sew on a button, sweep the floor, wash and dry dishes, grilled cheese sandwich, ironing a shirt, repot a house plant, cup and ball-on-a-string game, decoupage craft project, leather key fob project<br>• The possibility of creating a standardized tool that allowed persons to choose from among a variety of tasks of varying challenges was recognized as a "***standardization nightmare***" — standardization using traditional psychometric methods would require the recruitment of a large sample of people who were familiar with and willing/able to perform ***all*** of the newly developed tasks within a limited period of time<br>• ***The breakthrough***: Under the direction of Ben Wright and J. Michael Linacre, the first FACETS analysis of the Revised PSA-3 was carried out — six raters rated the same four persons' (each of whom had performed one or two of three common IADL tasks) performance on the 15 process skills included in the first PSA-3; the 5 category rating scale was used (optimal, competent, marginal [questionable], ineffective, and deficit)<br>• The importance of developing unidimensional scales of ADL ability was recognized as a critical element in future developments; any idea of including art, craft, work, or play activities was to be avoided | • 12 tasks: 9 IADL and 3 craft activities |

(continued)

**Table 14-1 (continued)**

| | | |
|---|---|---|
| 1989: Research Version 2 | • 15 ADL motor skill items were added to PSA-3, creating the first pilot version of the Assessment of Motor and Process Skills [current AMPS skill item name, if later revised and retained in AMPS]: Stabilizes, Aligns, Walks, Climbs (steps, ramps), Sits/Stands, Reaches, Turns [Bends], Coordinates, Manipulates, Maneuvers [Positions], Moves, Transports, Lifts, Calibrates, and Grips<br>• Based on the results of principal components factor and many-faceted Rasch analyses, Maneuvers [Positions] and Calibrates became ADL motor skills<br>• 17 ADL process skill items from the Revised PSA-3 were retained, Uses was divided into Uses and Chooses, and Negotiates [Navigates] was added, resulting in 19 ADL process skill items<br>• 17 new IADL tasks were developed, including two salad tasks that were later combined into a single green salad task, two repotting plant tasks that were later combined, and a pasta salad task that was later eliminated<br>• The 5-category rating scale was retained | • 17 IADL tasks |
| 1989: AMPS, Research Version 3 | • Number of ADL process skills increased to 20 with the addition of Benefits | • 17 IADL tasks<br>• 64 persons |
| 1989: AMPS, Research Version 4 | • The ADL motor skill items were revised: Climbs, and Sits/Stands were omitted, Flows and Endures were added, Turns was renamed Bends, and Maneuvers was renamed Positions<br>• A 4-category rating scale was determined to be the best alternative (competent, marginal [questionable], ineffective, and deficit) | • 27 IADL tasks<br>• 211 persons<br>• 18 raters |

(continued)

**Table 14-1 (continued)**

| | | |
|---|---|---|
| 1990/1991: AMPS, Research Edition 5 | • The number of ADL motor skill items remained 15 and the number of ADL process skill items remained 20 | • 31 tasks<br>• 850 persons<br>• 62 raters |
| 1992: AMPS, Research Edition 6 | • The number of ADL motor skills increased to 16, as both principal components factor and many-faceted Rasch analyses confirmed that Paces was both an ADL motor skill and an ADL process skill<br>• This version of the AMPS was the first one that was essentially the same as the published editions of the AMPS | • 55 tasks<br>• 1400 persons<br>• 200 raters |
| 1994: AMPS, Research Edition 7 | • First normative data published in the manual – three global age groups 18 to 39 years, 40 to 64 years, 65 to 90 years | • 56 IADL tasks<br>• 2500 persons<br>• 300 raters |
| 1995: AMPS, First Edition | • Release of the first version of the AMPS computer-scoring software; the computer-generated reports included a Raw Score Report, a Graphic Report, and a Summary Report of the person's performance skill strengths and weaknesses | • 56 IADL tasks<br>• 4697 persons<br>• 500 raters |
| 1997: AMPS, Second edition | • Release of Version 2.0 of the AMPS computer-scoring software<br>• The first seven PADL tasks were added<br>• The global normative data were updated, but continued to be only for global age groups: 3 to 15 years, 16 to 59 years, and 60 to 90 years | • 63 tasks; 56 IADL and 7 PADL<br>• 10,445 persons<br>• 1000 raters |

(continued)

**Table 14-1 (continued)**

| | | |
|---|---|---|
| 1999: AMPS, Third edition | • Release of Version 3.0 of the AMPS computer-scoring software<br>• The global normative data were updated, but continued to be only for global age groups: 3 to 8 years, 9 to 15 years, 16 to 59 years, and 60 to 90 years | • 76 tasks: 68 IADL and 8 PADL<br>• 25,565 persons<br>• 2610 raters |
| 2001: AMPS, Fourth edition | • AMPS 2001 computer-scoring software released<br>• The AMPS is first published in a two-volume format | • 83 tasks: 74 IADL and 9 PADL<br>• 46,886 persons<br>• 4,322 raters |
| 2002: AMPS, Fifth edition | • Extensive revisions were made to Volume 1 of the AMPS Manual | • 83 tasks: 74 IADL and 9 PADL |
| 2005 | • AMPS 2005 computer-scoring software released; a new Narrative Report which summarized both criterion- and norm-referenced results, is added | • 83 tasks: 74 IADL and 9 PADL |
| 2006: AMPS, Sixth edition | • Volume 2 of the AMPS Manual revised, Appendix added to Volume 1 | • 85 tasks: 75 IADL and 10 PADL<br>• 102,366 persons<br>• 9233 raters |
| 2010: AMPS, Seventh edition | • AMPS 2010 computer-scoring software released; a new Progress Report is added<br>• Extensive revisions were made to Volume 1 and Volume 2 of the AMPS Manual<br>• New normative values, by year 3 to 15 years, and 10 age groups between 16 and 103 years, are included in the manual | • 116 tasks: 100 IADL and 16 PADL<br>• 148,158 persons<br>• 13,070 raters |

for a single AMPS observation can be considered a different test or "version" of the AMPS. All possible versions of the AMPS are contained in a large item bank.

> *Data from different tests taken by different candidates can be combined and analyzed together, so long as there is some network of commonalities (candidates and/or items) linking the tests . . . . Once a bank of items has been calibrated, inclusion of items from the bank into each new test automatically equates that test to the common metric of the bank, and so to all other tests derived from the bank. (Wright, 1993, p. 299)*

Through a process of gathering data on thousands of linked AMPS task observations, we constructed a large item bank comprised of two scales, an ADL motor ability scale and an ADL process ability scale. At present, each AMPS scale has 116 tasks and each task has associated with it the 16 universal ADL motor skill items and the 20 universal ADL process skill items. The result is a large bank of test items that continues to grow as new tasks are developed and standardized. For example, since the AMPS scales now consist of 116 tasks, the resulting item bank contains 1856 ADL motor items and 2320 ADL process items. If we consider an individual assessment, based on observing the person perform two different AMPS tasks, with large variety of possible combinations, then there is an even larger subset of 32 ADL motor skill items and 40 ADL process skill items available from the entire AMPS item bank.

The equating of AMPS tasks, linked by common ADL skill items, makes it possible to compare the ability of a person who performed one set of tasks during the initial evaluation with the results of his or her performance on a different set of tasks upon reevaluation. In a like manner, equating allows us to compare performance among several persons who each performed a different set of AMPS tasks.

The list of AMPS tasks is continually growing as new clinical needs emerge and as the use of the AMPS spreads to new world regions. Bray, Fisher, & Duran (2000) investigated the addition of 21 AMPS tasks and described the process of developing, analyzing, and standardizing new tasks. In their investigation, the researchers sought to create new tasks that offered (a) an expanded range of task challenge, (b) enhanced variety of task choice, and (c) enhanced options for tasks traditionally completed by men.

When new tasks are created, many of the strategies and analyses utilized by Bray et al. (2000) are undertaken to verify the validity of adding the tasks to the AMPS. One consideration when adding a new task to the AMPS is the adherence of the task to the measurement model, such that all persons respond in a predictable manner when performing the task. The aim is to create a person-free measurement model — one where the tool's measurement properties remain the same regardless of the characteristics of the person(s) being assessed. Violations to the AMPS many-faceted Rasch measurement model assertions can occur if persons do not respond in a predictable/expected manner (e.g., persons in a particular world region find the task easier to perform than do persons in other world regions; men find the task easier to perform than do women). Such violations indicate the presence of differential task functioning and threaten the person-free measurement model, resulting in the potential for overall test bias (see Section 14.2).

### 14.1.3 Converting Ordinal Scores into Linear Measures

Rasch analysis computer programs convert summed ordinal item scores into equal-interval numbers that can be placed along a common linear continuum or scale. This conversion is accomplished through a logistic transformation of the proportion of persons obtaining a given item score (Andrich, 1988; Wright & Masters, 1982). In this process, the original raw item scores are analyzed "for the possibility of a single latent variable along which the intended measuring agents, the items [and tasks], can be calibrated and the intended objects of measurement, the subjects, can be measured" (Wright & Linacre, 1989, p. 858). When we analyze AMPS data, we consider not only the difficulty of the ADL skill items, the challenge of the tasks, and the ability of the person, but also the severity of the rater who scores the performance. Because we consider all of these facets simultaneously, we must use a many-faceted Rasch model (Linacre, 1993) and the FACETS Rasch analysis computer program (Linacre, 2009). This model provides the frame of reference for examining and accounting for all four of these facets simultaneously, resulting in the construction of a single common variable on which each facet is measured.

This feature has very important practical implications. Different persons are rated by different raters. If a rater is more severe, the persons that are scored by that rater will have lower item scores and will thus appear to be less able than if the same persons were rated by a more lenient rater. By calibrating the severity of each rater, we can use the FACETS program to subtract out the personal bias of each rater that is due to his or her overall scoring severity. The result is ***rater-free measurement*** and the ability to directly compare a person's ADL motor and ADL process ability measures that have been generated by two different AMPS raters.

Similarly, each person evaluated may perform a different combination of AMPS tasks. These tasks vary in the challenge that they offer to the person. If a person performs easy tasks, that person will obtain higher scores than if he or she performs hard tasks. The same situation occurs if two persons who have similar ability perform tasks that vary in task challenge. In order to compare the abilities of different persons, or to determine if a person improved between two AMPS observations, we must subtract out the relative challenge of each task when we estimate the ADL motor and ADL process ability measures.

The FACETS Rasch analysis computer program was the first available method that enabled us to design an ADL assessment that corrects person ADL ability measures for differences among raters due to scoring severity, while simultaneously accounting for variation in the challenge of the tasks performed by person. Therefore, the resulting person ADL ability measures are not biased by the severity of the particular rater who observed the performance, or by the challenge of the particular tasks the person performed (Lunz & Stahl, 1990; Lunz & Wright, 1989; Lunz, Wright, & Linacre, 1990). Without this measurement model and without the FACETS computer program, it would be practically impossible to generate linear measures of ADL ability from such a complex assessment model (e.g., an assessment model which now contains 116 ADL tasks of varying challenge, 16 ADL motor skill items of varying difficulty, 20 ADL process skill items of varying difficulty), and thus clinically impossible to determine how able a person is in comparison to others and how much their ADL ability has changed over time.

### 14.1.4 Summary of Objective Measurement Principles and the Development of the AMPS

The development of the AMPS was guided by the principles of modern test theory and the strong desire and need to develop objective, linear measures of ADL ability. A many-faceted Rasch measurement model was utilized within this developmental process. During the development and continued standardization of the AMPS, our evaluative process must verify that the assertions of the AMPS Rasch measurement model have not been violated. More specifically, we must verify that our theoretical model holds true for a variety of persons (e.g., persons of varying age, gender, diagnosis, nationality) within different contexts (e.g., homes, hospitals, world regions), and between diverse populations of examiners. If the assertions that are documented in Section 14.2 have not been violated, we can be assured that uinidimensional, linear, objective measures of ADL ability can be generated.

Within the field of occupational therapy, the utilization of objective measures (i.e., Rasch measurement methods) has begun to raise the quality and accuracy of our assessments and our predictions; however, continued efforts need to be made to ensure that researchers and occupational therapists are utilizing and promoting objective methodology in their practice. Objective measures can, and should, provide therapists with a common language and a sense of shared meaning of their assessment results. As consumers of health-care research, we encourage you to keep these principles in mind as you appraise the literature and seek to move forward with the implementation of evidence-based practice. In subsequent sections of this chapter, we will explore, in greater depth, how the assertions of the AMPS measurement model have been, and continue to be verified, and we will describe in greater detail how occupational therapists and health-care providers can use the ADL ability measures to effectively and consistently communicate the ADL ability of their clients.

## 14.2 Defining Key Terms and Concepts

In this section, we will define a number of terms and concepts that are important for understanding the background of the development of the AMPS and the current evidence related to the validity and reliability of the AMPS scales and measures. In so doing, we will assume knowledge of more general research-related terminology. As needed, one can refer back to this section when reading Chapter 15. This section may also be of use when reading AMPS-related research articles. The key terms and concepts are as follows:

- ***Rasch measurement models*** — Rasch models of measurement are a family of models that are based on the idea of first conceptualizing a logical pattern of responses, and then collecting data to verify that the expectations were met. As all Rasch measurement models are probabilistic, the expectations are based on assertions that acknowledge that not all people will always respond in exactly the same way. For example, the assertions of the simple Rasch model, where all items are scored dichotomously (e.g., pass or fail, agree or disagree), are that (a) items that are more difficult (e.g., preparing a complex meal) are ***more likely*** to be more difficult for all persons than are items that are less difficult (e.g., combing one's hair), and (b) persons with more of the trait being evaluated (e.g., ADL ability), are ***more likely*** to pass more difficult items than are persons with less of the trait.

- ***Many-faceted Rasch models*** — Many-faceted Rasch models are an extension of the simple Rasch model discussed above, as they include more than two facets. That is, the simple Rasch model considers items that represent the trait or construct to be measured (e.g., combing one's hair represents the construct of ADL) and persons who have more or less of the trait. In the simple Rasch model, items and persons are the facets — the simple Rasch model is thus a two-faceted model. In the case of the AMPS, we have four facets: (a) items (ADL actions) that vary in terms of how difficult they are to perform in a competent, skillful manner; (b) tasks that vary in the challenge they offer to the person who performs them; (c) raters who vary in severity when they score the AMPS; and (d) people who vary in terms of their ADL ability. This means that the basic assertions of the many-faceted Rasch model of the AMPS become:

  - Items that are more difficult (e.g., Positions, Benefits) are ***more likely*** to be more difficult for all persons than are items that are less difficult (e.g., Lifts, Uses).
  - Tasks that are more challenging (e.g., Pasta with meat, sauce, green salad, and beverage) are ***more likely*** to be more challenging for all persons than are tasks that are less challenging (e.g., Eating a snack and drinking a beverage).
  - Lenient raters are ***more likely*** to assign higher scores to all persons than are strict raters.
  - Persons with higher ADL ability are ***more likely*** to receive higher scores on all items/tasks than are persons of lower ADL ability.

- ***Calibration of items, tasks, raters, and persons*** — When data from a set of evaluations (e.g., AMPS observations) are entered into the computer and analyzed with specialized Rasch analysis computer programs (e.g., FACETS) (Linacre, 2009), the computer program calibrates the elements of each facet along a common linear continuum. That is, logistic transformation procedures are used to convert ordinal raw data (e.g., scores of 1, 2, 3 or 4) into linear (equal interval) calibrations. These calibrations are mathematical estimates of where on the linear continuum the element most likely is located. In the case of the AMPS, items are calibrated for their difficulties, tasks are calibrated in terms of their challenges, raters are calibrated in terms of their severities, and people are calibrated in terms of their degree of ADL ability. When Rasch analysis computer programs are used to convert ordinal item scores into equal-interval numbers, all elements that are calibrated are placed along the same linear continuum or scale. The linear continua of the AMPS represent ADL motor ability and ADL process ability.

- ***ADL skill item difficulty*** — With the AMPS, the ADL skill item difficulty calibration is the estimated location of each ADL skill item on the continuum of increasing ADL motor or ADL process ability, ranging from easy to hard. The final estimate of each ADL skill item difficulty calibration is based on a logistic transformation of all the ratings in the AMPS database that have been given to each ADL skill item. ADL skill items more often given low ratings have proportionally lower summed raw total scores, and are calibrated as more difficult ADL skills.

- ***ADL task challenge*** — The ADL task challenge calibration of the AMPS is the estimated location of each task on the same continuum of increasing ADL motor or ADL process ability, ranging from very easy to much harder than average task challenge. The final estimate of each task challenge calibration is based on a logistic transformation of the sum of all the item ratings that were assigned to each task in the AMPS when people performed that ADL task. ADL tasks more often given low ratings have proportionally lower summed raw total scores, and are calibrated as more challenging ADL tasks.

- ***Rater severity*** — The rater severity calibration is the estimated location of that rater on the same continuum of increasing ADL motor or ADL process ability, with the rater calibration values representing a continuum of severity from more lenient to more strict raters of ADL ability. The final estimate of each rater severity calibration is based on a logistic transformation of the sum of the item ratings that raters assigned to a common set of calibration cases scored during an AMPS training course. Raters who overall assign lower ratings have proportionally lower summed raw total scores, and are calibrated as more severe raters.

- ***Person ADL ability measure*** — The person ADL ability measure (person calibration) is the estimated location of that person on the continuum of increasing ADL motor or ADL process ability. The person ADL ability measure represents how skilled the person was when he or she performed AMPS tasks. The final estimate of each person's ADL motor and ADL process ability measures are also based on a logistic transformation of the sum of the ADL skill item ratings that the rater assigned when he or she scored that person. Unlike item, task, and rater calibrations, however, person ADL ability measures are adjusted (i.e., moved up or down along the common scale) to account for the challenges of the ADL tasks the person performed and the severity of the rater who scored the person's ADL task performance.

- ***Logits*** — The derived calibrations and measures discussed above are all expressed in equal-interval units of measurement called logits. The term "*logit*" comes from the idea that calibrations and measures are based on the natural logarithm of the ratio or odds (***log***-odds probability un***its*** or logits) of obtaining adjacent scores on a given ADL skill item (e.g., 1 vs. 2, 2 vs. 3, 3 vs. 4) when a person of a given ability is observed performing a given ADL task, and scored by a given rater. Just as a millimeter represents an equal interval unit of length, but the term "*millimeter*" only has meaning once one understands the concept of length, the term "*logit*" only has meaning when one understands that a ***logit refers to an equal interval unit of ADL ability*** when used in the context of the AMPS.

- ***Goodness-of-fit statistics*** — When ADL skill items, ADL tasks, raters, and persons are calibrated using FACETS, the program generates goodness-of-fit statistics that can be used to evaluate how well the data met the expectations of the many-faceted Rasch model of the AMPS discussed above. When the data fit the model, ADL skill items, ADL tasks, raters, and persons will demonstrate goodness of fit. If, however, an ADL skill item that calibrated as being easier (i.e., expected to get higher scores) is scored lower than expected, that rating will misfit. The same holds true if an ADL skill item that is hard is scored higher than expected. For example, if a person has been given scores of 2 or 1 on easy ADL skill items like Chooses, Attends, and Gathers, we do not suddenly expect that person to get a score of 4 on a relatively difficult ADL skill item such as Accommodates. In this case, the rating for Accommodates will misfit. Since that rating was given to a person by a rater, both the person and the rater will also misfit in relation to that assigned score. Since Rasch models are probabilistic, we do expect a certain amount of misfit; we become concerned when misfit is too extreme.

  The goodness-of-fit statistics generated by the FACETS program include mean square (*MnSq*) and standardized ($z$) fit statistics. The expected values for these statistics are *MnSq* = 1.0 and $z$ = zero. Since some variation is allowed, no commonly agreed upon criteria for acceptable goodness of fit have been established. Wright and Linacre (1994) have suggested that *MnSq* values as high as 1.7 are reasonable for observational data, and Wilson (2005) has stressed the importance of considering both the magnitude of the *MnSq* value and the statistical test of its magnitude (i.e., $z$). That is, $z$ values of 2 or more signal that there was too much variation in the data — there were too many unexpected ratings. Our criteria have consistently been set to be stricter than those recommended by Wright and Linacre. That is, ***our criteria for acceptable goodness of fit is MnSq ≤ 1.4 and/or a z value < 2; if both MnSq is > 1.4 and z is 2 or higher, the data have failed to demonstrate acceptable goodness of fit***. In general, if the criterion for $z$ is set at 2.0, then we expect that 95% of the data will demonstrate acceptable goodness of fit. That is, 95% of the items, tasks, raters, and persons should demonstrate acceptable goodness of fit. Among performance tests, however, it is not uncommon for the percentage of persons who misfit to be somewhat higher.

- ***Differential item (or task) functioning*** — The term "*differential item functioning*" or DIF, refers to the idea that the calibration values for items differ between two different groups of persons (e.g., gender, diagnosis, world region). That is, any ADL skill item that displays differential item functioning among groups, which are expected to have the same calibration values, will actually have different item calibration values. Conversely, when the item calibration values are the same among groups, the ADL skill items that are easy for one group are also easy for those in another group. The same principle can be applied to tasks. While we expect task challenges to be the same among groups, should a task be less challenging for one group than for another, it can be said to demonstrate differential task functioning. While the presence of differential item or task functioning is never ideal, it is not always severe enough to disrupt the measurement system. When it does, test bias occurs (for more discussion, see the explanation of differential test functioning below).

  There is no universally accepted standard for determining if differential item or task functioning is present. While Rasch computer programs often use *t* tests to evaluate for differential item or task functioning, large sample sizes are associated with small standard errors and too much power, and the risk is over identification of significant differences. The more common alternative, therefore, is the use of effect sizes, but again, there is no commonly accepted standard. While recommended values typically range from 0.40 logit to 0.60 logit, it has become most common to consider values less than 0.50 logit as evidence of no differential item or task functioning (Conrad, Dennis, Bezruczko, Funk, & Riley, 2007; Linacre, 1994; Tennant & Pallant, 2007; Tristán, 2006; Wilson, 2005). In many of our early analyses, we have used a stricter criterion, 0.43 logit, but now consider 0.50 to be more reasonable. Finally, because such evaluations for differential item or task functioning often do not include an evaluation of statistical significance, the term "*meaningful difference*" often has been used.

- ***Differential test functioning*** — When differential item or task functioning is present, it can disrupt the measurement system. That is, it may become the case that the ADL motor and ADL process measures of the AMPS for one group (e.g., gender, diagnosis, world region) are placed at a disadvantage when their AMPS results are compared to the AMPS results of people from another group. To ensure that this is not the case, it is important to ensure that any identified differential item or task functioning does not result in differential test functioning. When differential test functioning is present, the measurement system is disrupted and the result is test bias (Boomsma, van Duijn, & Snijders, 2001; Pae & Park, 2006; Wright & Stone, 1999).

The most common method for determining if there is differential test functioning is to plot ADL ability measures of persons from both groups that have been based on the item difficulty calibrations for one group against their ADL measures when they have been based on the item difficulty calibrations for the second group, and then evaluate if the paired ADL measures fall within a 95% confidence interval, indicating no evidence of differential test functioning (Wright & Stone, 1979). When there is no evidence of differential test functioning, the measurement system is not disrupted.

## 14.3 Summary

As a consumer of research literature, a basic knowledge of objective measurement will provide a foundation for not only understanding the theoretical development of the AMPS, but also for understanding, critiquing, and synthesizing research studies that are related to the AMPS, as well as other commonly used assessments. It is our hope that this brief introduction of the conceptualization of the AMPS and the list of key terms and definitions will aid you as you read about the validity and reliability of the AMPS (see Chapter 15).

# 15. CURRENT STANDARDIZATION SAMPLE, ITEM AND TASK CALIBRATION VALUES, AND VALIDITY AND RELIABILTY OF THE AMPS

by

Anne G. Fisher

Brenda K. Merritt

Within this chapter of the manual, we first provide the reader with information related to the current standardization sample for the AMPS, and the skill item and task calibration values that have been based on this sample and subsequently incorporated into the AMPS computer-scoring software. Following our presentation of this background information, we will summarize the existing evidence supporting the validity and reliability of the AMPS scales and ADL ability measures. This will include a summary of the key research that has been conducted to verify that the AMPS scales are valid and yield valid and reliable measures of the quality of ADL task performance of persons who are from different world regions or cultural backgrounds, and who have different diagnoses, genders, ethnicities, and ages. We will also discuss evidence that quality of ADL task performance develops and changes over the life span, and how the environment may impact ADL ability. We also present information related to the development of AMPS normative values and how these norms, as well as evidence related to the reliability of individual ADL ability measures, can be used in the process of interpreting the AMPS Graphic Report. Lastly, we will discuss the utility of using the AMPS ability measures to assist in the prediction of a person's need for assistance, as well as how the AMPS ability measures have been utilized as sensitive outcome measures within health-related research.

**Prior to reading this chapter, the reader is encouraged to read Chapter 14 in order to become familiar with**

- **The conceptual model used to guide the development of the AMPS**
- **Key concepts and definitions of terms related to many-faceted Rasch measurement methods**

## 15.1 Current Standardization Sample of the AMPS

### 15.1.1 Raters Who Scored the Persons in the Standardization Sample

There are 13,070 occupational therapists who scored persons in the current standardization sample (see Section 15.1.2). Each of these occupational therapists demonstrated that they scored the AMPS in a valid and reliable manner. Excluded from this group were 114 additional raters who demonstrated invalid (extremely lenient or extremely strict) rater calibration values on the ADL motor ($n$ = 42, 0.3% of total) and ADL process ($n$ = 72, 0.5% of total) scales of the AMPS. Among the 13,070 remaining raters, 95% demonstrated acceptable goodness of fit to the many-faceted Rasch model of the AMPS ($MnSq \leq 1.4$ and/or $z < 2$), supporting ***high inter- and intrarater reliability***.

### 15.1.2 Demographic Characteristics of the Current AMPS Standardization Sample

When the current AMPS standardization sample was compiled, there was a total of 157,060 persons available within the AMPS database. To create the standardization sample, all of the persons who were co-scored by multiple raters within the process of determining rater severity (i.e., rater severity calibration values) were excluded from the sample ($n$ = 260). This resulted in a potential standardization sample of 156,800 persons. Within this potential sample, 8642 persons (5.5%) were omitted due to rater scoring error (e.g., scores assigned by raters were invalid as the raters were too lenient or too strict). Rater scoring error was of two forms. Either the rater had an extreme rater calibration value that precluded valid estimation of ADL ability measures for any of the persons he or she had evaluated (see Section 15.1.1), or the rater had a valid rater calibration value but scored an occasional person too leniently or too strictly (i.e., in an invalid manner). Approximately 4% of the potential sample was omitted for this latter reason; 5% rater scoring error would be expected due to chance. Among the 148,158 remaining persons comprising the current AMPS standardization sample, 92% demonstrated acceptable goodness of fit to the AMPS motor scale and 90% demonstrated acceptable goodness of fit to the AMPS process scale ($MnSq \leq 1.4$ and/or $z < 2$).

The age of the persons within the standardized sample ranged from 3 to 103 years, and included persons with an assortment of diagnoses (including healthy, well persons). The standardization sample was divided into the following "diagnostic" groups:

- **Well** — Healthy, well persons with no apparent problems or history of medical, psychiatric, physical, or cognitive conditions
- **Frail old** — Older adults 60 years of age or older, who have no known medical, psychiatric, physical, or cognitive conditions, but who demonstrate possible physical or medical problems, possible problems with safety or judgment, need of supervision because of safety risks, and/or need of assistance with complex ADL tasks
- **Mild** — Persons at risk for or who have been diagnosed with mild disabilities, including sensory integrative dysfunction (without mental retardation or intellectual disability), developmental coordination disorder, specific learning disability (e.g., dyslexia, reading disorder), speech and language disorder, and/or disorder of attention (with or without hyperactivity)
- **Other DD** —Persons with neurological developmental disorders, including cerebral palsy, spina bifida, multiple developmental disorders (e.g., mental retardation and cerebral palsy), and/or unspecified developmental disorders
- **CVA** — Persons with right- or left-sided cerebral vascular accident
- **Other neurol** — Person with other types of neurological disorders, including traumatic brain injury, multiple sclerosis, Parkinsonism, progressive neurological disorders, nonhemispheric stroke, unspecified and/or multiple acquired neurological disorders
- **MR** — Persons with mental retardation (intellectual disabilities) without major concomitant medical, psychiatric, or physical conditions
- **Mskl** — Persons with musculoskeletal disorders, including rheumatoid arthritis, osteoarthritis, spinal cord injury, hip fracture or replacement, and/or limb injury or deformity

- **Medical/sensory** — Persons with medical conditions (e.g., cardiovascular, respiratory, burns, AIDS/HIV) or sensory disorders (e.g., visual, auditory, vestibular)

- **Schiz/thought** — Persons with schizophrenia or other type of thought disorder

- **Other psych/autism** — Persons with other psychiatric disorders (e.g., affective, anxiety, personality, depression, bipolar, substance abuse, unspecified) or a disorder on the autism spectrum

- **Dementia** — Persons with dementia, including those with Alzheimer, multi-infarct, unspecified, or mixed type

- **Other memory** — Persons with memory disorders not associate with dementia

- **Multiple/unknown** — Persons with two or more diagnoses from different categories (e.g., neurological and medical, musculoskeletal and cognitive), or whose diagnosis or type of disorder is not known

Persons within the standardized sample were from various world regions (i.e., Nordic countries, United States, Canada, United Kingdom, Republic of Ireland, Middle Europe, other European countries, Australia, New Zealand, Asia, South America, Middle East, and unknown); and 44% ($n$ = 65,110) of the persons were men and 56% ($n$ = 82,989) were women (see Tables 15-1 and 15-2). Their ADL motor ability measures ranged from 3.95 logits to -3.00 logits, and their ADL process ability measures ranged from 3.00 logits to -2.00 logits; positive values indicate more ADL ability (i.e., better observed quality of ADL task performance).

## 15.2 Current ADL Skill Item Difficulty and ADL Task Challenge Calibration Values

The standardization sample described in Section 15.1.2 was utilized to generate current calibration values for the AMPS skill items and tasks that are incorporated into the AMPS computer-scoring software. These calibration values are shown in Tables 15-3 to 15-5.

## Table 15-1 AMPS Standardization Sample: "Diagnostic" Group by Age

| Age group (yrs) | "Diagnostic" group* | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Well | Frail old | Mild | Other DD | CVA | Other neurol | MR | Mskl | Medical/ sensory | Schiz/ thought | Other psych/ autism | Dementia | Memory | Multiple/ unknown | Total |
| 3 | 523 | 0 | 11 | 32 | 0 | 1 | 3 | 4 | 2 | 0 | 2 | 0 | 0 | 18 | 596 |
| 4 | 761 | 0 | 53 | 122 | 5 | 20 | 11 | 23 | 12 | 1 | 14 | 0 | 0 | 40 | 1062 |
| 5 | 880 | 0 | 103 | 192 | 8 | 25 | 28 | 21 | 8 | 1 | 26 | 0 | 0 | 63 | 1355 |
| 6 | 902 | 0 | 162 | 218 | 9 | 47 | 44 | 18 | 10 | 0 | 39 | 0 | 0 | 68 | 1517 |
| 7 | 901 | 0 | 190 | 197 | 6 | 53 | 30 | 33 | 11 | 0 | 27 | 0 | 0 | 73 | 1521 |
| 8 | 774 | 0 | 158 | 201 | 6 | 46 | 49 | 31 | 12 | 2 | 32 | 0 | 0 | 98 | 1409 |
| 9 | 776 | 0 | 109 | 194 | 12 | 62 | 38 | 35 | 9 | 2 | 39 | 0 | 0 | 87 | 1363 |
| 10 | 724 | 0 | 94 | 176 | 6 | 69 | 38 | 31 | 15 | 1 | 38 | 0 | 0 | 96 | 1288 |
| 11 | 696 | 0 | 74 | 155 | 6 | 65 | 31 | 28 | 6 | 2 | 46 | 0 | 2 | 90 | 1201 |
| 12 | 584 | 0 | 57 | 148 | 14 | 64 | 46 | 27 | 8 | 3 | 44 | 0 | 0 | 80 | 1075 |
| 13 | 515 | 0 | 50 | 146 | 11 | 75 | 28 | 24 | 14 | 6 | 38 | 0 | 0 | 104 | 1011 |
| 14 | 395 | 0 | 45 | 130 | 6 | 95 | 37 | 38 | 12 | 3 | 61 | 0 | 0 | 88 | 910 |
| 15 | 368 | 0 | 37 | 113 | 11 | 99 | 45 | 33 | 14 | 21 | 71 | 0 | 1 | 93 | 906 |
| 16 to 29 | 279 | 0 | 149 | 1044 | 277 | 2611 | 992 | 1091 | 167 | 2475 | 2240 | 0 | 4 | 2946 | 14275 |
| 30 to 39 | 222 | 0 | 45 | 380 | 525 | 2403 | 717 | 1378 | 194 | 2434 | 1829 | 5 | 10 | 2902 | 13044 |
| 40 to 49 | 162 | 0 | 26 | 259 | 1539 | 3183 | 562 | 1750 | 245 | 1981 | 1627 | 34 | 11 | 3910 | 15289 |
| 50 to 59 | 213 | 0 | 9 | 176 | 3487 | 3651 | 394 | 2093 | 407 | 1290 | 1353 | 248 | 35 | 5182 | 18538 |
| 60 to 61 | 547 | 75 | 1 | 15 | 852 | 620 | 41 | 451 | 142 | 151 | 187 | 88 | 15 | 1060 | 4245 |
| 62 to 64 | 574 | 85 | 1 | 19 | 1395 | 955 | 57 | 645 | 205 | 212 | 253 | 158 | 24 | 1674 | 6257 |
| 65 to 69 | 708 | 147 | 0 | 35 | 2434 | 1397 | 55 | 1105 | 435 | 236 | 400 | 417 | 54 | 3234 | 10657 |
| 70 to 74 | 601 | 190 | 0 | 8 | 2770 | 1372 | 22 | 1364 | 609 | 125 | 380 | 832 | 79 | 4313 | 12665 |
| 75 to 79 | 374 | 206 | 0 | 4 | 2696 | 1287 | 11 | 1619 | 819 | 83 | 341 | 1285 | 112 | 5462 | 14299 |
| 80 to 103 | 294 | 373 | 0 | 2 | 2974 | 1371 | 7 | 3289 | 1944 | 57 | 337 | 2350 | 230 | 10447 | 23675 |
| **Total** | 12773 | 1076 | 1374 | 3966 | 19049 | 19571 | 3286 | 15131 | 5300 | 9086 | 9424 | 5417 | 577 | 42128 | **148158** |

* See Section 15.1.2 for further clarification

**Table 15-2 AMPS Standardization Sample: World Region by "Diagnostic" Group and Gender**

| "Diagnostic" group* | Nordic countries | United States & Canada | United Kingdom & Ireland | German-speaking Europe | Other European | Australia & New Zealand | Asia | South America | Middle East | Unknown | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Well | 3389 | 3552 | 2333 | 64 | 722 | 1025 | 1598 | 11 | 48 | 31 | 12773 |
| Frail old | 251 | 196 | 251 | 0 | 58 | 101 | 212 | 3 | 1 | 1 | 1074 |
| Mild | 475 | 208 | 192 | 90 | 107 | 72 | 177 | 1 | 36 | 18 | 1376 |
| Other DD | 1186 | 387 | 723 | 26 | 401 | 303 | 893 | 11 | 12 | 24 | 3966 |
| CVA | 7990 | 1417 | 1770 | 242 | 3260 | 814 | 3420 | 23 | 96 | 17 | 19049 |
| Other neurol | 7669 | 2651 | 2816 | 311 | 2641 | 1563 | 1782 | 20 | 75 | 43 | 19571 |
| MR | 628 | 285 | 1696 | 6 | 108 | 146 | 395 | 10 | 8 | 4 | 3286 |
| Mskl | 5945 | 2414 | 1868 | 109 | 1759 | 1191 | 1704 | 24 | 57 | 60 | 15131 |
| Medical/sensory | 1435 | 1370 | 1108 | 16 | 440 | 547 | 350 | 7 | 15 | 12 | 5300 |
| Schiz/thought | 2096 | 1064 | 2652 | 25 | 303 | 1704 | 1224 | 6 | 10 | 2 | 9086 |
| Other psych/autism | 2903 | 1420 | 2832 | 29 | 498 | 1394 | 327 | 5 | 11 | 5 | 9424 |
| Dementia | 1603 | 1055 | 1607 | 12 | 379 | 283 | 463 | 2 | 6 | 7 | 5417 |
| Other memory | 163 | 102 | 183 | 1 | 21 | 60 | 44 | 0 | 3 | 0 | 577 |
| Multiple/unknown | 11888 | 7344 | 9725 | 366 | 5655 | 4141 | 2784 | 51 | 83 | 91 | 42128 |
| **Gender** | | | | | | | | | | | |
| Male | 20314 | 9870 | 13716 | 576 | 6905 | 6287 | 6973 | 84 | 212 | 173 | 65110 |
| Female | 27289 | 13578 | 16029 | 721 | 9445 | 7052 | 8397 | 90 | 246 | 142 | 82989 |
| Unknown | 18 | 15 | 11 | 0 | 2 | 5 | 3 | 0 | 3 | 0 | 57 |
| **Total** | 47621 | 23463 | 29756 | 1297 | 16352 | 13344 | 15373 | 174 | 461 | 315 | 148158 |

* See Section 15.1.2 for further clarification

**Table 15-3 ADL Motor and ADL Process Skill Item Calibration Values (logits)**

| ADL motor items | | | ADL process items | | |
|---|---|---|---|---|---|
| Easier | | | Easier | | |
| ↑ | 0.6 | Endures | ↑ | 1.3 | Uses |
| | 0.4 | Lifts | | 0.7 | Chooses |
| | 0.4 | Aligns | | 0.6 | Sequences |
| | 0.4 | Moves | | 0.6 | Searches/Locates |
| | 0.2 | Transports | | 0.5 | Attends |
| | 0.1 | Flows | | 0.3 | Inquires |
| | 0.1 | Grips | | 0.3 | Gathers |
| | 0.1 | Reaches | | 0.3 | Heeds |
| | -0.1 | Bends | | 0.1 | Terminates |
| | -0.1 | Manipulates | | 0.0 | Navigates |
| | -0.1 | Walks | | 0.0 | Handles |
| | -0.2 | Stabilizes | | -0.1 | Adjusts |
| | -0.2 | Coordinates | | -0.1 | Continues |
| | -0.4 | Paces | | -0.2 | Restores |
| | -0.5 | Calibrates | | -0.2 | Initiates |
| ↓ | -0.8 | Positions | | -0.2 | Organizes |
| Harder | | | | -0.3 | Paces |
| | | | | -0.6 | Notices/Responds |
| | | | | -1.2 | Benefits |
| | | | ↓ | -1.6 | Accommodates |
| | | | Harder | | |

**Table 15-4 AMPS Motor Task Challenge Calibration Values (logits)**

| | | |
|---|---|---|
| **Very easy ADL tasks** | | |
| 1.1 | P-12 | Eating a snack with a utensil |
| 0.9 | P-13 | Eating a snack and drinking a beverage |
| 0.9 | P-14 | Brushing or combing hair |
| **Much easier than average ADL tasks** | | |
| 0.6 | P-15 | Washing and drying the face |
| X.X* | P-17 | Washing and drying hands |
| 0.5 | P-1 | Eating a meal |
| 0.5 | P-2 | Brushing teeth |
| 0.5 | P-9 | Eating an Asian meal with chopsticks |
| **Easier than average ADL tasks** | | |
| 0.4 | P-6 | Upper body dressing — garment within reach |
| 0.4 | P-16 | Shaving the face using electric razor |
| 0.4 | S-1 | Feeding a cat — dry cat food and water |
| 0.4 | S-3 | Feeding a dog — dry dog food and water |
| 0.3 | A-1 | Beverage from the refrigerator — one person |
| 0.3 | A-2 | Hot or cold instant drink — one person |
| X.X* | J-10 | Putting away clean dishes from a dishwasher |
| 0.3 | L-1 | Folding a basket of laundry |
| 0.3 | M-1 | Setting a table — one or two persons |
| 0.3 | P-3 | Upper body grooming/bathing |
| 0.3 | P-4 | Putting on socks and shoes — slip-on or prefastened |
| 0.2 | A-5 | Single-cup espresso or coffee — two persons |
| 0.2 | F-12 | Presliced meat and/or cheese sandwich with vegetable(s) — one person |
| 0.2 | J-2 | Hand washing dishes |
| 0.2 | L-3 | Loading and starting a washing machine |
| 0.2 | L-5 | Ironing a shirt — ironing board already set up |
| 0.2 | N-2 | Watering plants and removing dead leaves |
| 0.2 | O-1 | Polishing shoes |
| 0.2 | P-5 | Putting on socks and shoes — fastened or tied |
| 0.2 | P-11 | Upper/lower body dressing — garments set out |
| **Average ADL tasks** | | |
| 0.1 | A-3 | Pot of boiled/brewed coffee or tea — one or two persons |
| 0.1 | C-1 | Cold cereal and beverage — one person |
| 0.1 | C-3 | Sour milk (“fil”), “kefir,” or yogurt with cereal and beverage — one person |

(continued)

X.X* = task challenge is approximate; insufficient data available for stable calibration

**Table 15-4 AMPS Motor Task Challege (continued)**

| | | |
|---|---|---|
| **Average ADL tasks (continued)** | | |
| 0.1 | D-6 | Boiled egg(s) served in egg cup(s) |
| 0.1 | F-2 | Presliced meat or cheese sandwich — one person |
| 0.1 | F-7 | Open-face meat or cheese sandwich with sliced vegetable — one person |
| 0.1 | F-8 | Open-face sandwich with soft spread and sliced vegetable — one person |
| 0.1 | F-9 | Jam sandwich — one person |
| X.X* | F-13 | Open-face chocolate-hazelnut spread sandwich — one person |
| 0.1 | H-1 | Cottage cheese and fruit salad — one person |
| 0.1 | I-8 | Instant noodles, soup, or beans — one person |
| 0.1 | I-18 | Heating a precooked meal or dessert in a microwave — one or two persons |
| 0.1 | K-1 | Making a bed with standard sheets and blanket or "duvet" |
| 0.1 | K-3 | Making a freestanding bed, "duvet" edges folded under |
| 0.1 | L-2 | Hand washing laundry |
| 0.1 | M-2 | Setting a table, Swedish-style — four persons |
| 0.1 | P-7 | Upper and lower body dressing — garments stored |
| 0.1 | P-10 | Showering |
| 0.1 | S-2 | Feeding a cat — moist cat food and water |
| 0.1 | S-4 | Feeding a dog — moist dog food and water |
| 0.0 | A-4 | Stove-top espresso coffee — one to four persons |
| 0.0 | B-1 | Toast and instant coffee, tea, instant soup, or hot chocolate — one person |
| 0.0 | C-2 | Hot cooked cereal and beverage — one person |
| 0.0 | F-1 | Peanut butter and jelly sandwich — one person |
| 0.0 | F-4 | Grilled cheese sandwich and beverage — one person |
| 0.0 | F-5 | Open-face cheese or liverpaste sandwich on unsliced soft bread, and boiled/brewed coffee or tea — one or two persons |
| 0.0 | F-6 | Open-face cheese or liverpaste sandwich on presliced soft bread, and boiled/brewed coffee or tea — one or two persons |
| 0.0 | F-10 | Grilled cheese on toast and beverage — one person |
| X.X* | F-11 | Presliced meat or cheese with vegetable and beverage — two to four persons |
| 0.0 | G-1 | Boiled/brewed coffee or tea and cookies/biscuits served at table ("fika") — two to four persons |
| 0.0 | G-2 | Boiled/brewed coffee or tea and cookies/biscuits served on a tray — two or three persons |
| X.X* | H-5 | Tossed green salad with packaged lettuce and dressing, served in a large bowl — two to four persons |
| | | (continued) |

X.X* = task challenge is approximate; insufficient data available for stable calibration

**Table 15-4 AMPS Motor Task Challege (continued)**

| | | |
|---|---|---|
| **Average ADL tasks (continued)** | | |
| 0.0 | I-1 | Canned soup and crackers or presliced bread — one or two persons |
| 0.0 | I-2 | Fried green plantains ("tostones") |
| 0.0 | I-3 | Fried ripe plantains |
| 0.0 | I-4 | Beans and toast — one person |
| 0.0 | I-9 | Quick noodles cooked in a pot — one person |
| 0.0 | I-15 | Prepare tomato sauce for pasta — two persons |
| 0.0 | I-17 | Heating a frozen meal or dessert in a microwave — one person |
| 0.0 | J-1 | Sweeping the floor |
| 0.0 | J-8 | Cleaning windows |
| 0.0 | J-9 | Hand washing, drying, and putting away dishes |
| 0.0 | K-2 | Making a bed against a wall, "duvet" edges folded under |
| X.X* | K-7 | Making bed (with mattress on the floor against a wall) with standard sheets and blanket or "duvet" |
| X.X* | K-8 | Spreading bedding on the floor, Japanese-style — one person |
| 0.0 | N-1 | Repotting a small houseplant |
| 0.0 | P-8 | Upper body grooming and total body dressing |
| 0.0 | Q-1 | Sweeping outside |
| 0.0 | R-1 | Shopping |
| -0.1 | B-2 | Toast and boiled/brewed coffee or tea — one person |
| -0.1 | C-4 | Hot cooked cereal, open-face cheese sandwich, and beverage — one person |
| -0.1 | D-1 | Scrambled or fried eggs, toast, and beverage — one person |
| -0.1 | D-2 | Scrambled or fried eggs, toast, and boiled/brewed coffee or tea — one person |
| -0.1 | D-3 | Scrambled or fried eggs, meat, and boiled/brewed coffee or tea — one person |
| -0.1 | D-5 | Scrambled or fried eggs, toast, and espresso coffee — one person |
| -0.1 | E-1 | French toast and beverage — two persons |
| -0.1 | E-3 | Thick (e.g., American) pancakes and beverage — one to three persons |
| -0.1 | F-3 | Tuna, chicken, or crab salad sandwich — one person |
| -0.1 | H-2 | Fresh fruit salad — two persons |
| -0.1 | H-3 | Tossed salad with dressing, served in individual bowls — two or three persons |
| -0.1 | H-4 | Green salad, served in large bowl, with dressing on side — two or three persons |

(continued)

X.X* = task challenge is approximate; insufficient data available for stable calibration

**Table 15-4 AMPS Motor Task Challege (continued)**

| | | |
|---|---|---|
| **Harder than average ADL tasks** | | |
| -0.1 | I-5 | Vegetable preparation — one to four persons |
| -0.1 | I-14 | Pasta with sauce and beverage — two persons |
| X.X* | I-20 | Gazpacho (blended, cold tomato soup) — two to four persons |
| -0.1 | L-4 | Ironing a shirt — setting up the ironing board |
| -0.1 | L-7 | Hanging washed, spin-dried laundry to dry |
| -0.2 | D-4 | Omelette or scrambled eggs with added ingredients, toast, and beverage — one person |
| -0.2 | G-3 | Cake, muffins, or brownies |
| -0.2 | I-6 | Pasta with sauce, green salad, and beverage — two persons |
| -0.2 | I-11 | Fried rice — one or two persons |
| -0.2 | I-21 | Miso soup — one or two persons |
| X.X* | I-22 | Rice, soup, and a side dish — one person |
| -0.2 | J-5 | Mopping the floor |
| -0.2 | K-4 | Changing standard sheets on a freestanding bed |
| -0.2 | L-6 | Ironing multiple garments and putting garments away |
| -0.2 | Q-2 | Raking grass cuttings or leaves |
| -0.2 | Q-3 | Weeding |
| -0.3 | D-7 | Spanish omelette with added ingredients — two persons |
| -0.3 | E-2 | Thin (e.g., European) pancakes and beverage — one to three persons |
| -0.3 | I-7 | Pasta with meat, sauce, green salad, and beverage — two persons |
| -0.3 | I-10 | Fried meat and vegetable dish with a bowl of rice — one person |
| -0.3 | I-16 | Pasta with meat, sauce, and beverage — two persons |
| -0.3 | J-3 | Vacuuming, moving no furniture |
| -0.3 | J-4 | Vacuuming, moving lightweight furniture |
| -0.3 | K-5 | Changing sheets and “duvet” cover on a bed against wall |
| -0.3 | K-6 | Changing sheets and “duvet” cover on a freestanding bed |
| -0.4 | I-13 | Vegetable soup — one to four persons |
| -0.4 | J-7 | Cleaning a bathroom |
| -0.4 | Q-4 | Vacuuming the inside of an automobile |
| X.X* | Q-5 | Repair a bicycle tube puncture or hole |
| **Much harder than average ADL tasks** | | |
| -0.5 | I-12 | Vegetable soup, vegetables sautéed — one to four persons |
| -0.5 | I-19 | Meatballs with boiled potatoes, sauce, boiled vegetable, and beverage — two to four persons |
| -0.5 | J-6 | Vacuuming two rooms on different levels |

X.X* = task challenge is approximate; insufficient data available for stable calibration

**Table 15-5 AMPS Process Task Challenge Calibration Values (logits)**

| | | |
|---|---|---|
| **Very easy ADL tasks** | | |
| 0.8 | P-4 | Putting on socks and shoes — slip-on or prefastened |
| 0.8 | P-12 | Eating a snack with a utensil |
| 0.8 | P-13 | Eating a snack and drinking a beverage |
| **Much easier than average ADL tasks** | | |
| 0.7 | A-1 | Beverage from the refrigerator — one person |
| 0.7 | P-5 | Putting on socks and shoes — fastened or tied |
| 0.7 | P-6 | Upper body dressing — garment within reach |
| 0.7 | P-14 | Brushing or combing hair |
| 0.6 | P-1 | Eating a meal |
| 0.6 | P-2 | Brushing teeth |
| 0.5 | K-1 | Making a bed with standard sheets and blanket or “duvet” |
| 0.5 | P-9 | Eating an Asian meal with chopsticks |
| 0.5 | P-15 | Washing and drying the face |
| 0.5 | P-16 | Shaving the face using electric razor |
| X.X* | P-17 | Washing and drying hands |
| 0.5 | S-1 | Feeding a cat — dry cat food and water |
| 0.5 | S-3 | Feeding a dog — dry dog food and water |
| **Easier than average ADL tasks** | | |
| 0.4 | A-2 | Hot or cold instant drink — one person |
| 0.4 | K-2 | Making a bed against a wall, “duvet” edges folded under |
| 0.4 | K-3 | Making a freestanding bed, “duvet” edges folded under |
| 0.4 | L-1 | Folding a basket of laundry |
| 0.4 | L-3 | Loading and starting a washing machine |
| 0.4 | P-11 | Upper/lower body dressing — garments set out |
| X.X* | J-10 | Putting away clean dishes from a dishwasher |
| X.X* | K-7 | Making bed (with mattress on the floor against a wall) with standard sheets and blanket or “duvet” |
| X.X* | K-8 | Spreading bedding on the floor, Japanese-style — one person |
| 0.3 | M-1 | Setting a table — one or two persons |
| 0.3 | O-1 | Polishing shoes |
| 0.3 | S-2 | Feeding a cat — moist cat food and water |
| 0.3 | S-4 | Feeding a dog — moist dog food and water |
| 0.2 | A-5 | Single-cup espresso or coffee — two persons |
| 0.2 | C-1 | Cold cereal and beverage — one person |
| 0.2 | J-1 | Sweeping the floor |
| 0.2 | J-2 | Hand washing dishes |
| 0.2 | L-2 | Hand washing laundry |

(continued)

X.X* = task challenge is approximate; insufficient data available for stable calibration

**Table 15-5 AMPS Process Task Challege (continued)**

| | | |
|---|---|---|
| **Easier than average ADL tasks (continued)** | | |
| 0.2 | N-2 | Watering plants and removing dead leaves |
| 0.2 | P-3 | Upper body grooming/bathing |
| 0.2 | P-7 | Upper and lower body dressing — garments stored |
| **Average ADL tasks** | | |
| 0.1 | C-3 | Sour milk (“fil”), “kefir,” or yogurt with cereal and beverage |
| 0.1 | F-1 | Peanut butter and jelly sandwich — one person |
| 0.1 | F-2 | Presliced meat or cheese sandwich — one person |
| 0.1 | F-7 | Open-face meat or cheese sandwich with sliced vegetable — one person |
| 0.1 | F-8 | Open-face sandwich with soft spread and sliced vegetable — one person |
| 0.1 | F-9 | Jam sandwich — one person |
| 0.1 | F-12 | Presliced meat and/or cheese sandwich with vegetable(s) — one person |
| X.X* | F-13 | Open-face chocolate-hazelnut spread sandwich — one person |
| 0.1 | I-8 | Instant noodles, soup, or beans — one person |
| 0.1 | I-18 | Heating a precooked meal or dessert in a microwave — one or two persons |
| 0.1 | J-8 | Cleaning windows |
| 0.1 | K-4 | Changing standard sheets on a freestanding bed |
| 0.1 | L-5 | Ironing a shirt — ironing board already set up |
| 0.1 | L-7 | Hanging washed, spin-dried laundry to dry |
| 0.1 | M-2 | Setting a table, Swedish-style — four persons |
| 0.1 | P-10 | Showering |
| 0.1 | Q-1 | Sweeping outside |
| 0.0 | A-3 | Pot of boiled/brewed coffee or tea — one or two persons |
| X.X* | F-11 | Presliced meat or cheese with vegetable and beverage — two to four persons |
| 0.0 | H-1 | Cottage cheese and fruit salad — one person |
| 0.0 | I-17 | Heating a frozen meal or dessert in a microwave — one person |
| 0.0 | J-3 | Vacuuming, moving no furniture |
| 0.0 | J-4 | Vacuuming, moving lightweight furniture |
| 0.0 | J-5 | Mopping the floor |
| 0.0 | J-9 | Hand washing, drying, and putting away dishes |
| 0.0 | K-5 | Changing sheets and “duvet” cover on a bed against wall |
| 0.0 | K-6 | Changing sheets and “duvet” cover on a freestanding bed |
| 0.0 | L-4 | Ironing a shirt — setting up the ironing board |

(continued)

X.X* = task challenge is approximate; insufficient data available for stable calibration

**Table 15-5 AMPS Process Task Challege (continued)**

| | | |
|---|---|---|
| **Average ADL tasks (continued)** | | |
| 0.0 | P-8 | Upper body grooming and total body dressing |
| 0.0 | Q-2 | Raking grass cuttings or leaves |
| 0.0 | R-1 | Shopping |
| -0.1 | D-6 | Boiled egg(s) served in egg cup(s) |
| -0.1 | G-1 | Boiled/brewed coffee or tea and cookies/biscuits served at table ("fika") — two to four persons |
| -0.1 | G-2 | Boiled/brewed coffee or tea and cookies/biscuits served on a tray — two or three persons |
| -0.1 | I-3 | Fried ripe plantains |
| -0.1 | I-5 | Vegetable preparation — one to four persons |
| -0.1 | I-9 | Quick noodles cooked in a pot — one person |
| -0.1 | I-15 | Prepare tomato sauce for pasta — two persons |
| X.X* | I-20 | Gazpacho (blended, cold tomato soup) — two to four persons |
| -0.1 | N-1 | Repotting a small houseplant |
| -0.1 | Q-3 | Weeding |
| **Harder than average ADL tasks** | | |
| -0.2 | A-4 | Stove-top espresso coffee — one to four persons |
| -0.2 | B-1 | Toast and instant coffee, tea, instant soup, or hot chocolate — one person |
| -0.2 | B-2 | Toast and boiled/brewed coffee or tea — one person |
| -0.2 | C-2 | Hot cooked cereal and beverage — one person |
| -0.2 | F-4 | Grilled cheese sandwich and beverage — one person |
| -0.2 | F-5 | Open-face cheese or liverpaste sandwich on unsliced soft bread, and boiled/brewed coffee or tea — one or two person |
| -0.2 | F-6 | Open-face cheese or liverpaste sandwich on presliced soft bread, and boiled/brewed coffee or tea — one or two persons |
| -0.2 | F-10 | Grilled cheese on toast and beverage — one person |
| X.X* | H-5 | Tossed green salad with packaged lettuce and dressing, served in a large bowl — two to four persons |
| -0.2 | I-1 | Canned soup and crackers or presliced bread — one or two persons |
| -0.2 | I-2 | Fried green plantains ("tostones") |
| -0.2 | L-6 | Ironing multiple garments and putting garments away |
| -0.3 | F-3 | Tuna, chicken, or crab salad sandwich — one person |
| -0.3 | H-2 | Fresh fruit salad — two persons |

(continued)

X.X* = task challenge is approximate; insufficient data available for stable calibration

**Table 15-5 AMPS Process Task Challege (continued)**

| | | |
|---|---|---|
| **Harder than average ADL tasks (continued)** | | |
| -0.3 | H-4 | Green salad, served in large bowl, with dressing on side — two or three persons |
| -0.3 | I-4 | Beans and toast — one person |
| -0.3 | J-6 | Vacuuming two rooms on different levels |
| -0.3 | Q-4 | Vacuuming the inside of an automobile |
| -0.4 | C-4 | Hot cooked cereal, open-face cheese sandwich, and beverage — one person |
| -0.4 | D-1 | Scrambled or fried eggs, toast, and beverage — one person |
| -0.4 | D-3 | Scrambled or fried eggs, meat, and boiled/brewed coffee or tea — one person |
| -0.4 | E-1 | French toast and beverage |
| -0.4 | H-3 | Tossed salad with dressing, served in individual bowls — two or three persons |
| -0.4 | I-14 | Pasta with sauce and beverage — two persons |
| -0.4 | I-16 | Pasta with meat, sauce, and beverage — two persons |
| -0.4 | I-21 | Miso soup — one or two persons |
| -0.4 | J-7 | Cleaning a bathroom |
| X.X* | Q-5 | Repair a bicycle tube puncture or hole |
| **Much harder than average ADL tasks** | | |
| -0.5 | D-2 | Scrambled or fried eggs, toast, and boiled/brewed coffee or tea — one person |
| -0.5 | D-4 | Omelette or scrambled eggs with added ingredients, toast, and beverage — one person |
| -0.5 | D-5 | Scrambled or fried eggs, toast, and espresso coffee — one person |
| -0.5 | D-7 | Spanish omelette with added ingredients — two persons |
| -0.5 | E-2 | Thin (e.g., European) pancakes and beverage — one to three persons |
| -0.5 | E-3 | Thick (e.g., American) pancakes and beverage — one to three persons |
| -0.5 | G-3 | Cake, muffins, or brownies |
| -0.5 | I-10 | Fried meat and vegetable dish with a bowl of rice — one person |
| -0.5 | I-11 | Fried rice — one person |
| X.X* | I-22 | Rice, soup, and a side dish — one person |
| -0.6 | I-6 | Pasta with sauce, green salad, and beverage |
| -0.6 | I-13 | Vegetable soup — one to four persons |
| -0.7 | I-7 | Pasta with meat, sauce, green salad, and beverage |
| -0.7 | I-12 | Vegetable soup, vegetables sautéed — one to four persons |
| -0.7 | I-19 | Meatballs with boiled potatoes, sauce, boiled vegetable, and beverage — two to four persons |

X.X* = task challenge is approximate; insufficient data available for stable calibration

### 15.2.1 Goodness of Fit of the AMPS Skill Items and Tasks

When the data for the current standardization sample were subjected to many-faceted Rasch analyses, the results revealed that all of the ADL motor and ADL process skill items demonstrate acceptable goodness of fit to the many-faceted Rasch model of the AMPS. Similarly, of the 102 ADL tasks with sufficient data to calibrate, all demonstrated acceptable goodness of fit. The criteria for acceptable goodness of fit were set at $MnSq \leq 1.4$ and/or $z < 2$. These ***results support the unidimensionality of the AMPS scales***.

### 15.2.2 Additional Evidence to Support Unidimensionality of the AMPS Scales

The results of principal-components factor analyses have further supported the unidimensionality of the AMPS scales, one defined by the ADL motor items and one defined by the ADL process items. As the skill item Paces loaded on both the ADL motor and ADL process factors, it is included in both AMPS scales.

## 15.3 Current Evidence for Validity Between Genders

### 15.3.1 Equivalence of Current Normative Means Between Genders

To confirm that the normative values for ADL ability included in the AMPS computer-scoring software can be validly used with both men and women, we evaluated the standardization sample for differences between healthy, well men and women with regard to mean ADL motor or ADL process ability. More specifically, two gender by age group analyses of variance (ANOVA) for the AMPS ***normative well sample*** (see Section 15.1.2 and Tables 15-1 and 15-2) revealed no significant gender main effect for ADL motor ability ($F[1,12727] = 1.339$, $p = .247$), but a significant gender main effect for ADL process ability ($F[1,12727] = 98.232$, $p < .001$). While there was a significant main effect for ADL process ability, the absolute differences in mean ADL motor and ADL process ability by age group (see Table 15-6) revealed that the maximum difference between gender groups was only 0.12 logit (3-year-olds) on the AMPS motor scale and only 0.20 logit on the ADL process scale (14-year-olds). These values are below or equal to the mean standard errors (*SE*s) for their respective scales (see Section 15.13.5), and support that there are no practical, clinically

**Table 15-6 Gender Differences in Mean ADL Motor and Mean ADL Process Abilities (logits) for Well Persons in the AMPS Standardization Sample**

| Age group (yrs) | ADL motor ability | | | ADL process ability | | |
|---|---|---|---|---|---|---|
| | Female | Male | Absolute difference | Female | Male | Absolute difference |
| 3 | 0.97 | 0.85 | 0.12 | 0.07 | -0.07 | 0.14 |
| 4 | 1.16 | 1.12 | 0.04 | 0.29 | 0.22 | 0.08 |
| 5 | 1.31 | 1.26 | 0.05 | 0.49 | 0.39 | 0.10 |
| 6 | 1.45 | 1.39 | 0.06 | 0.62 | 0.56 | 0.07 |
| 7 | 1.58 | 1.57 | 0.01 | 0.75 | 0.67 | 0.08 |
| 8 | 1.67 | 1.65 | 0.02 | 0.84 | 0.75 | 0.08 |
| 9 | 1.84 | 1.77 | 0.07 | 0.99 | 0.90 | 0.08 |
| 10 | 1.87 | 1.89 | 0.02 | 1.04 | 0.95 | 0.09 |
| 11 | 1.98 | 2.03 | 0.05 | 1.09 | 1.07 | 0.02 |
| 12 | 2.11 | 2.14 | 0.03 | 1.17 | 1.10 | 0.08 |
| 13 | 2.20 | 2.17 | 0.03 | 1.25 | 1.15 | 0.10 |
| 14 | 2.27 | 2.23 | 0.05 | 1.38 | 1.18 | 0.20 |
| 15 | 2.36 | 2.31 | 0.05 | 1.40 | 1.20 | 0.19 |
| 16 to 29 | 2.96 | 2.92 | 0.04 | 2.12 | 1.94 | 0.18 |
| 30 to 39 | 3.02 | 2.96 | 0.06 | 2.19 | 2.08 | 0.11 |
| 40 to 49 | 2.83 | 2.94 | 0.11 | 2.01 | 2.05 | 0.03 |
| 50 to 59 | 2.65 | 2.69 | 0.04 | 1.91 | 1.82 | 0.09 |
| 60 to 61 | 2.37 | 2.37 | 0.00 | 1.73 | 1.63 | 0.10 |
| 62 to 64 | 2.30 | 2.33 | 0.03 | 1.69 | 1.61 | 0.08 |
| 65 to 69 | 2.24 | 2.26 | 0.02 | 1.65 | 1.60 | 0.04 |
| 70 to 74 | 2.13 | 2.17 | 0.04 | 1.64 | 1.58 | 0.06 |
| 75 to 79 | 2.03 | 2.06 | 0.03 | 1.53 | 1.49 | 0.04 |
| 80 to 103 | 1.90 | 1.89 | 0.01 | 1.51 | 1.49 | 0.03 |

observable differences in mean ADL motor or ADL process ability between men and women as long as standardized administration procedures are utilized and both males and females perform ADL tasks that are familiar and chosen. The finding of no practical gender differences in mean ADL motor or ADL process ability ***supports the use of gender-combined norms*** when evaluating a person's performance in relation to age expectations. See Section 15.9, Table 15-9 for more information about the development of the normative ADL motor and ADL process values by age group.

### 15.3.2 Differences in ADL Ability Between Men and Women With and Without Disabilities

Further supporting the equivalence of ADL ability between genders is evidence that, despite statistical differences, any differences in ADL ability between genders is not clinically meaningful, even among persons with disabilities. For example, Merritt and Fisher (2003) found that the mean differences in ADL ability between men ($n$ = 9,250) and women ($n$ = 9,250) (matched for age, functional level, and diagnostic category) was ≤ 0.07 logit on the ADL motor scale and ≤ 0.13 logit on the ADL process scale. Similarly, in a study evaluating differences in a sample of persons with right- or left-hemispheric stroke, Rexroth, Fisher, Merritt, and Gliner (2005) found that men had significantly higher mean ADL motor ability ($p < 0.05$, $d = -0.17$) and women had significantly higher ADL process ability ($p \leq 0.01$, $d = 0.09$); however, these differences were not large enough to be clinically observable.

### 15.3.3 Equivalence of Item and Task Calibration Values Between Genders

While the results of an initial study suggested that the motor skill item Lifts may demonstrate differential item functioning (DIF) between men and women, the presence of possible DIF on the AMPS motor scale did not result in differential test functioning (Duran & Fisher, 1996). Moreover, later evidence based on a larger sample, revealed not only an absence of DIF, but also an absence of differential task functioning associated with gender (Merritt & Fisher, 2003).

To further confirm that the AMPS skill item and task calibration values included in the AMPS computer-scoring software can be validly used with both men and women, we evaluated data for the entire standardization sample (see Section 15.1.2) for evidence of differential item or task functioning between men and women. When the ADL motor and ADL process skill items were evaluated for presence of DIF, the maximum absolute difference between gender-specific item calibration values was 0.29 logit on the ADL motor scale and 0.14 logit on the ADL process scale. When the tasks were evaluated for differential task functioning by gender, the maximum absolute difference in task calibrations by gender were 0.05 logit on the ADL motor scale and 0.07 on the ADL process scale. These results support (a) the absence of both differential item and task functioning, and (b) the use of gender-combined item and task calibration values without risk of test bias associated with gender.

**The AMPS is free of gender bias, as indicated by:**

- **No practical (i.e., observable) differences in mean ADL motor or ADL process ability among well males and females**
- **No differential item functioning associated with gender**
- **No differential task functioning associated with gender**

**These results support the validity of using**

- **Gender-combined norms when interpreting the results of an AMPS observation**
- **Gender-combined item and task calibration values in the AMPS computer-scoring software to calculate a person's ADL motor or ADL process ability measures**

#### 15.3.4 Relevance of AMPS Tasks for Men and Women

Despite questions from occupational therapists about the relevance of household and cooking tasks for men, analysis of the data has revealed that only a few AMPS tasks were performed more often by either men or women. More specifically, when Merritt and Fisher (2003) compared the frequency with which each AMPS task was performed by males versus females *n* = 9,250 men and 9,250 women), for 70% of the tasks then included in the AMPS, at least 40% of the persons who performed each task were men. This finding suggests that there most likely are AMPS tasks that are relevant for men to perform for an AMPS observation. See Chapter 6, Section 6.6 for more information related to the determination of whether or not the AMPS is appropriate for a specific person.

## 15.4 Current Evidence for Equivalence of Item and Task Calibration Values Among World Regions

Two earlier studies compared item difficulty and task challenge calibrations across world regions (i.e., examined for differential item and task functioning). In the first of these studies, Magalhães, Fisher, Bernspång, and Linacre (1996) compared matched (age, gender, and diagnosis) groups of North American (*n* = 324) and Swedish (*n* = 265) persons who performed familiar and culture-relevant AMPS tasks. They found that, except for the motor skill item Endures, the ADL motor and ADL process skill item difficulty calibrations remained stable across these two world regions.

In the second study, Goldman and Fisher (1997) replicated and extended the study by Magalhães et al. (1996) by examining for differential item and task functioning among 840 persons from North America, 2150 persons from Scandinavia, and 836 persons from the United Kingdom. Goldman and Fisher found that none of the ADL motor and ADL process skill items and none of the "culture-general" tasks (tasks that are commonly performed in at least two of the three world regions studied) displayed either differential item or task functioning.

While the results of these two earlier studies suggested that AMPS data from North America, Scandinavia, and the United Kingdom can be combined to create a single version of the AMPS that can be used for international comparisons, we have now been able to compare six world regions using data from the current AMPS standardization sample (see Section 15.1.2). More specifically, the following six world regions had sufficient data to examine for differential item and task functioning: North America, United Kingdom and Republic of Ireland, Nordic countries, other European countries, Australia and New Zealand, and Asian countries. For the ADL motor skill items, a total of 240 comparisons were made (i.e., 15 region pairs [for example, North America vs. United Kingdom and Republic of Ireland, North America vs. Nordic countries, North America vs. other European countries, etc.] were compared for each of 16 items = 15 region pairs X 16 items = 240), and only one comparison exceeded the 0.50 logit value that signals the possible presence of DIF. That is, the region-specific item calibration values for Asia versus Australia and New Zealand differed by 0.52 logit; Endures calibrated as being easier for persons from Asia (1.01 logits) than for those from Australia and New Zealand (0.49 logit). Upon further examination, we found that the presence of DIF between Asia versus Australia and New Zealand did not result in differential test function (i.e., biased estimates of ADL motor ability). When the ADL process skill items were examined for the presence of DIF by region, none of the 300 comparisons (15 region comparisons for each of 20 ADL process skill items) exceeded the 0.50 logit criterion. Finally, among the ADL tasks included in the AMPS, 26 are widely used across all six regions listed above, and none of these 26 ADL tasks displayed differential task functioning. Considered together, these results suggest that the AMPS is free of bias associated with global world region, and that the combined international item and tasks calibrations are valid for cross-regional comparisons.

**The AMPS is free of differential item and differential task functioning by world region, supporting the use of combined international item and task calibration values when calculating a person's ADL motor and ADL process ability measures.**

## 15.5 Current Evidence for Validity Among Ethnic or Cultural Groups

### 15.5.1 Differential Item and Task Functioning Between Black and White Americans

Stauffer, Fisher, and Duran (2000) examined if the AMPS is a valid, non-biased tool when used with Black and White Americans. Their sample consisted of 466 Black and 466 White Americans matched for diagnosis, age, functional level, and gender. When comparing the stability of the ADL motor and ADL process skill item calibrations between Black and White Americans, it was found that the hierarchies remained stable, as none of the skill item calibrations differed by more than 0.43 logit. Similarly, among the 35 AMPS task challenge calibrations that were compared, none of the ADL motor or ADL process task challenge calibrations differed between the two groups. Such findings provide initial evidence that the ***item difficulty and the task challenge hierarchies are free of differential item and task functioning between White and Black Americans***, and thus valid ADL ability measures can be generated for these two groups.

### 15.5.2 Differences in ADL Ability Between Black and White Americans

Stauffer et al. (2000) also evaluated for differences in ADL ability between Black and White Americans, and found that while their mean ADL motor abilities did not differ significantly, the Black Americans had significantly lower ADL process ability (0.2 logit). The authors speculated that differences in health status rather than test bias were the reason for the difference in mean ADL process ability between the two groups, however additional research is needed to verify this speculation.

### 15.5.3 Relevance of AMPS Tasks and Rater Reliability When the AMPS is Applied Cross-culturally

Several studies have focused on examining the validity of calibrating culture-specific tasks into the AMPS measurement model or, in other words, have focused on verifying the validity of the AMPS scales and ADL ability measures when the AMPS is applied cross-culturally. Dickerson and Fisher (1995) found that the four ***Cuban-American tasks*** included in Volume 2, Chapter 2, AMPS Task Descriptions demonstrated good fit to the measurement model and could be calibrated with other existing North American tasks in the AMPS system (***scale validity***). Moreover, ***Cuban-born persons*** living in the United States demonstrated ***person response validity*** when they performed North American as well as Cuban-American tasks.

Similar results were obtained by Bernspång and Fisher (1995b) and Clawson (1995) when the AMPS was used to evaluate ***Swedish or Mexican-born persons***, respectively. An interesting finding, however, emerged from Clawson's (1995) use of the AMPS with Mexican-born persons living in the United States. A high proportion of her participants had unexpectedly low scores on the process skill item Uses. In all cases, these unexpectedly low scores were assigned according to the scoring criteria for the process skill item Uses included in Volume 2, Chapter 8, AMPS Skill Items. For example, two participants used knives to open cans. Others spread butter or margarine with spoons or large rubber spatulas. In fact, these participants demonstrated acceptable fit to the measurement model only when the skill item Uses was eliminated from the analysis. While the inclusion of the skill item Uses in the computation of ADL ability measures did not have a significant effect on the estimation of their overall ADL process ability (i.e., did not disrupt the measurement system; no differential test functioning was present), this observation clarifies ***how important it is for the AMPS rater to be aware of cultural norms related to tool use if the AMPS is to be scored in a valid and reliable manner***. Moreover, if future research verifies that the use of knives to open cans is an accepted practice in Mexican culture, Mexican-born persons should not be given lower scores on the process skill item Uses for using knives to open cans.

Goto, Fisher, and Mayberry (1996) examined the ***cross-cultural validity*** of the AMPS as well as the ***reliability of raters*** from different cultures when rating Japanese persons performing AMPS ADL tasks. They found a 95% overall rate of acceptable goodness of fit of the Japanese participants to the existing AMPS many-faceted Rasch model defined by persons from North America, Europe, and Australasia. The six raters from the United States, Sweden, England, and Japan, also demonstrated acceptable goodness of fit and their overall rate of misfitting ratings was less than 2.5% ($t \leq 3$), indicating good interrater reliability. The Japanese rater and one American rater, however, varied significantly in rater severity. Further examination of the raw data revealed that the non-Japanese ("Western") raters tended to be stricter when they rated the Japanese participants than when they rated "Western" persons during AMPS training and calibration courses.

**The finding that there may be risk for cross-cultural bias when AMPS raters score the AMPS skill items (Clawson, 1995; Goto et al., 1996) underscores the importance of each rater being aware of his or her own cultural values, and the need for all raters to become familiar with cultural differences in how people perform ADL tasks.**

**Whenever a rater observes and scores the performance of a person from a culture different than his or her own, the rater must**

- **Ensure that both the rater and the person are fully familiarized with the tasks and equipment to be used during the evaluation**
- **Be familiar with what is considered logical, efficient, and usual by the person's cultural group**
- **Be knowledgeable enough to recognize legitimate cultural variations in ADL task performance**

## 15.6 Current Evidence for Validity and Differences Among Diagnostic Groups

### 15.6.1 Differential Item and Task Functioning Between Persons with Right Versus Left Hemispheric Stroke

Two early studies were conducted to evaluate for DIF between persons with right versus left stroke (Bernspång & Fisher, 1995a; Solomon, 1995). Bernspång and Fisher examined the data of 71 persons with left hemispheric stroke and 71 persons with right hemispheric stroke, and Solomon evaluated the data of 391 persons with right hemispheric stroke and 386 persons with left hemispheric stroke. Bernspång and Fisher found that the ADL process skill item difficulty calibrations remained stable between the two groups, while those with right hemispheric stroke experienced the skill items Paces, Transports, and Coordinates as more difficult, and the group with left hemispheric stroke experienced the skill item Calibrates as more difficult. In contrast, with a larger sample, Solomon found that only the process skill item Uses differed meaningfully between the two groups. The process skill item Uses was relatively easier for persons with right hemispheric stroke.

In a more recent study with an even larger sample (Rexroth et al., 2005), researchers found that there were no significant differences in ADL skill item calibrations between persons with right hemispheric stroke ($n$ = 1,939) and left hemispheric stroke ($n$ = 1,939). Thus, the ADL motor and ADL process skill hierarchies were found to be stable, indicating that persons with right and left stroke have similar abilities when performing all of the goal-directed actions (ADL skills) defined within the AMPS. This study provides evidence that the ability to perform ADL actions is the same despite known differences in underlying cognitive and physical impairments (e.g., hemiparesis, apraxia, perceptual disorders). As such, ***occupational therapists are cautioned against making assumptions*** of what a person can or cannot do based on the side of hemispheric stroke, and should therefore, always assess each person at the level of occupation (versus impairment).

### 15.6.2 Differences in Mean ADL Ability Between Persons with Right Versus Left Hemispheric Stroke

In addition to evidence supporting the absence of DIF between persons with right or left hemispheric stroke, there is evidence to support that there are no significant differences in mean ADL motor or ADL process ability between the two groups, despite hemisphere-specific differences in underlying cognitive and physical impairments (Bernspång & Fisher, 1995a; Solomon, 1995). While Rexroth et al. (2005) found that persons with right hemispheric stroke had significantly higher mean ADL motor and ADL process ability measures, the magnitude of the effect was small, and thus the researchers concluded that the differences were not clinically meaningful.

**Despite differences in underlying impairments,**

- **The relative challenge of the ADL motor and ADL process skill items do not differ between persons with right versus left hemispheric stroke**
- **Persons with right versus left hemispheric stroke do not differ in a clinically meaningful manner in mean ADL motor or ADL process ability**

**Thus, regardless of the side of the lesion, the observed ADL skill challenges and strengths are similar between groups of persons with right and left hemispheric stroke.**

### 15.6.3 Differences in Skill Item Difficulty Between Well Persons and Persons with Dementia

Analysis of data from 66 persons with mild-to-moderate ***dementia of the Alzheimer's type*** (DAT) and a matched (age and gender) group of well individuals revealed significant differences in item difficulty calibrations between the two groups. Persons with specific memory and related cognitive deficits associated with DAT had relatively intact abilities on ADL process skill items related to ***knowing how*** (e.g., Uses, Handles, Organizes, Gathers), but relatively lower abilities on process skill items related to ***knowing what*** (e.g., Heeds, Inquires, Chooses, Initiates, Benefits, Searches/Locates, Paces) (Oakley, Fisher, Sunderland, & Hill, 1992).

Another study involved a comparison of 341 persons with ***DAT*** and 287 well, older persons. In this study, Cooke, Fisher, Mayberry, and Oakley (2000) compared actual differences in ADL process skill item difficulty calibrations between the two groups. Actual item difficulty calibrations were calculated by adjusting the relative difficulties of the items to account for overall group differences in mean ADL process ability. In support of the results of the earlier study by Oakley et al. (1992), Cooke et al. (2000) found that persons with DAT experienced ADL process skill items related to ***knowing what*** (e.g., Heeds, Inquires, Chooses, Benefits, Notices/Responds) as more difficult, and experienced ADL process skill items related to ***knowing how*** (e.g., Uses, Gathers, Organizes, Handles, Navigates) as less difficult. By comparing the actual item difficulty calibrations, Cooke et al. found that Handles, Organizes, and Navigates appeared to remain as "intact" in persons with DAT as in nondisabled persons despite an overall loss in ADL process ability among persons with DAT.

In a more recent study, Oakley, Duran, Fisher, and Merritt (2003) examined the differences in actual ADL motor skill item calibrations between (a) well older adults who were living independently in the community ($n$ = 378), (b) persons with DAT who required minimal assistance to live in the community ($n$ = 189), and (c) persons with DAT who required moderate to maximal assistance to live in the community ($n$ = 378). When comparing actual ADL motor skill item calibrations, the researchers determined that the group of persons with DAT who required minimal assistance to live in the community demonstrated more effort with the ADL motor skill items Paces and Flows than did the group of well older adults. In contrast, the group with DAT who required moderate to maximal assistance to live in the community had more difficulty performing all of the ADL motor skill items ***except*** Endures, Calibrates, and Grips. Furthermore, the ADL motor skills that were most adversely impacted within the group with DAT requiring moderate to maximal assistance were Paces, Flows, Walks, Transports, and Coordinates.

**While DAT is characterized as a cognitive disorder, person with DAT often demonstrate increased clumsiness or physical effort when performing ADL tasks, not just decreased efficiency. Such evidence serves to remind occupational therapists who work with persons with DAT that it is important to observe, document, and subsequently plan interventions that are geared toward improving both ADL motor and ADL process ability.**

### 15.6.4 Differences in Mean ADL Ability Among Well Persons, Persons with Memory Impairments, and Persons with Dementia

Oakley et al. (2003) found that mean ADL motor ability differed significantly among persons with DAT who require minimal assistance, persons with DAT who require moderate assistance, and well, independent persons ($p < 0.05$). Moreover, ***the DAT group who only required minimal assistance to live in the community demonstrated significantly lower mean ADL motor ability measures than did their healthy, well peers***. This evidence supports the sensitivity of the AMPS to detect differences in ADL ability among groups.

Additional support for the sensitivity of the AMPS ability measures was provided by the results of a study that compared older adults with ***dementia***, persons with ***age-associated memory impairment*** (AAMI), and ***community-living, nondisabled*** older adults (Robinson, 1994). The three groups were found to differ significantly in mean ADL motor ability. The community-living nondisabled older adults and the group with AAMI did not differ significantly in ADL process ability; however both of these groups were significantly more able than was the group with dementia.

Doble, Fisher, Fisk, and MacPherson (1992) also examined the validity of the AMPS when used to evaluate older persons with ***dementia***. They found statistically significant differences in ADL motor and ADL process ability between persons with dementia and an age-matched group of community-living older persons. In a subsequent study, Doble and her colleagues (Rockwood, Doble, Fisk, MacPherson, & Lewis, 1996) compared 39 persons with ***DAT*** to a group of 24 persons with ***cognitive impairments, but no dementia***, and to a group of 48 persons (well and frail) with no cognitive impairments. The group with DAT had significantly lower mean ADL motor ability than either of the other two groups, and the two groups without dementia did not differ significantly in mean ADL motor ability. In contrast, all three groups differed significantly in mean ADL process ability; the group with DAT had the lowest mean ADL process ability and the group with no cognitive impairments the highest.

**Persons with dementia of the Alzheimer type:**

- **Demonstrate lower observable occupational skill in relation to actions dependent on "knowing what" to do (e.g., finding and choosing correct task objects, performing the task specified)**
- **Demonstrate skill equal to their well, age-matched peers in relation to "knowing how" to interact with task objects (e.g., using, handling, and organizing task objects)**
- **Have significantly lower mean ADL motor and ADL process ability measures than their healthy peers**

### 15.6.5 Differences in Skill Item Difficulty Between Well Persons and Persons with Multiple Sclerosis

A comparison of 80 persons with ***multiple sclerosis*** and a matched (age and gender) group of well individuals revealed that persons with multiple sclerosis have (a) relatively intact ability on the ADL motor skill items ***primarily related to obtaining and holding objects*** (e.g., Coordinates, Grips, Positions, Manipulates, Calibrates); (b) relatively lower ability on the ADL motor skill items related to ***body position*** and ***mobility*** (e.g., Walks, Transports, Stabilizes); (c) relatively intact ability on the ADL process skill items related to ***knowing what*** (Heeds, Inquires, Chooses, Initiates, Accommodates); and (d) relatively lower ability on the ADL process skill items related to ***knowing how and the details of task performance*** (Uses, Handles, Organizes, Gathers, Attends, Notices/Responds) (Harrison, Johansson-Charles, & Doble, 1993).

### 15.6.6 Differences in Mean ADL Ability Between Well Persons and Persons with Multiple Sclerosis

Harrison et al. (1993) and Doble, Fisk, Fisher, Ritvo, and Murray (1994) found that persons with multiple sclerosis had significantly lower mean ADL motor and ADL process ability measures than did the persons in their matched well samples. When persons with multiple sclerosis were divided into three groups based on overall level of independence (live independently in the community: $n = 25$; live in the community with minimal assistance: $n = 32$; live in the community with moderate to maximum assistance: $n = 22$), all three groups differed significantly from each other in mean

ADL motor and ADL process ability measures (Harrison et al., 1993). These results again support the sensitivity of the AMPS ability measures for detecting differences in ADL ability among groups.

### 15.6.7 Differences in Skill Item Difficulties Between Persons with Mild and Moderate Intellectual Disabilities (Mental Retardation)

Kottorp, Bernspång, and Fisher (2003) conducted a study comparing the relative and actual skill item hierarchies between persons with mild intellectual disabilities ($n$ = 178) and moderate intellectual disabilities ($n$ = 170). When examining the stability of the ***relative*** ADL motor and ADL process skill item hierarchies between the two groups (e.g., DIF), it was found that only two skill items were significantly different (Endures and Uses), however these slight differences did not disrupt the measurement model, and thus did not impact the validity of generating ADL motor and ADL process ability measures for these two groups (i.e., the AMPS measures were free from differential test functioning).

When comparing ***actual*** skill item differences (e.g., skill item calibrations which take into account mean ADL ability differences), Kottorp et al. (2003) found that the skill items Endures, Manipulates, Calibrates, Coordinates, Uses, Navigates, Notices/Responds, and Terminates were equally difficult between the persons with mild and moderate intellectual disabilities, despite differences in overall ADL motor or ADL process ability. The remaining ADL motor and ADL process skill items were more difficult for those with moderate intellectual disabilities.

### 15.6.8 Differential Item Function Among Persons with Different Types of Psychiatric Disorders

A recent study (Moore, Merritt, & Doble, 2010) examined the stability of the skill item hierarchies between age-matched groups of persons who have (a) schizophrenia ($n$ = 200), (b) bipolar disorder depressed episode ($n$ = 158), or (c) bipolar disorder manic episode ($n$ = 200). Only one skill item demonstrated DIF among the three groups: the process skill item Attends was more difficult for the bipolar manic group than for the bipolar depressed group. The presence of DIF for the skill item Attends did not disrupt the estimation of the persons' ADL process skill ability measures, thus providing evidence that the AMPS is free of differential test functioning and can be used without bias.

- **Persons with schizophrenia, bipolar depressed episode, and bipolar manic episode demonstrate very similar ADL motor and ADL process skill item profiles**
- **There is no evidence to support test bias among these three groups**

### 15.6.9 Differences in Mean ADL Ability Among Groups with Psychiatric Disorders, and Between Well Persons and Persons with Psychiatric Disorders

Pan and Fisher (1994) found that persons with psychiatric disorders receiving occupational therapy services through inpatient and partial hospitalization programs ($n$ = 30) had significantly lower mean ADL motor and ADL process ability measures than did a matched (age, ethnicity, and gender) group of well persons. Moreover, when Girard, Fisher, Short, and Duran (1999) compared the ADL motor and ADL process abilities of persons with schizophrenia, persons with depression, and nondisabled persons to test the hypothesis that ADL motor and ADL process abilities fall along a continuum of ability, they found significant differences in mean ADL motor and mean ADL process ability across all three groups ($p \leq 0.05$). More specifically, the healthy group's ADL motor and ADL process abilities were the highest, the group with schizophrenia had the lowest ADL ability, and the group with depression was in the middle.

When evaluating individual differences in ADL ability, however, Girard et al. (1999) noted that there was overlap, such that there were incidences when a more able person with schizophrenia had similar ability measures as those who were healthy and/or depressed. Thus, ***while the differences in mean ADL motor and ADL process may document a trend or continuum of decreasing ability from persons who are healthy to persons with depression to persons with schizophrenia, mean differences do not fully reflect within group variations in ADL ability***. The researchers thus caution against making assumptions or predictions of ADL ability based on the person's diagnosis. Rather, care should be taken to observe and document each person's strengths and problems of occupational performance.

Finally, in a more recent study with a larger sample, Moore et al. (2010) compared overall ADL ability across three psychiatric diagnoses (schizophrenia, bipolar manic episode, and bipolar depressed episode). They found no clinically meaningful differences in mean ADL motor or mean ADL process ability between the three groups.

- **Persons with psychiatric disorders have significantly lower ADL motor and ADL process abilities than do their age-matched healthy, well peers**
- **Persons with schizophrenia have lower mean ADL motor and ADL process abilities than do persons with depression**
- **Persons with schizophrenia, bipolar manic episodes, and bipolar depressed episode do not differ in mean ADL motor or ADL process abilities**

### 15.6.10 Differences in Mean ADL Ability Among Other Diagnostic Groups

Women with ***systemic lupus erythematosas*** (SLE) ($n$ = 15) had significantly lower mean ADL motor and ADL process ability measures than did their healthy, well peers ($n$ = 15), $t$ = -4.83, $p < 0.001$ and $t$ = -4.72, $p < 0.001$, for ADL motor and ADL process ability, respectively (Poole, Atanasoof, Pelsor, & Sibbitt, 2006). The authors stated that "the AMPS appears to be a sensitive instrument of functional ability in SLE even with minimal disease activity and normal levels of cognition" (p. 657).

Another study focused on the ADL ability of persons with ***chronic heart failure*** (CHF), ranging in age from 66 to 91 years of age ($n$ = 40) (Norberg, Boman, & Löfgren, 2008). The results supported a relationship between severity of symptoms of CHF such that those with little or no limitations in physical activities had significantly higher ADL motor and ADL process ability measures than did those who had marked difficulty performing physical activities or were unable to perform physical activities without discomfort. The mean ADL motor and ADL process ability measures of the persons with CHF were also below healthy age expectations, with significant differences being found in the three older age groups (71–75 years, 76–82 years, and

83–93 years). Evidence thus suggests that the AMPS can be used to detect changes in the quality of ADL task performance for persons with CHF; such findings may be useful in documenting the need for services to promote safe and efficient ADL task performance.

### 15.6.11 Current Evidence for Equivalence of Task and Item Calibration Values Across Diagnostic Groups

Many of the studies presented in Section 15.6 provide evidence that the AMPS skill items display DIF among diagnostic groups, but when evaluated further, such DIF has not been shown to result in differential test functioning. The presence of DIF can be viewed as evidence that the AMPS item hierarchy profiles are sensitive indicators of diagnostic differences that can inform clinical practice. Conversely, the presence of DIF can be a threat to the AMPS measurement system if it can be shown to result in differential test functioning and test bias. Therefore, to implement a more comprehensive investigation into such a threat, we examined for both differential task and differential item functioning among the global diagnostic groups listed in Section 15.1.2.

When we examined for ***differential task functioning***, there were 26 tasks that had been widely used across the six world regions compared in Section 15.4. None of these 26 AMPS tasks displayed differential task functioning by diagnostic group on either the ADL motor or ADL process scales of the AMPS.

In contrast, when we examined for ***DIF***, 483 of 2176 (22.2%) paired diagnostic comparisons for each ADL motor skill item, and 146 of 2720 (5.4%) paired diagnostic comparisons for each ADL process skill item displayed DIF; 5% might be expected by chance alone. While the level of DIF among the ADL process skill items was close to expectations, the level of DIF among the ADL motor skill items clearly exceeded expectations. The most common ADL motor skill items to display DIF were Walks, Calibrates, Paces, and Endures, and only Aligns and Moves displayed DIF less than 5% of the time. Among the ADL process skill items, 7 of 20 skill items displayed DIF more than 5% of the time (Attends, Chooses, Handles, Heeds, Navigates, Paces, and Searches/Locates), and among those, the majority of DIF was associated with Navigates.

The ADL motor and ADL process skill item calibration values by diagnostic group are shown in Tables 15-7 and 15-8. Skill item calibration values that differ by

**Table 15-7 ADL Motor Item Difficulty Calibration Values by "Diagnostic" Group**

| ADL motor skill | "Diagnostic" group* | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total sample | Well | Old at risk/ frail | Mild | DD | RCVA | LCVA | BI | MS | Other neur | Schiz/ thought | Other psych | MR | Mem/ dem | SCI | RA/ OA | Other mskl | Med/ sens | Mult/ not known |
| Endures | 0.63 | 1.10 | 0.37 | 1.20 | 1.46 | 0.93 | 0.93 | 0.96 | 0.82 | 0.76 | 0.39 | 0.31 | 0.92 | 0.58 | 1.15 | 0.48 | 0.32 | -0.16 | 0.49 |
| Lifts | 0.40 | 0.10 | 0.18 | 0.34 | 0.56 | 0.64 | 0.55 | 0.55 | 0.39 | 0.53 | 0.57 | 0.47 | 0.51 | 0.44 | 0.07 | -0.14 | 0.06 | 0.25 | 0.43 |
| Aligns | 0.40 | 0.74 | 0.30 | 0.73 | 0.30 | 0.42 | 0.54 | 0.44 | 0.19 | 0.41 | 0.43 | 0.47 | 0.45 | 0.31 | 0.18 | 0.42 | 0.25 | 0.28 | 0.32 |
| Moves | 0.36 | 0.15 | 0.46 | 0.19 | 0.23 | 0.34 | 0.39 | 0.46 | 0.33 | 0.38 | 0.63 | 0.59 | 0.48 | 0.49 | 0.18 | 0.25 | 0.18 | 0.30 | 0.35 |
| Transports | 0.18 | 0.54 | 0.33 | 0.67 | 0.25 | -0.11 | -0.06 | 0.21 | 0.02 | 0.07 | 0.65 | 0.60 | 0.54 | 0.26 | 0.13 | 0.02 | -0.04 | 0.12 | 0.08 |
| Flows | 0.13 | 0.37 | 0.43 | -0.07 | -0.19 | 0.06 | -0.04 | -0.05 | -0.12 | 0.04 | -0.02 | -0.03 | 0.11 | 0.05 | 0.44 | 0.21 | 0.36 | 0.40 | 0.15 |
| Grips | 0.11 | -0.50 | 0.01 | -0.41 | -0.01 | 0.15 | 0.11 | 0.06 | 0.34 | 0.19 | -0.12 | -0.08 | -0.13 | 0.17 | 0.15 | -0.03 | 0.27 | 0.35 | 0.30 |
| Reaches | 0.07 | 0.22 | -0.23 | 0.60 | 0.25 | 0.10 | 0.09 | 0.23 | 0.02 | 0.12 | 0.37 | 0.32 | 0.36 | 0.14 | -0.32 | -0.28 | -0.29 | -0.06 | -0.03 |
| Bends | -0.06 | 0.52 | -0.64 | 0.72 | 0.17 | -0.07 | -0.02 | -0.03 | -0.18 | -0.10 | 0.08 | 0.09 | 0.14 | -0.14 | -0.37 | -0.42 | -0.38 | -0.26 | -0.21 |
| Manipulates | -0.11 | -0.54 | 0.01 | -0.77 | -0.50 | -0.18 | -0.27 | -0.17 | 0.03 | -0.07 | -0.35 | -0.27 | -0.54 | -0.04 | 0.10 | -0.01 | 0.32 | 0.26 | 0.05 |
| Walks | -0.12 | 0.88 | 0.25 | 0.88 | -0.19 | -0.67 | -0.59 | -0.25 | -0.75 | -0.49 | 0.50 | 0.45 | 0.42 | -0.09 | -0.79 | -0.05 | -0.14 | -0.23 | -0.39 |
| Stabilizes | -0.16 | -0.02 | -0.41 | 0.19 | -0.09 | -0.13 | -0.10 | -0.15 | -0.46 | -0.30 | 0.22 | 0.08 | 0.12 | -0.25 | -0.43 | -0.12 | -0.28 | -0.40 | -0.28 |
| Coordinates | -0.18 | -0.57 | 0.07 | -0.87 | -0.51 | -0.57 | -0.51 | -0.27 | 0.18 | -0.16 | -0.37 | -0.29 | -0.45 | 0.09 | 0.10 | 0.19 | 0.21 | 0.32 | 0.00 |
| Paces | -0.41 | -0.32 | -0.02 | -0.81 | -0.36 | -0.16 | -0.14 | -0.60 | -0.07 | -0.38 | -0.90 | -0.84 | -0.89 | -0.88 | 0.36 | 0.24 | 0.02 | -0.36 | -0.38 |
| Calibrates | -0.46 | -1.19 | -0.48 | -1.21 | -0.58 | 0.00 | -0.15 | -0.57 | -0.02 | -0.29 | -1.08 | -0.95 | -0.96 | -0.45 | -0.09 | -0.06 | -0.09 | -0.25 | -0.20 |
| Positions | -0.78 | -1.20 | -0.74 | -1.13 | -0.78 | -0.73 | -0.71 | -0.78 | -0.54 | -0.71 | -0.88 | -0.83 | -0.89 | -0.64 | -0.68 | -0.54 | -0.78 | -0.58 | -0.64 |

* See Section 15.1.2 for further clarification

## Table 15-8 ADL Process Item Difficulty Calibration Values by "Diagnostic" Group

| ADL process skill | "Diagnostic" group | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total sample | Well | Old at risk/ frail | Mild | DD | RCVA | LCVA | BI | MS | Other neur | Schiz/ thought | Other psych | MR | Mem/ dem | SCI | RA/ OA | Other mskl | Med/ sens | Mult/ not known |
| Uses | 1.33 | 1.24 | 1.50 | 1.36 | 1.44 | 1.46 | 1.19 | 1.35 | 1.50 | 1.37 | 0.99 | 1.11 | 1.27 | 1.37 | 1.47 | 1.51 | 1.48 | 1.41 | 1.39 |
| Chooses | 0.67 | 0.78 | 0.67 | 0.83 | 0.92 | 0.78 | 0.64 | 0.60 | 0.75 | 0.62 | 0.65 | 0.60 | 0.74 | 0.24 | 1.02 | 0.81 | 0.75 | 0.72 | 0.67 |
| Sequences | 0.55 | 0.50 | 0.53 | 0.53 | 0.58 | 0.60 | 0.59 | 0.48 | 0.72 | 0.56 | 0.36 | 0.40 | 0.43 | 0.37 | 0.77 | 0.63 | 0.65 | 0.64 | 0.62 |
| Searches/ Locates | 0.55 | 0.84 | 0.61 | 0.87 | 0.89 | 0.52 | 0.48 | 0.46 | 0.66 | 0.42 | 0.61 | 0.56 | 0.85 | 0.04 | 0.85 | 0.73 | 0.64 | 0.44 | 0.50 |
| Attends | 0.47 | 0.00 | 0.28 | -0.01 | 0.23 | 0.69 | 0.80 | 0.39 | 0.67 | 0.60 | 0.41 | 0.29 | 0.06 | 0.44 | 0.78 | 0.49 | 0.53 | 0.58 | 0.54 |
| Inquires | 0.31 | 0.38 | 0.41 | 0.41 | 0.53 | 0.42 | 0.41 | 0.23 | 0.44 | 0.30 | 0.26 | 0.22 | 0.46 | -0.10 | 0.57 | 0.38 | 0.32 | 0.34 | 0.29 |
| Gathers | 0.30 | 0.21 | 0.22 | 0.30 | 0.35 | 0.33 | 0.33 | 0.29 | 0.27 | 0.33 | 0.22 | 0.26 | 0.38 | 0.40 | 0.24 | 0.16 | 0.20 | 0.22 | 0.33 |
| Heeds | 0.28 | 0.42 | 0.34 | 0.34 | 0.30 | 0.36 | 0.36 | 0.24 | 0.43 | 0.26 | 0.24 | 0.26 | 0.05 | -0.19 | 0.64 | 0.46 | 0.42 | 0.35 | 0.24 |
| Terminates | 0.07 | -0.08 | 0.10 | -0.09 | -0.10 | 0.11 | 0.17 | 0.06 | 0.20 | 0.11 | -0.13 | -0.09 | -0.25 | 0.13 | 0.20 | 0.23 | 0.20 | 0.15 | 0.13 |
| Navigates | -0.04 | -0.16 | -0.30 | 0.21 | -0.03 | -0.21 | -0.24 | 0.06 | -0.45 | -0.10 | 0.32 | 0.26 | 0.34 | 0.78 | -0.68 | -0.44 | -0.40 | -0.24 | -0.01 |
| Handles | -0.04 | -0.21 | -0.05 | -0.18 | -0.27 | -0.34 | -0.34 | 0.02 | -0.18 | -0.02 | 0.15 | 0.14 | 0.12 | 0.70 | -0.58 | -0.20 | -0.15 | 0.04 | 0.02 |
| Adjusts | -0.08 | 0.10 | 0.01 | 0.25 | 0.05 | -0.14 | -0.10 | -0.06 | -0.12 | -0.10 | -0.12 | -0.10 | -0.07 | -0.06 | -0.13 | -0.08 | -0.10 | -0.13 | -0.12 |
| Continues | -0.14 | -0.29 | -0.34 | -0.24 | -0.07 | -0.11 | -0.08 | -0.21 | -0.12 | -0.12 | -0.14 | -0.22 | -0.12 | -0.13 | -0.24 | -0.20 | -0.19 | -0.18 | -0.09 |
| Restores | -0.18 | -0.28 | -0.13 | -0.35 | -0.25 | -0.08 | -0.08 | -0.09 | -0.06 | -0.12 | -0.29 | -0.16 | -0.30 | -0.41 | 0.01 | -0.07 | -0.05 | -0.21 | -0.21 |
| Initiates | -0.19 | -0.23 | -0.15 | -0.36 | -0.19 | -0.03 | -0.11 | -0.27 | -0.06 | -0.17 | -0.30 | -0.30 | -0.40 | -0.51 | 0.17 | -0.04 | -0.04 | -0.13 | -0.21 |
| Organizes | -0.24 | -0.41 | -0.37 | -0.37 | -0.34 | -0.30 | -0.25 | -0.19 | -0.34 | -0.20 | -0.18 | -0.20 | -0.14 | 0.12 | -0.45 | -0.36 | -0.36 | -0.24 | -0.19 |
| Paces | -0.33 | -0.03 | -0.10 | -0.22 | -0.41 | -0.40 | -0.40 | -0.29 | -0.71 | -0.37 | -0.13 | -0.20 | -0.14 | 0.08 | -0.61 | -0.50 | -0.48 | -0.55 | -0.41 |
| Notices/ Responds | -0.58 | -0.56 | -0.57 | -0.63 | -0.58 | -0.70 | -0.58 | -0.52 | -0.43 | -0.59 | -0.67 | -0.62 | -0.63 | -0.57 | -0.38 | -0.41 | -0.44 | -0.56 | -0.60 |
| Benefits | -1.15 | -0.87 | -1.08 | -1.08 | -1.28 | -1.25 | -1.23 | -1.07 | -1.27 | -1.19 | -0.96 | -0.94 | -1.13 | -1.12 | -1.42 | -1.23 | -1.18 | -1.07 | -1.26 |
| Accommodates | -1.55 | -1.36 | -1.48 | -1.53 | -1.77 | -1.68 | -1.65 | -1.45 | -1.68 | -1.59 | -1.31 | -1.31 | -1.53 | -1.44 | -1.83 | -1.57 | -1.58 | -1.46 | -1.67 |

* See Section 15.1.2 for further clarification

≥ 0.50 logit between groups can be considered to be significant; the group with the lower item calibration value demonstrated more relative difficulty performing the ADL action without increased clumsiness or physical effort, decreased efficiency, or safety risk.

Because the presence of DIF can signal a risk to the measurement system and/or the presence of test bias, we proceeded to examine for ***differential test functioning*** to ensure that the AMPS measurement system was not disrupted and that there was no bias within the AMPS associated with diagnosis. ***The results of all possible paired analyses*** (e.g., using item calibration values for the well persons vs. using the item calibration values for the sample with mild disabilities) ***revealed no evidence of disruption of the measurement system or differential test functioning***. That is, the maximum standardized difference ($Z$) among all of the persons in the standardization sample was 0.34 (critical value required to identify significant differential test functioning is $Z = 2.00$, $p \leq .05$), and the difference in ADL motor or ADL process ability measures based on all possible paired comparisons never exceeded 0.10 logit.

**Despite differences in ADL motor and ADL process item calibration values among diagnostic groups, there is no evidence of test bias within the AMPS associated with a person's diagnosis.**

## 15.7 Current Evidence for Validity of the AMPS with Children

### 15.7.1 Goodness of fit of Children to the Many-faceted Rasch Model of the AMPS

While the AMPS was designed to be used with children and adolescents, the AMPS has been standardized primarily on adults. This situation has occurred because approximately 90% of the persons in the AMPS database, the source of the sample used to standardize the AMPS, were 16 years of age or older (see Table 15-1). One source of evidence to support the use of the AMPS with children and adolescents can be obtained by evaluating whether children and adolescents fit the existing many-faceted Rasch model (Linacre, 1993) that has been defined primarily by an adult standardization sample. In order for children and adolescents to show acceptable goodness of fit to the many-faceted Rasch models that define the AMPS motor and process skill scales, tasks and items that are more likely to be easier (or harder) for adults must also be more likely to be easier (or harder) for children and adolescents, and strict (or lenient) raters must reliably be strict (or lenient) when rating either adults or children and adolescents. While it is generally considered ideal to have at least a 95% overall rate of acceptable fit, it is often the case that the level of person misfit found among performance assessments is as high as 10%. For the AMPS, the level of misfit among the total current standardization sample was 8% on the ADL motor scale and 10% on the ADL process scale.

When Poulson (1996) examined the goodness of fit statistics for a sample of 162 well, nondisabled children and adolescents between 3 and 15 years of age, he found that 90% of his sample demonstrated acceptable goodness of fit to the many-faceted measurement model for the AMPS motor scale, and 95% of his sample demonstrated acceptable goodness of fit to the measurement model for the AMPS process skill scale defined by the adult standardization sample. Thus, the children and adolescents demonstrated an acceptable rate of goodness of fit to the AMPS motor skill scale that was slightly lower than the 92% seen in the current standardization sample.

In contrast, when we examined the data for the 15,214 children included in the current standardization sample, only 870 (5.7%) failed to demonstrate acceptable goodness of fit on the ADL motor scale and 1448 (9.5%) failed to demonstrate acceptable goodness of fit on the ADL process scale. In both cases, the level of misfit was less than among the total sample of 148,158 persons, supporting person response validity when the AMPS is used with children. When we included only well, nondisabled children, the rate of unacceptable fit was even lower, 4.4% on the ADL motor scale and 7.9% on the ADL process scale.

Poulson (1996) supplemented his evaluation of person fit with a second method for evaluating goodness of fit — overall percentage of misfitting ratings (Fisher, 1993, 1994; Goto et al., 1996). A previous analysis of the data for 609 nondisabled adults (16 to 59 years of age) and 488 well older adults (60 years of age and above) had revealed an overall rate of misfitting ratings ($t \geq 3$) on the AMPS motor and process skill scales of 2.0 to 2.5% (Fisher, 1995b). Based on the rationale that well, nondisabled children and adolescents should be expected to meet similar standards of fit as do adults and older persons, Poulson set his criterion for an acceptable rate of misfitting ratings at

≤ 2.5% at $t \geq 3$. He found that the rate of misfitting ratings was 2.2% on the AMPS motor skill scale and 1.6% on the AMPS process skill scale. Poulson concluded that the overall proportion of misfitting ratings for both AMPS scales was acceptable.

Finally, Poulson found that within a subsample of the younger children (3 to 7 years of age) 23% of the misfitting item ratings were associated with the motor skill item Lifts; the younger children tended to get unexpectedly lower scores on Lifts. Further analysis, however, revealed that misfit on the skill item Lifts did not meaningfully impact the final ADL motor ability measures of the children in the younger age group. Poulson concluded, therefore, that there was no differential test functioning, and that the AMPS motor and process skill scales are valid for use with children and adolescents. His conclusion is supported by our more recent findings presented above.

**When evaluated with the AMPS, children demonstrate valid patterns of response across the AMPS items and tasks.**

### 15.7.2 Evidence that ADL Ability Increases with Age Among Typically-developing Children

Another source of evidence that the ADL motor and ADL process ability measures are valid can be derived from evidence that the quality or effectiveness (e.g., level of clumsiness or physical effort, degree of efficiency) of the goal-directed motor and process actions in children increase with age. The assumption that ADL motor and ADL process ability increases with age in young children has been supported by Bäckström (1995), who evaluated nondisabled children, ages 6, 9, and 12 years of age. She found that the ***mean ADL motor and ADL process ability*** of her sample subgroups ***increased significantly with age***.

When we examined the data for the 8,799 well, typically-developing children, 3 to 15 years of age in the current standardization sample, we found a significant relationship between age and both ADL motor and ADL process ability ($r$ = .64). Moreover, ANOVAs with post hoc Tukey HSD tests revealed that all age groups 3 to 12 years of age differ significantly in mean ADL motor ability, and all age groups 3 to

11 years of age differ significantly in mean ADL process ability ($p \leq .05$). After age 12 on the ADL motor scale and age 11 on the ADL process scale, ADL ability continues to increase, but adjacent age groups no longer differ significantly. See Section 15.9 for further discussion of changes in ADL ability across the lifespan; normative mean ADL motor and ADL process abilities for the well persons 3 to 103 years of age are shown in Table 15-9.

### 15.7.3 Evidence that Typically-developing Children and Children with Disabilities Differ in Mean ADL Ability

In a relatively small study, Payne and Howell (2005) compared the ADL motor and ADL process ability measures across a ***typical client caseload of children*** ($n$ = 33) that included both typically-developing children and children with disabilities. Of the diagnostic groups that had sufficient data for comparative analyses, the ADL ability measures were compared to normative data (Fisher, 2006a), revealing the following results (note, some participants had dual diagnoses and their data were compared under both diagnoses, thus there were 32 diagnostic comparisons):

- All six of their typically-developing children had ADL motor and ADL process ability measures that were within ±1 standard deviation (*SD*) of the normative means.

- The majority of the children with attention deficit hyperactivity disorder (ADHD) ($n$ = 2) and developmental coordination disorder (DCD) ($n$ = 13) had ADL motor and ADL process ability measures that were within ±1 *SD* of the respective normative means, with one participant with ADHD and two participants with DCD obtaining ADL motor and ADL process ability measures that were more than 1 *SD* below the normative means.

- Three of the five children with an autism spectrum disorder had ADL motor and ADL process ability measures that were within ±1 *SD* of their healthy peers, two children with an autism spectrum disorder had ADL motor ability measures that were more than 1 *SD* below the normative mean, one child with an autism spectrum disorder had an ADL process ability measure that was more than 1 *SD* below the normative mean, and one child with an autism spectrum disorder had an ADL process ability measure that was more than 2 *SD* below the normative mean.

- All three of the children with cerebral palsy (CP) had ADL motor ability measures that were more than 2 *SD* below the normative mean. In contrast, two children with CP had ADL process ability measures that were more than 1 *SD* below the normative mean, and the remaining child with CP had an ADL process ability measure that was more than 2 *SD* below the normative mean.

**Payne and Howell (2005) noted the following advantages to using the AMPS with children:**

- **When used with older, more able children, *the typical ceiling effects often encountered with the use of other assessments were not observed with the AMPS***
- **"Therapists found that there was often no need to perform additional standardized assessments with these children because the deficits in performance components and their impact on function were clear from their performance of the AMPS tasks" (p. 279)**
- **Both the children and their parents expressed greater affinity to the AMPS — the *AMPS appeared to more readily acknowledge both the strengths and weaknesses, within the context of real life tasks* (versus focusing solely on the child's problems)**

White and Mulligan (2005) also documented that children with ***ADHD*** ($n = 12$) obtained significantly lower mean ADL motor and ADL process ability measures than did a healthy cohort ($n = 21$) ($t$ (31) = 1.79, $p = 0.04$ and $t$ (31) = 3.47 $p = 0.001$, respectively).

### 15.7.4 Age-related Differences in Mean ADL Ability Among Children with Cerebral Palsy

Finally, Van Zelst, Miller, Russo, Murchland, and Crotty (2006) investigated differences in ADL ability among children with cerebral palsy (CP) across different age groups. The younger children (3 to 8 years of age, $n = 34$) in their sample had significantly higher mean ADL motor skill ability measures than did their older cohort (9 to 12 years of age, $n = 20$) (0.46 logit versus 0.09 logit, respectively, $p = 0.041$). In contrast, the mean ADL process ability measures did not differ between the two age groups ($p = 0.885$). Additionally, the ADL ability measures of all of their participants were below the ADL motor and ADL process cutoff measures and below age expectations. The results of this study indicate that not only do children with cerebral palsy experience difficulty performing ADL tasks, they also may not be as likely to develop greater ADL skill with age. Such findings warrant further studies aimed at evaluating the longitudinal changes in ADL ability among persons with CP.

## 15.8 Current Evidence of the Effect of the Environment on ADL Ability

To date, three studies have been implemented comparing the performance of persons tested in their home environment and again in a clinical setting (Darragh, Sample and Fisher, 1998; Nygård, Bernspång, Fisher, & Winblad, 1994; Park, Fisher, & Velozo, 1994). Each study sample was approximately 20 people, with a total of 61 persons across all three studies. The persons tested had memory disorders, were older adults living in the community, or were persons with acquired brain injuries. Overall, the results of these studies revealed that while approximately 80% of the paired ADL ability measures remained unchanged between home and clinic, 20% had higher ADL process ability measures and 5% had higher ADL motor ability measures when tested at home than when tested in a clinical setting, and 5% had lower ADL process ability measures when tested in the home versus the clinic setting (none had ADL motor ability measures that were lower when tested in the home). Considered together, these finding suggest a need to be cautious when using the results of an AMPS observation implemented in a clinical setting to predict performance in the person's home environment. Ideally, the occupational therapist should seek to evaluate their clients in the most contextually relevant setting(s) (e.g., the environment where the person will be performing ADL tasks). Another precaution is in order. These results were all based on community-living persons who went to a clinical setting for the AMPS observation. To date, no research has been implemented that compares the results for persons who are currently hospitalized and tested in both the hospital and their planned home environment.

**ADL task performance in one setting (home, clinic) cannot always be generalized to another setting.**

**While there is evidence that there is a 95% chance that a person's ADL motor ability will remain stable between home and clinic settings, there is a 25% chance that the person's ADL process ability will differ.**

**Therefore, the occupational therapist should ideally evaluate the person in the setting(s) in which there is concern about ADL task performance.**

## 15.9 Current Evidence that ADL Ability Changes Across the Lifespan: Development of AMPS Norms

ADL motor and ADL process skills are learned and practiced abilities (Connolly & Dalgleish, 1989). If we consider the constructs of ADL motor and ADL process ability, which pertain to the ***quality*** of ADL task performance, we can begin to develop expectations with regard to changes in ADL ability across the lifespan. More specifically, if we consider healthy, nondisabled persons, we would expect young children to be less skilled than adults and adults to be more skilled than older adults. In support of this idea, Dickerson and Fisher (1993) found significant age effects between ***young and older well*** women on both AMPS scales; the older adults had lower mean ADL ability than did the younger adults. These results contradict research that suggests that healthy, well older persons do not differ significantly from younger adults on performance of familiar, practiced, and ecologically valid tasks (see Dickerson & Fisher for a review). It should be noted, however, that some of the older persons tested by Dickerson and Fisher performed just as well as the most able of the younger participants. Thus, while overall significant differences were detected between the two groups of women, upon closer examination of the data, the researchers often found similar individual measures between the younger and older women.

To further verify this prediction, Hayase et al. (2004) conducted a study of healthy, well persons ranging in age from 3 to 93 ($n$ = 4398). They found that ADL motor and ADL process ability increase sharply from 3 to 6 years of age and continue to improve until 15 years of age. Between the ages of 15 and 50 years, ADL ability plateaus and then gradually declines after 50 years of age.

Our examination of the mean ADL motor and ADL process ability measures of the well persons in the current standardization sample confirmed the findings of Hayase et al. (2004). More specifically, with a larger sample, there is additional evidence that ADL ability increases during childhood and young adulthood and declines during older adulthood (see Figure 15-1 and Table 15-9; see also Section 15.7.2 for more information about developmental changes in children). Thus, as children develop and gain more experience performing ADL tasks, the quality of their ADL task performances increases.

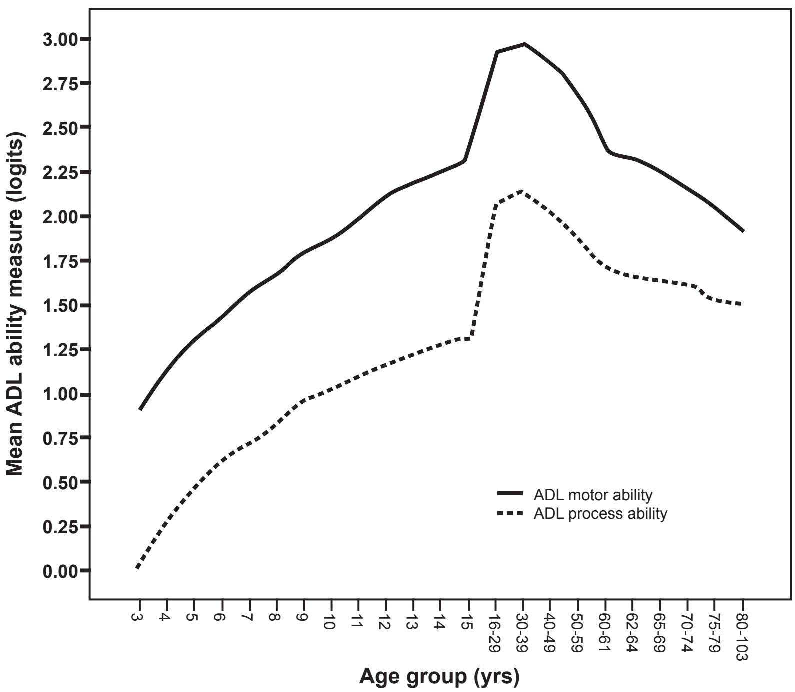


***Figure 15-1.*** **Mean ADL ability of well persons across the lifespan.**

**Table 15-9 Gender and Mean ADL Motor and ADL Process Ability Measures (logits) for the Well Persons in the AMPS Standardization Sample (*n* = 12,773)**

| | | Gender | | ADL motor ability | | ADL process ability | |
|---|---|---|---|---|---|---|---|
| Age (yrs) | *n* | Male | Female | *M* | *SD* | *M* | *SD* |
| 3 | 523 | 227 | 296 | 0.92 | 0.40 | 0.01 | 0.41 |
| 4 | 761 | 355 | 406 | 1.14 | 0.41 | 0.26 | 0.43 |
| 5 | 880 | 397 | 483 | 1.29 | 0.44 | 0.45 | 0.43 |
| 6 | 902 | 437 | 465 | 1.42 | 0.45 | 0.59 | 0.41 |
| 7 | 901 | 431 | 470 | 1.57 | 0.44 | 0.71 | 0.40 |
| 8 | 774 | 322 | 452 | 1.66 | 0.48 | 0.80 | 0.41 |
| 9 | 776 | 344 | 432 | 1.81 | 0.46 | 0.95 | 0.41 |
| 10 | 724 | 317 | 407 | 1.87 | 0.47 | 1.00 | 0.41 |
| 11 | 696 | 256 | 440 | 1.99 | 0.49 | 1.09 | 0.40 |
| 12 | 584 | 258 | 326 | 2.12 | 0.50 | 1.14 | 0.42 |
| 13 | 515 | 207 | 308 | 2.19 | 0.51 | 1.21 | 0.41 |
| 14 | 395 | 188 | 207 | 2.25 | 0.51 | 1.28 | 0.44 |
| 15 | 368 | 176 | 192 | 2.34 | 0.48 | 1.30 | 0.41 |
| 16 to 29 | 279 | 73 | 206 | 2.95 | 0.52 | 2.08 | 0.43 |
| 30 to 39 | 222 | 102 | 120 | 2.99 | 0.47 | 2.14 | 0.50 |
| 40 to 49 | 162 | 65 | 97 | 2.88 | 0.49 | 2.03 | 0.48 |
| 50 to 59 | 213 | 65 | 148 | 2.66 | 0.46 | 1.88 | 0.40 |
| 60 to 61 | 547 | 180 | 367 | 2.37 | 0.51 | 1.70 | 0.45 |
| 62 to 64 | 574 | 197 | 377 | 2.31 | 0.51 | 1.66 | 0.41 |
| 65 to 69 | 708 | 286 | 422 | 2.25 | 0.53 | 1.63 | 0.46 |
| 70 to 74 | 601 | 222 | 379 | 2.15 | 0.53 | 1.62 | 0.47 |
| 75 to 79 | 374 | 130 | 244 | 2.04 | 0.51 | 1.51 | 0.43 |
| 80 to 103 | 294 | 88 | 206 | 1.90 | 0.50 | 1.50 | 0.42 |

As persons enter into adulthood, ADL ability increases even more. It is possible that as persons transition into adulthood, there are increased demands and expectations to effectively perform the wider array of ADL tasks that are necessary for independent living in the community. Interestingly, it is at this point in the lifespan where we see a rather steep increase in both ADL motor and ADL process ability; clearly additional research is warranted to verify if the steep increase in quality of ADL task performance is due to transitioning into adulthood and increased demands for independent living. Lastly, as persons enter into their 40s and 50s, it appears that the natural aging process begins to result in a decline in the quality of ADL task performance, and that the quality of ADL task performance continues to decline with advancing age.

### 15.9.1 Developing Normative Expectations Used to Interpret the Results of an AMPS Observation

The mean ADL motor and ADL process ability measures shown in Table 15-9 provide the basis for interpreting a person's ADL motor and ADL process ability measures from a norm-based perspective. When these mean values were calculated, limited sample sized required that well persons 16 years of age and above be grouped into larger age groups. When these age groups were formed, we ensured that there was no statistical difference between the yearly means for the persons who were included within a combined age group. Thus, for example, there was no significant difference among any of the yearly age groups 16 to 29 years of age.

The Results and Interpretation Report, generated by the AMPS computer-scoring software, documents the expected range of ADL ability measures for healthy, well, age-matched peers of the person who was evaluated using the AMPS, and indicates whether or not the person tested has ADL motor and/or ADL process ability measures that are within that expected range of ADL ability. The expected range is delineated by ±2 *SD* from the age-matched mean for the well sample (see Table 15-9); 95% of healthy well persons of the same age are expected to have ADL ability measures within this range. See Chapter 10 for more information on interpreting AMPS results.

- **Examining a person's ADL motor and ADL process ability measures in relation to normative age expectations can help occupational therapists describe and document the level of ADL ability of the persons who they evaluate in relation to where we would expect healthy persons of the same age to fall on the respective AMPS scales**
- **When a person's ADL ability measures are below age expectations (i.e., below -2 *SD*), we can use this data to justify a person's need for additional occupational therapy services to lessen the gap between the their ADL ability and normative expectations**

### 15.9.2 Using the AMPS Normative Values to Monitor Rater Severity

Within the discussion of normative expectations in overall ADL ability, it is important to remember that the ADL motor and ADL skill items of the AMPS are scored based on the criterion of competence and not on the basis of normative expectations. Thus, we expect children, older adults and persons with disabilities to demonstrate less skill when enacting the ADL motor and ADL process skills when engaged in ADL task performances. Therefore, if children, older adults, or persons with disabilities obtain ADL ability measures that are higher than expected, and/or similar to those expected of healthy, well adults, this could be an indication of measurement error — the person was given higher ratings on ADL motor or ADL process skill items than would be expected for a child, an older adult, or a person with a disability, respectively. When the rater finds that he or she is assigning many high scores to these persons, the AMPS rater should evaluate if (a) the AMPS tasks the person performed were of sufficient challenge, (b) the ADL skill items scores were accurately based on the scoring criteria in the AMPS manual (see Chapter 8 and Volume 2, Chapter 8, AMPS Skill Items), and/or (c) errors occurred when the person's data were entered into the AMPS computer-scoring software.

**Comparing a person's ADL motor and ADL process ability measures to normative ADL ability expectations:**

- **Provides evidence of a person's quality of ADL task performance in relation to his or her healthy peers**
- **Can help individual AMPS raters to monitor their scoring severity and their adherence to the standardized administration and scoring procedures of the AMPS**

## 15.10 Current Evidence for Differences in Mean ADL Ability by Functional Level: Development of the AMPS Scale Cutoff Criteria

Within the early analyses of AMPS data, researchers determined that the ADL motor and ADL process cutoff measures could be used to help demarcate those who demonstrate competent ADL task performance (i.e., no signs of clumsiness or physical effort, decreased efficiency, safety risk, or need for assistance) from those who are less competent when engaging in ADL tasks. These cutoff measures were set at 2.0 logits on the ADL motor scale and 1.0 logit on the ADL process scale (Fisher 2006a); those above the cutoff measures demonstrate competent ADL task performance. Over time, the evidence also revealed that the ADL process ability cutoff might also be useful for identifying persons who likely need some assistance to live in the community (Fisher, 2006a; Hartman, Fisher, & Duran, 1999). In the following sections, we will summarize the evidence that has supported the use of the AMPS cutoff measures to identify the need for assistance to live in the community. Please note that while more recent evidence suggests changing the ADL cutoff measures to enhance the accuracy of predicting the need for assistance in the community, the original (i.e., quality of performance) cutoff measures remain unchanged (i.e., persons above 2.0 logits on the ADL motor scale and 1.0 logit on the ADL process scale demonstrate competent ADL motor and ADL process performance, respectively).

### 15.10.1 Using the AMPS Competence Cutoff Measures to Identify Persons Who Need Assistance with Community Living

Across multiple research studies, it has been found that the ***AMPS process skill scale has higher discriminating power*** than does the ADL motor scale when differentiating between individuals who are able to live independently in the community from those who require assistance to live in the community (Aggson, 1996; Kottorp, 2008; Hartman et al., 1999; Merritt, 2010, in press). While ADL process ability is more closely associated with global functional level (i.e., independent in the community, in need of minimal assistance in the community, or in need of moderate to maximal assistance to live in the community), analyses indicate that global functional level is significantly associated with both ADL motor ability ($F_{2,64463}$ = 8,728, $p < 0.01$) and ADL process ability ($F_{2,64463}$ = 15,916, $p < 0.01$), with both analyses revealing large effect sizes ($Eta^2$ = 0.21 and 0.33, respectively) (see Table 15-10) (Merritt, in press). Upon further examination of the means and standard deviations, we found an overlap in ADL ability among the persons in the three functional level groups. Such variations may be a reflection of the diverse abilities of persons within each functional level group, error in the occupational therapists' ratings of functional level, and/or unexplained variation in ADL ability measures within and between the three functional level groups. Regardless, the statistical analyses reveal that as ADL ability decreases, there is a significantly greater need for assistance to live in the community.

**Table 15-10 Mean ADL Motor and ADL Process Ability by Global Functional Level**

| | Global functional level | | |
|---|---|---|---|
| | **Independent (*n* = 17,877)** | **Minimal assistance (*n* = 35,948)** | **Maximal assistance (*n* = 48,541)** |
| Age (years) | | | |
| *M* | 53.9 | 54.8 | 54.3 |
| *SD* | 18.1 | 23.0 | 26.8 |
| ADL motor ability (logits) | | | |
| *M* | 1.92 | 1.37 | 0.75 |
| *SD* | 0.74 | 0.78 | 0.98 |
| ADL process ability (logits) | | | |
| *M* | 1.53 | 1.01 | 0.37 |
| *SD* | 0.53 | 0.51 | 0.70 |

As there is not a "precise" cutoff measure which can be used to differentiate between those who are independent from those who need some assistance, Bernspång and Fisher (1995a) identified a ***risk zone*** for the AMPS scales delineated by the cutoff ability measures ± 0.3 logit. When analyzing the data of persons with right or left CVA (*n* = 147), well independent persons (*n* = 71), and a small group of persons who had been identified as being frail or at risk of functional decline (*n* = 12), Bernspång and Fisher found that 8 of the 12 persons (67%) who were frail or at risk had ADL motor and ADL process ability measures within the risk zones. In a second analysis of 40 persons in the AMPS database who were similarly identified as being frail or at risk of functional decline, it was once again found that 67% had ADL process ability measures located within the risk zone. Additionally, Bernspång and Fisher found that none of the well, independently-living persons had ADL motor or ADL process ability measures below the risk zones (i.e. below 1.7 or 1.3 logits on the motor and process scales, respectively). Lastly, within the group of persons with stroke, only 5% had ADL motor ability measures above 2.3 logits and 30% had ADL process ability measures above 1.7 logits. The persons who obtained measures above the ADL process risk zone demonstrated greater ability to compensate for any residual limitations and were able to live independently in the community.

Additional studies evaluated the AMPS competence cutoff measures as indicators of the need for assistance to live in the community. With a sample of older adults with and without DAT, Hartman et al. (1999) found a 94% overall correct rate of classification of independent or in need of assistance (based on whether their ADL measures were above or below the cutoff measure of 1.0 logit) for the AMPS process skill scale, and an overall correct classification rate of 64% on the ADL motor scale. When Hartman et al. considered the process scale risk zone proposed by Bernspång and Fisher (1995a), they found that 65% of their participants who had been incorrectly classified fell into the risk zone.

Similarly, with a group of persons with brain injury, Aggson (1996) found that 91% of persons below the AMPS motor skill cutoff measure and 89% of the persons below the AMPS process skill cutoff measure needed assistance to live in the community. When Aggson considered both AMPS scales simultaneously, her overall rate of correct classification increased to 82%.

Kottorp (2008) explored the use of the AMPS ability measures when predicting level of care needed for community living among a group of persons with mental retardation ($n$ = 380). With this population of persons, Kottorp found that only the ADL process ability measures of the AMPS contributed significantly to the prediction of the level of assistance required for community living, while ADL motor ability did not contribute to determining level of care needed. As such, Kottorp documented growing evidence that the ADL ability measures obtained from AMPS evaluations can be used to contribute to the determination of how much care a person likely needs to live safely in the community.

### 15.10.2 Developing New Cutoff Values for Predicting Level of Assistance Needed for Community Living

More recently, Merritt (in press) generated receiver operating characteristic (ROC) curves to determine the accuracy of using ADL motor and ADL process ability measures to determine the need for assistance to live in the community. Overall, ADL motor and ADL process ability showed fair to good discriminating value as evidenced by the area under the ROC curves (0.78 for ADL motor and 0.84 for ADL process ability). Further examination of the ROC curve analyses revealed that using the current ADL motor cutoff of 2.00 logits to determine the need for assistance was associated with unacceptable sensitivity estimates at 0.40 (i.e., only 40% of those who are independent in the community were classified correctly).

Given the fact that the original ADL motor cutoff was developed to identify persons who perform ADL tasks with clumsiness or physical effort, the finding that the

ADL motor cutoff was too high to accurately demarcate persons who are independent in the community from those who need assistance was not surprising. Clearly, persons who demonstrate slight or minimal clumsiness or physical effort are often capable of independent living — such persons would fall below the ADL motor cutoff, yet would demonstrate the ability to live independently in the community.

**Cutoff Measures for Predicting Need for Assistance**

Through continued examination of ROC curve analyses, the ideal ADL motor cutoff for determining the need for assistance was found to be 1.50 logits (sensitivity = 0.67 and specificity = 0.72) (Merritt, in press). In contrast, examination of the ROC curve revealed that the current ADL process cutoff measure of 1.00 logit was associated with acceptable sensitivity and specificity estimates of 0.81 and 0.70, respectively (i.e., 81% of the independent sample were correctly classified and 70% of the sample in need of assistance were correctly classified). Hence, Merritt (in press) proposed ***a new ADL motor cutoff of 1.50 logits*** for assisting with prediction of assistance, and the ***continued use of the 1.00 logit ADL process cutoff*** for both identifying those with competent ADL process ability and those who are likely in need of assistance with community living.

**Combined Use of ADL Motor and ADL Process Measures to Enhance Prediction of Independence Versus Need for Assistance with Community Living**

Merritt (in press) conducted additional analyses to see if matched decisions resulted in higher accuracy estimates when identifying persons who need assistance to live in the community. ***When both the ADL motor ability and the ADL process ability measures are either below (or above) the cutoff measures of 1.5 and 1.0 logits, respectively, the accuracy estimates were enhanced***. When matched decisions occurred (*n* = 41,664), 86% of the independent sample were correctly identified as being independent in the community, and 83% of those in need of assistance were below the both of the cutoff measures and were correctly identified as needing assistance.

**Combined Use of ADL Motor and ADL Process Measures to Enhance Prediction of Level of Assistance Needed for Community Living**

In a follow up study, Merritt (2010) investigated the utility of using the ADL motor and ADL process ability measures to classify persons as either (a) independent, (b) in need of minimal assistance to live in the community, or (c) in need of maximal assistance to live in the community. Once again, ROC curves were generated to determine accuracy estimates of the proposed cutoff measures. Based on these analyses, an additional ADL motor cutoff measure, demarking the ***need for maximal assistance***, was set at 1.00 logit (sensitivity = 0.70, sensitivity = 0.66). Likewise, an additional ADL process cutoff measure, demarking the need for maximal assistance, was set at 0.70 logit (sensitivity = 0.79, specificity = 0.69).

Thus, two ADL motor cutoff measures were proposed, one at 1.5 logits (measures above this cutoff indicate independence in the community) and one at 1.00 logit (measures below this cutoff indicate the need for moderate to maximal assistance to live in the community). Similarly, two ADL process ability measures were proposed, one at 1.00 logit and the other at 0.70 logit. Figures 15-2 through 15-4 graphically display the accuracy estimates for predicting level of dependence for each of the three functional levels: independent, need for minimal assistance or supervision, need for at least moderate assistance. Comparison of predication rates based on ADL motor ability only, ADL process ability only, and combined consideration of both clearly supports enhanced prediction when both ADL motor and ADL process ability measures are considered.

Merritt's analyses (2010) reveal that the ***AMPS ability measures are relatively accurate when used as evidence to support clinical decisions of either independent in the community or the need for maximal assistance to live in the community***. In contrast, ***evidence does not support the use of the AMPS ability measures in accurately identifying those who need minimal assistance to live in the community***. Figure 15-3 suggests that while ADL process ability measure is likely the best predictor of need for minimal assistance, only 33% of the sample was correctly classified.

Merritt (2010) postulated that one explanation for the relative inaccuracies for determining the need for minimal assistance may lie in the fact that the AMPS was not designed to determine level of dependence in the community, rather it was ***designed to measure a person's quality of ADL task performance***, thus, a certain amount of inaccuracy and error is expected. Additionally, there is no possibility to determine

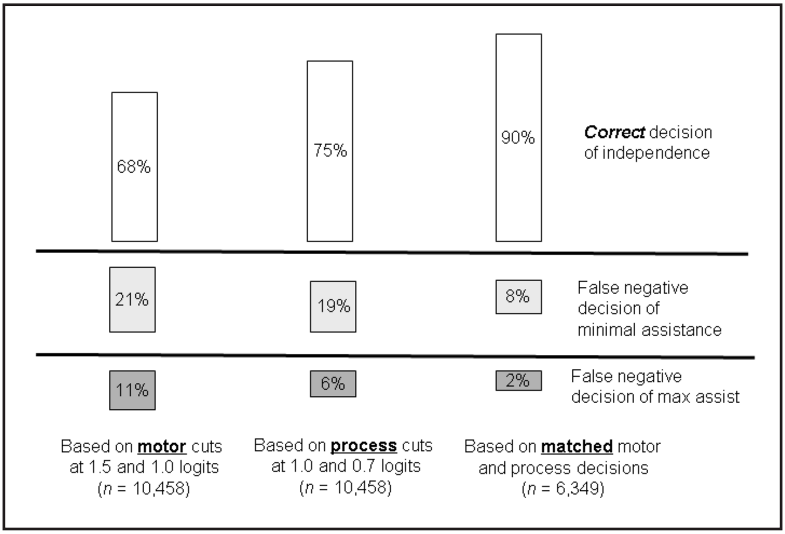


***Figure 15-2.*** **Accuracy of correctly categorizing persons who are *independent* in the community (Merritt, 2010. Reprinted with permission of the author).**

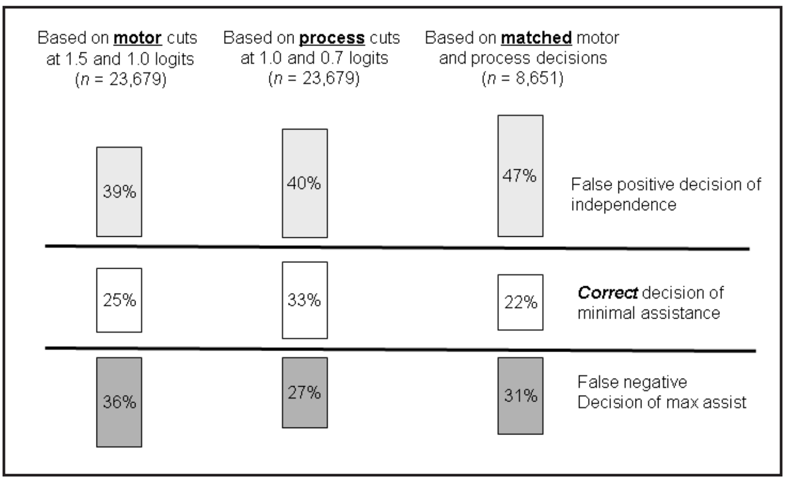


***Figure 15-3.*** **Accuracy of correctly categorizing persons who need *minimal assistance* to live in the community (Merritt, 2010. Reprinted with permission of the author).**

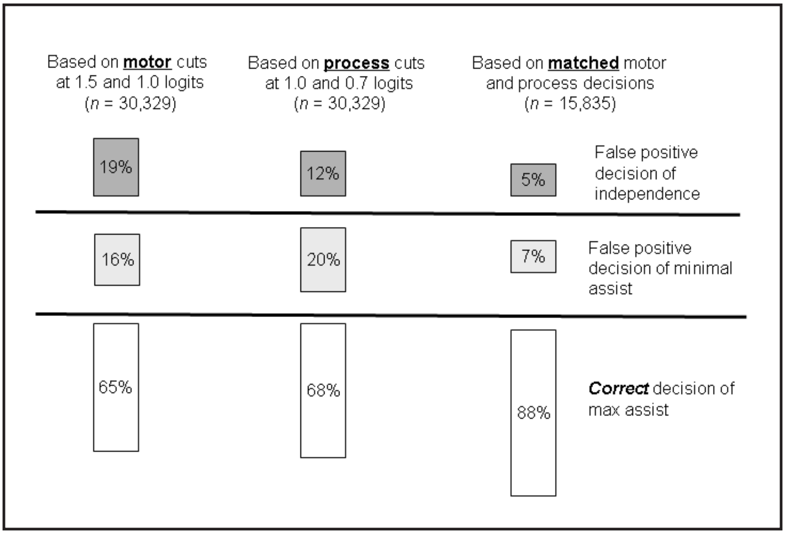


***Figure 15-4.*** **Accuracy of correctly categorizing persons who need *at least moderate assistance* to live in the community (Merritt, 2010. Reprinted with permission of the author).**

if the functional level ratings of the participants in her study were accurate, and thus, there may be error in the functional level ratings assigned by the raters.

### 15.10.3 Interpreting the Results of an AMPS Observation Based on New Cutoffs for Predicting Need and/or Level of Assistance Needed for Community Living

Merritt's (2010, in press) results provide us with new guidelines we can apply as we interpret the results of a person's AMPS observation. More specifically, these guidelines can be applied to ***support the occupational therapist's professional judgment of a person's need for assistance to live in the community***. These guidelines are as follows:

- New proposed ADL motor and ADL process cutoff measures demarking independence (1.5 and 1.0 logits, respectively) and the need for moderate to maximal assistance (1.0 and 0.7 logit, respectively) have been proposed.

- While matched ADL motor and ADL process decisions result in the most accurate predictions of level of dependence, when decision points do not match, the ***ADL process ability is the most accurate predictor of the need for assistance***.

- When ADL ability measures fall between the proposed cutoff measures (i.e., within the "risk zone") it is even more important to gather additional evidence to determine the level of assistance a person requires.

### 15.10.4 Summary of Current Evidence Related to the ADL Competence Cutoffs and Cutoffs for Predicting Need and/or Level of Assistance for Community Living

In summary, different cutoff values are needed for different decisions. The ***ADL competence cutoffs*** (identified on the Graphic Report) can be used to identify persons who likely demonstrate some clumsiness and/or increased physical effort performing ADL tasks (ADL motor cutoff = 2.0 logits) and/or persons who likely demonstrate some spatial-temporal inefficiencies when performing ADL tasks (ADL process cutoff = 1.0 logit). The newly developed ***"level of assistance" cutoffs*** of 1.5 logits and 1.0 logit on the ADL motor scale and 1.0 logit and 0.7 logit on the ADL process scale can be used to help support the occupational therapist's professional judgment of a person's need for assistance for community living, and possibly, the level of assistance that might be needed. Thus, in conjunction with the occupational therapist's professional judgment, current research supports a constellation of interpretations with regard to a person's ADL motor and ADL process ability measures, these interpretations are summarized in Table 15-11.

## 15.11 Current Evidence for Reliability of the AMPS Measures

In this section, we will present evidence to support different types of reliability for the ADL motor and ADL process ability measures of the AMPS. Reliability pertains to the stability of measures when testing procedures are repeated with a group of individuals or a single individual (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Within a traditional measurement perspective, the two most

**Table 15-11 Summary of the Clinical Interpretations of the ADL Motor and ADL Process Cutoff Measures**

| Person ability | Clinical interpretation |
|---|---|
| **ADL motor scale** | |
| **Person ability** | **Clinical interpretation** |
| • Above 2.0 logits | • Little to no increased clumsiness or physical effort |
| • Below 2.0 logits | • Evidence of increased clumsiness or physical effort |
| • Above 1.5 logits | • Less likely to need assistance to live in the community |
| • Below 1.5 logits | • More likely to need assistance to live in the community |
| **ADL process scale** | |
| **Person ability** | **Clinical interpretation** |
| • Above 1.0 logit | • Likely efficient; competent use of time, space, and tools |
| • Below 1.0 logit | • Evidence of some inefficiencies and/or undesirable use of time, space, and/or tools |
| • Above 1.0 logit | • Less likely to need assistance to live in the community |
| • Below 1.0 logit | • More likely to need assistance to live in the community |
| **ADL motor and ADL process scales** | |
| **Person ability** | **Clinical interpretation** |
| • Motor ability above 1.5 logits ***and*** • Process ability above 1.0 logit | • Best prediction of independent in the community (86% correct classification) |
| • Motor ability below 1.5 logits ***and*** • Process ability below 1.0 logit | • Best prediction of the need for assistance to live in the community (83% correct classification) |
| • Motor ability below 1.0 logit ***and*** • Process ability below 0.7 logit | • May need maximal assistance to live in the community |
| • When ADL motor and ADL process measures do not fall within the same decision zones, ADL process ability is the strongest predictor of the need for assistance to live in the community. | |

common estimates of reliability are the standard error of measurement and the reliability coefficient (Haertel, 2006). The ***standard error of measurement*** (*SE*) of each person's ADL motor or ADL process ability measure is perhaps the most useful reliability index as it provides an estimation of the potential within-person variation of a person's measure expressed in the same units that are used in the scale (e.g., logits) (Feldt & Brennan, 1989). More specifically, the *SE* "represents the standard deviation of a hypothetical set of repeated measurements on a single individual" (Feldt & Brennan, p. 105), such that there is a 95% chance that his or her measure falls within ±2 *SE* from his or her obtained measure. Moreover, if a person's ADL motor or ADL process ability measures differ by more than 2 *SE*, the two measures can be said to differ significantly ($p \leq .05$) (Harvill, 1991). Finally, the smaller the *SE*, the more likely the generated measures will be reliable and sensitive indices of change (Bond & Fox, 2007).

The use of ***reliability coefficients*** is a more traditional method of reporting the stability of a set of scores (Haertel, 2006). Unlike the *SE*, which is based on a single test score, reliability coefficients are based on correlations between two replications of the evaluation procedures administered to the same group under different conditions (e.g., different occasions, different forms of the same instrument). The most common methods for estimating reliability coefficients are test-retest, parallel forms, and split-half methods. While ***test-retest reliability*** pertains to the stability of test scores from the same test (same form) administered at two separate occasions, ***parallel forms reliability*** pertains to the stability of test scores obtained from two different forms of the same test. One of the most common methods for estimating the reliability coefficient for test scores of groups is based on estimating the ***reliability based on all possible split-halves of a set of test items*** in the form of Cronbach's coefficient alpha (Allen & Yen, 2002; Ary, Jacobs, & Razavieh, 2002; Crocker & Algina, 1986; Cronbach, 1951; Feldt & Brennan, 1989; Haertel, 2006).

When Rasch analysis computer programs are used, they generate reliability estimates in the form of a ***Rasch equivalent of Cronbach's alpha*** and an ***SE for each person's estimated measure***. Rasch analysis computer programs also generate reliability estimates in the form of a ***separation index***. The Rasch equivalent of Cronbach's alpha, commonly referred to as the ***separation reliability coefficient***, is closely related to a separation index which reflects the replicability of the person placements along the measurement scale. The separation index can also be reported as

***G***, which reflects how well a set of items spread the group of individuals tested into statistically distinct levels (strata) of ability when the strata are separated by 3 *SE* units. Finally, traditional estimates of Cronbach's alpha tend to be higher than the Rasch equivalent estimates, as the Rasch equivalent form does not take into account extreme scores that tend to inflate traditional estimates of reliability (Bond & Fox, 2007; Smith, 2001).

### 15.11.1 Test-retest reliability

Test-retest reliability coefficients provide indices of the stability of the ADL ability measures of the AMPS over time. Test-retest reliability is evaluated by assessing persons twice within a relatively short period of time during which we do not expect a change in ability.

Data from a small heterogeneous sample of 17 older participants (mean age = 81, *SD* = 8 years, range = 67 to 97) revealed ***test-retest reliabilities of r = .91 for the AMPS motor skill scale and r = .90 for the AMPS process skill scale*** (unpublished data). The participants included both well persons and persons with neurologic, musculoskeletal, or cognitive disabilities. These participants were retested 1 to 2 weeks after their first AMPS observation. In a more comprehensive study, Rockwood et al. (1996) obtained ***test-retest reliability coefficients of r = .90 and r = .87 for the AMPS motor skill scale and AMPS process skill scale***, respectively. These participants were older adults with dementia, memory impairments without dementia, musculoskeletal disorders, or medical disorders. The mean interval between the two AMPS observations was 4.5 days (*SD* = 3.5, range = 1 to 18 days). With samples of persons with stable functional abilities, the results of these two studies reveal high test-retest reliability and provide evidence that AMPS ADL ability measures remain consistent from one test session to the next.

### 15.11.2 Parallel Forms Reliability

***Parallel forms reliability*** coefficients provide indices of the stability of AMPS ADL ability measures when different sets of ADL tasks (strictly parallel alternative forms) are utilized to generate ADL ability measures. That is, if a person performs different ADL tasks each time he or she is tested, each pair of tasks is an alternate form. Because the different (alternate) "forms" have equivalent means, *SD*s, and *SE*s, they can be considered to be parallel forms (Feldt & Brennan, 1989; Haertel, 2006). For the AMPS, parallel forms reliability is particularly important to evaluate because occupational therapy practitioners often reevaluate persons over the course of intervention and allow the persons to choose different sets of AMPS tasks to perform.

Kirkley and Fisher (1999) examined the alternate forms reliability[10] of 91 persons who had performed four different AMPS tasks within a 7 day period. Their participants included well older adults as well as persons at least 16 years of age with developmental, cognitive, musculoskeletal, psychiatric, neurologic, or medical disorders. When Kirkley and Fisher compared the ADL motor and ADL process ability measures based on tasks 1 and 2 to the ADL motor and ADL process ability measures based on tasks 3 and 4, they obtained ***alternate forms reliability coefficients of r = .91 for the AMPS motor skill scale and r = .85 for the AMPS process skill scale***. When they compared the ADL ability measures for task 1 only to the ADL ability measures for task 2 only, the reliability coefficients dropped to $r = .81$ and $r = .71$ for the AMPS motor skill scale and the AMPS process skill scale, respectively. This attenuation in reliability coefficients underscores the importance of having persons perform at least two AMPS tasks during a single AMPS observation.

For both scales, 80% of the paired ability measures remained stable within ± 0.5 logit. For the 20% that varied more than ± 0.5 logit, three potential reasons were identified:

- ***Actual fluctuations ADL ability*** — real day to day variations in ADL ability measures, often associated with the person's diagnoses, that were detected when the AMPS was re-administered time 2 — approximately 25% of the significant differences in ability measures Kirkley and Fisher observed (5% of the total sample)

- ***Performance of AMPS tasks that were too easy*** — error associated with allowing the person to choose and perform poorly targeted tasks (see Chapter 5, Section 5.2.3) —approximately 40% of the significant differences in ability measures Kirkley and Fisher observed (8% of the total sample)

[10] Kirkley and Fisher (1999) used the term alternate forms in their paper so we have retained the use of that term here. The terms parallel forms and alternate forms are often used interchangeably, but strictly speaking, the term parallel forms is more common, provided the forms compared are equivalent psychometrically (Feldt & Brennan, 1989; Haertel, 2006).

- ***Indentified error*** — error appeared to be random and commonly was associated with a single AMPS task performance that was noticeably worse than the other three — approximately 35% of the significant differences in ability measures Kirkley and Fisher observed (7% of the total sample)

**If two AMPS ADL ability measures differ by more than 0.5 logit, there is a 93% chance that the difference between the two measures is significant.**

***The following caution is in order*:**
**When a person performs two AMPS tasks, and the person's performance on one of the two tasks is markedly better or worse than the other, the AMPS rater will have to determine if**

- **One task was too easy (*poorly targeted* — raw scores predominately 4s)**
  - **If one task was poorly targeted, a third task performance should be observed and scored. The third task should be one that is more difficult, based on the AMPS task challenge calibration values reported in Tables 15-4 and 15-5**
- **The person "just" performed more poorly on one task than the other, but neither task was poorly targeted (i.e., raw scores on both tasks included several 2s and/or 1s)**
  - **If the person's performance on one task was markedly below that of his or her performance on the other task, but neither task was poorly targeted, consider offering a third task option which is at about the same level of difficulty as the tasks the person already performed**
- **If one task performance appears to be invalid, do not include data for that task performance in the determination of the person's ADL ability measures**

**See Section 15.11.6 for more information about determining if two AMPS measures differ significantly**

### 15.11.3 Rasch Equivalent of Cronbach's Alpha: Separation Reliability

When the data for the current standardization sample were analyzed (see Section 15.2.1), the many-faceted Rasch equivalent of Cronbach's alpha was $R$ = .92 for the ADL motor measures and $R$ = .91 for the AMPS process measures, again supporting very high reliability of the AMPS measures.

### 15.11.4 Separation Index and *G*

Again based on the current standardization sample, the separation index for the AMPS motor scale was 3.43, indicating that the sample could be reliably divided into at least 4.9 distinct strata separated by 3 *SE* (Fisher, 1992; Smith, 2001). For the AMPS process scale, the separation index was 3.21, indicating that the sample could be reliably divided into at least 4.6 distinct strata.

**When different methods have been used to estimate the reliability coefficient of the ADL motor and ADL process ability measures, the results have consistently supported high reliability:**

| | ADL motor | ADL process |
|---|---|---|
| • Test-retest: | $r$ = .91<br>$r$ = .90 | $r$ = .87<br>$r$ = .90 |
| • Parallel forms: | $r$ = .91 | $r$ = .85 |
| • Rasch equivalent of Cronbach's alpha: | $r$ = .92 | $r$ = .91 |

### 15.11.5 Standard Error of Measurement (*SE*)

As noted in Section 15.11, the *SE* is often considered the most useful reliability index as it can be used to evaluate not only the reliability of a person's obtained ADL measures, but it also provides an index that reflects the sensitivity of the measures in the context of evaluating change. The mean *SE* for those persons in the current standardization sample who performed two AMPS tasks ($n$ = 146,418) was 0.25 logit on the ADL motor scale and 0.20 logit on the ADL process scale.

While the mean *SE* provides an overall estimate of the reliability of the individual ADL motor and ADL process ability measures, the magnitude of the *SE* varies along the length of a scale. The traditional *SE* (calculated indirectly from the reliability coefficient, *r*, for an entire sample) will be greatest near the center of a scale and gradually decrease toward the ends of the scale. Rasch-based *SE*s (calculated directly for each person) vary such that the *SE* becomes larger at extreme upper and lower ends of the scale (Harvill, 1991; Smith, 2001).

We, therefore, also calculated the mean *SE* for equal intervals along each AMPS scale (see Table 15-12). For those persons whose ADL motor or ADL process ability measures fall within the range delineated by grey shading, the overall standardization sample mean *SE* provides a reasonable estimate of the person's actual *SE*; for those outside that range, larger *SE* values likely apply.

### 15.11.6 Using the Standard Error of Measurement (*SE*) to Evaluate for Significant Changes in ADL Ability

The mean standardization sample *SE* values of 0.25 logit for the ADL motor scale and 0.20 logit for the ADL process scale provide reasonable approximations for 81.5% and 92.4% of the standardization sample's ADL motor or ADL process ability measures, respectively (see Table 15-12). Therefore, these values have been incorporated into the AMPS computer-scoring software, and are used when the occupational therapist requests a Progress Report.

More specifically, when a person has been tested twice, and the two measures are compared (or, if two different persons are compared), the occupational therapist can use the person's *SE* values to determine if a significant change is likely to have occurred. According to Harvill (1991), if two individual measures differ by at least 2 *SE*, then there is a reasonable chance that the two measures differ (i.e., there is a significant difference).[11]

[11] A more conservative approach to evaluating for significant differences, based on standardized *Z*, were $Z = (\text{measure 1} - \text{measure 2})/\text{Sqrt}((SE\ 1)^2 + (SE\ 2)^2)$ is sometimes suggested. When *Z* is used, two measures must differ by almost 3 *SE* to be considered to differ significantly. We have chosen to use Harvill's criterion as its validity has been cross-validated by the results of Kirkley and Fisher's (1999) alternate forms reliability study, where it was found that there is a 93% chance of a significant difference between two AMPS measures that differ by 0.5 logit, provided the AMPS tasks the person performed were of an appropriate challenge. Kirkley and Fisher's 0.5 logit criterion can be compared to the 0.5 logit and 0.4 logit criteria used in the AMPS computer-scoring software for identifying a significant difference between two ADL motor and two ADL process ability measures, respectively.

**Table 15-12 Standard Error of Measurement (*SE*) for ADL Motor and ADL Process Ability Measures Based at Different Locations Along the AMPS Scales***

| | ADL motor | | ADL process | |
|---|---|---|---|---|
| **ADL measure (range)** | ***n*** | ***SE*** | ***n*** | ***SE*** |
| -3.00 to -2.51 | 116 | 0.39 | – | – |
| -2.50 to -2.01 | 350 | 0.41 | – | – |
| -2.00 to -1.51 | 926 | 0.42 | 681 | 0.34 |
| -1.50 to -1.01 | 1979 | 0.41 | 1664 | 0.28 |
| -1.00 to -0.51 | 4159 | 0.37 | 4382 | 0.25 |
| -0.50 to 0.00 | 10123 | 0.31 | 12422 | 0.21 |
| 0.01 to 0.50 | 18319 | 0.25 | 29332 | 0.19 |
| 0.51 to 1.00 | 27053 | 0.23 | 43893 | 0.19 |
| 1.01 to 1.50 | 30759 | 0.22 | 34544 | 0.20 |
| 1.51 to 2.00 | 26503 | 0.23 | 15169 | 0.21 |
| 2.01 to 2.50 | 16701 | 0.25 | 3593 | 0.27 |
| 2.51 to 3.00 | 7167 | 0.32 | 738 | 0.35 |
| 3.01 to 3.50 | 2065 | 0.43 | – | – |
| 3.51 to 4.00 | 198 | 0.56 | – | – |
| Total | 146418 | 0.25 | 146418 | 0.20 |

* When a person's ADL motor or ADL process ability measure falls within the range delineated by grey shading, the average ADL motor and ADL process *SE* values used in the AMPS computer-scoring software provide reasonable estimates for determining if a person's measures changed significantly between two different AMPS observations (see Section 15.11.6 for more details).

When a Progress Report is requested, the AMPS computer-scoring software will compute the difference between each pair of ADL motor ability measures and each pair of ADL process ability measures, and determine if a significant change has occurred. The person's two ADL motor and two ADL process ability measures will also be plotted graphically on the Progress Report, with each measure depicted by a small square, and error bars delineating ±1 *SE*. Provided that both of the person's ADL motor and ADL process ability measures fall within the range depicted by the grey shaded area in Table 15-12, it is likely that the person's paired ADL measures differ significantly when the two error bars do not overlap.

If one or more of the person's ADL motor or ADL process ability measures fall outside the range depicted by the grey shaded area on Table 15-12, the occupational therapist will need to calculate the magnitude of difference between two ADL measures required for likelihood of statistical significance. For example:

*Klaus evaluated Marieke on two different occasions, once before he implemented any interventions, and again, after 2 weeks of occupational therapy intervention. Marieke's ADL motor measure for Time 1 was -1.4 logits. When Klaus looked up that value in Table 15-12, he found that an ADL motor measure in that range was associated with an average SE of 0.41 logit. Marieke's ADL motor ability measure Time 2 was -0.1, and Klaus noted, by again looking at Table 15-12, that the associated SE was 0.31 logit. By then summing these two SEs (i.e., 0.41 + 0.31 logit), Klaus determined that Marieke's two ADL motor measures would need to differ by at least 0.72 logit if he was to consider them to differ significantly. Since Marieke's ADL motor ability measures increased by 1.3 logits, a value that is greater than the 0.72 needed for evidence that her ADL ability measures differ significantly, Klaus concluded that Marieke likely had made significant improvements. Before making his final decision, Klaus considered the warnings in Section 15.11.2, and ensured that both tasks Marieke had performed were of sufficient challenge and that there was no marked discrepancy between Marieke's two sets of AMPS raw item scores.*

*When Klaus considered Marieke's ADL process ability measures of -0.4 logit Time 1 and 0.1 logit Time 2, a change of 0.5 logit, he noted that both measures were within the range delineated in grey in Table 15-12. Klaus reasoned, therefore, that the results reported by the AMPS computer-scoring software for Marieke's ADL process ability measures will match those he would have obtained by using the SE values in Table 15-12. That is, the respective SEs in Table 15-12 are 0.21 and 0.19 logit, which sum to 0.40 logit. This value is exactly the same as the sum of the mean ADL process SEs used in the AMPS computer-scoring software (i.e., 0.20 + 0.20 logit = 0.40 logit). Again, Klaus concluded that Marieke's ADL process ability had improved significantly.*

## 15.12 Current Evidence that the ADL Ability Measures of the AMPS Demonstrate Expected Relationships with Other Test Scores

A common method used to demonstrate that a test is evaluating the intended construct is to examine the extent to which the measures or scores from the new tool correlate with other existing tools known to test the same construct. In the case of the AMPS, there are no equivalent assessment methods. Our efforts to demonstrate concurrent validity, therefore, have required that we compare the ADL ability measures of the AMPS to the results of a variety of other tests, including global tests of ADL and tests that are designed to evaluate discrete body function deficits (i.e., underlying cognitive, biomechanical, neuromuscular, or psychosocial impairments). Global ADL scales commonly are designed to evaluate how much assistance a person needs with global ADL tasks (e.g., eating, bathing, meal preparation). These evaluations are often supplemented by discrete evaluation of body functions. Generally, only low positive relationships ($r$ = .30 to .50) between these discrete evaluations and global ADL test scores have been reported (cf. Bernspång, Asplund, Eriksson, & Fugl-Meyer, 1987; Jongbloed, Brighton, & Stacey, 1988; Judge, Schechtman, Cress, & the FICSIT Group, 1996; Pincus et al., 1989; Reed, Jagust, & Seab, 1989; Skurla, Rogers, & Sunderland, 1988; Teri, Borson, Kiyak, & Yamagishi, 1989).

The AMPS represents an evaluation method that is at a level somewhere between global ADL and discrete body function. Like global ADL evaluations, the AMPS is a test of ADL task performance. Unlike these global scales, however, the AMPS is used to evaluate the quality of the observable goal-directed actions of task performance. Problems with any of these actions must not be equated with underlying cognitive, biomechanical, neuromuscular, or psychosocial impairments. Since the AMPS is not designed to evaluate the same constructs as are tests of global ADL ability or discrete impairments, we expect moderate positive relationships ($r$ = .50 to .70) between AMPS ability measures and results of such assessments (Hinkle, Wiersma, & Jurs, 1988). Moreover, higher relationships should be limited to the AMPS scales that are most related to the domain being tested by the concurrent measure.

The results of validity studies exploring the relationship between the ADL motor and ADL process ability measures and scores of other tests are summarized in Table 15-13. Overall, the results of these studies have supported our assertion that there is a low to moderate relationship between the AMPS scales and discrete tests of physical or mental status. More specifically, the relationship between tests of mental status and the AMPS motor scale is, generally, low and not significant, as is the relationship between tests of physcial status and the AMPS process scale. Similarly, the results of several different studies have verified that the ADL motor and ADL process ability measures of the AMPS have significant, yet moderate, associations with other assessments of ADL. These moderate associations indicate that the ADL ability measures of the AMPS are unique and different constructs; the quality of ADL task performance, as measured with the AMPS, is a unique construct that is not assessed with other ADL assessments. In contrast to other assessments of ADL ability, which are often used to globally measure whether or not a person can independently and/or safely complete ADL tasks, the AMPS is used to measure the quality of performance as a person enacts 36 universal, observable, goal-directed actions of performance. Through documenting and scoring the effort, efficiency, safety, and independence (i.e., quality of performance) that a person demonstrates as he or she enacts the smallest units of performance (i.e., ADL motor and ADL process skills), occupational therapists are able to measure two unique constructs — ADL motor and ADL process ability.

## 15.13 Current Evidence Supporting the Use of the ADL Ability Measures of the AMPS as Outcome Measures

In this section, we will explore how the AMPS has been utilized as an outcome measure in health-related research. Prior to summarizing current research findings, we will provide the reader with a brief background of some of the key measurement principles related to outcome measures. Among those principles are issues related to reliability, sensitivity, and critiquing and interpreting research findings.

**Table 15-13 Relationship between ADL Motor and ADL Process Ability Measures of the AMPS, and Scores from Tests of Global ADL Ability, Self-reported ADL Ability, Environmental Safety, Underlying Body Functions, and Functional Health**

| Author, year and participants | Instrument | Relation to AMPS measures | |
|---|---|---|---|
| | | ADL motor | ADL process |
| ***Global ADL*** | | | |
| Bryze, 1991 | Scales of Independent Behavior (SIB) (Bruininks, Woodcock, Weathermen, & Hill, 1985) | | |
| • Persons with developmental disabilities | • Broad Independence scale | *r* = .62 | *r* = .66 |
| | • Motor Skills scale | *r* = .85 | NR** |
| | • Personal Living scale | *r* = .50 | *r* = .71 |
| Doble, Fisk, MacPherson, Fisher, & Rockwood, 1997 | OARS (Fillenbaum, 1988) — IADL scale | | |
| • Community dwelling elderly with and without Alzheimer's disease | • Nondemented community dwelling older adults | *r* = .68 | *r* = .42 |
| | • Community dwelling older adults with dementia | *r* = .32 | *r* = .58 |
| Rockwood et al., 1996 | OARS (Fillenbaum, 1988) | | |
| • Older adults with biomechanical, neuromuscular, and cognitive (dementia, memory loss without dementia) disorders | • PADL scale | NR** | *r* = .55 |
| | • IADL scale | *r* = .50 | *r* = .56 |

(continued)

*NS = results were reported as not significant; correlation coefficient was not reported

**NR = correlation coefficient was not reported

**Table 15-13 (continued)**

| Author, year and participants | Instrument | Relation to AMPS measures | |
|---|---|---|---|
| | | ADL motor | ADL process |
| ***Global ADL* (continued)** | | | |
| Robinson & Fisher, 1996<br>• Persons with dementia or memory impairments | Functional Independence Measure™ (FIM™) (Keith, Granger, Hamilton, & Sherwin, 1987) | | |
| | • Motor scale | *r* = .62 | *r* = -.07 |
| | • Social-cognition scale | *r* = .13 | *r* = .62 |
| Månsson & Lexell, 2004<br>• Persons with multiple sclerosis | FIM™ (Keith et al., 1987) | | |
| | • Motor domain scores | *r* = .38 | *r* = .46 |
| | • Cognitive domain scores | *r* = .14 | *r* = .36 |
| Rockwood et al., 1996<br>• Older adults with biomechanical, neuromuscular, and cognitive (dementia, memory loss without dementia) disorders | Barthel Index (Mahoney & Barthel, 1965) | *r* = .53 | *r* = .54 |
| ***Self-reported ADL Ability*** | | | |
| Doble et al., 1994<br>• Persons with multiple sclerosis | Sickness Impact Profile (Bergner, Babbit, & Pollard, 1976) | | |
| | • Ambulation subscale | *r* = -.65 | *r* = -.27 |
| | • Body Care subscale | *r* = -.69 | *r* = -.23 |
| | • Mobility subscale | *r* = -.45 | *r* = -.08 |
| | • Home Management subscale | *r* = -.19 | *r* = -.20 |

(continued)

*NS = results were reported as not significant; correlation coefficient was not reported
**NR = correlation coefficient was not reported

**Table 15-13 (continued)**

| Author, year and participants | Instrument | Relation to AMPS measures | |
|---|---|---|---|
| | | **ADL motor** | **ADL process** |
| ***Self-reported ADL Ability* (continued)** | | | |
| Albert et al., 2006<br>• Nondemented older adults | Self-reported IADL (i.e., using the telephone, completing light house work, preparing light meals, doing light shopping, handling finances, and managing finances) | *r* = -.34 | NS* |
| ***Environmental safety*** | | | |
| McNulty & Fisher, 2001<br>• Persons with psychiatric disorders | • Safety Assessment of Function and the Environment for Rehabilitation (SAFER) (Community Occupational Therapists & Associates, 1991) | *r* = .58 –.63 | *r* = .60 –.67 |
| | • Predicting Safety — AMPS conducted in home | | |
| | ○ Sensitivity | NR** | .78 |
| | ○ Specificity | NR** | .82 |
| | ○ Overall predictive value | NR** | .80 |
| ***Cognition when impact of cognitive impairments on ADL is considered*** | | | |
| Rockwood et al., 1996<br>• Older adults with biomechanical, neuromuscular, and cognitive (dementia, memory loss without dementia) disorders | Global Deterioration Scale (GDS) (Reisberg, Ferris, de Leon, & Crook, 1982) | *r* = -.43 | *r* = -.80 |

(continued)

*NS = results were reported as not significant; correlation coefficient was not reported
**NR = correlation coefficient was not reported

**Table 15-13 (continued)**

| Author, year and participants | Instrument | Relation to AMPS measures | |
|---|---|---|---|
| | | **ADL motor** | **ADL process** |
| ***Cognition when impact of cognitive impairments on ADL is considered* (continued)** | | | |
| Bouwens et al., 2008<br>• Cognitively impaired older adults (e.g., Alzheimer's disease, vascular dementia, depression) | GDS (Reisberg, et al., 1982) | NR** | *r* =.66 |
| Rockwood et al., 1996<br>• Older adults with biomechanical, neuromuscular, and cognitive (dementia, memory loss without dementia) disorders | Brief Cognitive Rating Scale, Axis V (Reisberg & Ferris, 1988) | *r* = -.53 | *r* = -.75 |
| ***Mental status*** | | | |
| Doble et al., 1997<br>• Community dwelling elderly with and without Alzheimer's disease | Mini-Mental State Exam (MMSE) (Folstein, Folstein, & McHugh, 1975) | *r* = .22 – .37 | *r* = .64 |
| Robinson & Fisher, 1996<br>• Older adults with dementia or memory loss | MMSE (Folstein, et al., 1975) | *r* = -.01 | *r* = .67 |
| Poole et al., 2006<br>• Women with systemic lupus erythematosus (SLE) | MMSE (Folstein, et al., 1975) | NS* | *r* = .6 |

(continued)

*NS = results were reported as not significant; correlation coefficient was not reported
**NR = correlation coefficient was not reported

**Table 15-13 (continued)**

| Author, year and participants | Instrument | Relation to AMPS measures | |
|---|---|---|---|
| | | ADL motor | ADL process |
| ***Mental status* (continued)** | | | |
| Mori & Sugimura, 2007<br>• Elderly persons with and without dementia | MMSE (Folstein, et al., 1975) | *r* = .01 | *r* = .38 – .45 |
| Bouwens et al., 2008<br>• Cognitively impaired older adults (e.g., Alzheimer's disease, vascular dementia, depression) | • MMSE (Folstein, et al., 1975)<br>• Cognitive component (CAMCOG) of the Cambridge Examination for Mental Status in the Elderly (CAMDEX) (Roth et al., 1986) | NR**<br>NR** | *r* = .54<br>*r* = .58 |
| Robinson & Fisher, 1996<br>• Older adults with dementia or memory loss | CAMCOG (Roth et al., 1986) | *r* = -.04 | *r* = .66 |
| Mori & Sugimura, 2007<br>• Elderly persons with and without dementia | Rivermead Behavioral Memory Test (RBMT) (Wilson, Cockburn, & Baddeley, 2003)<br>• Prospective memory<br>• Retrospective memory<br>• Prospective and retrospective memory | <br><br>*r* = .12<br>*r* = .16<br>*r* = .03 | <br><br>*r* = .06<br>*r* = .41 – .48<br>*r* = .39 – .42 |
| Marom, Jarus, & Josam, 2006<br>• Persons post stroke | Large Allen Cognitive Levels (LACL) (Allen, 1985) | *r* = .57 | *r* = .66 |

(continued)

*NS = results were reported as not significant; correlation coefficient was not reported
**NR = correlation coefficient was not reported

**Table 15-13 (continued)**

| Author, year and participants | Instrument | Relation to AMPS measures | |
|---|---|---|---|
| | | **ADL motor** | **ADL process** |
| ***Discrete physical impairments*** | | | |
| Doble et al., 1994<br>• Persons with multiple sclerosis | Expanded Disability Status Scale (EDSS) (Kurtzke, 1983) | *r* = -.79 | *r* = -.55 |
| Månsson & Lexell, 2004<br>• Persons with multiple sclerosis | Expanded Disability Status Scale (EDSS) (Kurtzke, 1983) | *r* = -.01 | *r* = -.21 |
| Mallinson, Cella, Cashy, & Holzner, 2006<br>• Persons receiving chemotherapy cancer treatment | Multidimensional Fatigue Symptom Inventory (MFSI) (Stein, Jacobsen, Blanchard, & Thors, 2004). | | |
| | • MFSI motor items | *r* = -.38 – -.46 | *r* = -.15 – -.28 |
| | • MFSI cognitive items | *r* = -.23 – -.33 | *r* = -.11 – -.21 |
| Norberg, Boman, & Löfgren, 2010<br>• Older persons with chronic heart failure | Swedish version (Fürst & Åhsberg, 2001) of the Multidimensional Fatigue Inventory (MFI-20) (Smets, Garssen, Bonke, & De Haes, 1995) | Persons above the 2.0 logits competence cutoff had significantly lower fatigue scores than those below | No difference in fatigue scores between those above and below the 1.0 logit competence cutoff |

(continued)

*NS = results were reported as not significant; correlation coefficient was not reported
**NR = correlation coefficient was not reported

**Table 15-13 (continued)**

| Author, year and participants | Instrument | Relation to AMPS measures | |
|---|---|---|---|
| | | ADL motor | ADL process |
| ***Discrete physical impairments* (continued)** | | | |
| Mallinson et al., 2006<br>• Persons receiving chemotherapy cancer treatment | • Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F) subscale (Yellen, Cella, Webster, Blendowski, & Kaplan, 1997) | $r$ = .42 – .45 | $r$ = .15 – .40 |
| Poole et al., 2006<br>• Women with systemic lupus erythematosus | Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) (Bombardier, Gladman, Urowitz, Caron, & Chang, 1992) | $r$ = -.84 | NS* |
| ***Functional health and activity limitations*** | | | |
| Mallinson et al., 2006<br>• Persons receiving chemotherapy cancer treatment | • Physical functioning subscale (PF-10) of the SF-36 (Ware, Snow, Kolinski, & Gandeck, 1993) | $r$ = .56 – .71 | $r$ = .28 – .44 |

*NS = results were reported as not significant; correlation coefficient was not reported
**NR = correlation coefficient was not reported

**Considered together, the studies summarized in Table 15-13 provide strong evidence that:**

- **ADL motor and ADL process ability measures of the AMPS are indicators of the quality of ADL task performance**
- **ADL motor and process ability measures of the AMPS represent unique constructs that are not assessed with other ADL assessments.**
- **ADL motor and ADL process ability measures of the AMPS cannot be equated with body functions (e.g., physical capacity, strength, memory)**
- **Tests of body functions cannot be used to validly and accurately predict ADL task performance**

### 15.13.1 Reliability and Sensitivity of Outcome Measures

Two important features of an outcome measure are reliability and sensitivity. First, when considering the reliability of an ADL outcome measure, we must ask ourselves if the differences between two sets of ADL measures represent actual differences in ability or measurement error. If researchers have confidence that a change in a person's ADL measure represents an actual change in ADL ability, then they can have greater confidence when interpreting the results of a study (e.g., there is greater confidence that an intervention strategy resulted in improved ADL task performance). The reliability of the AMPS measures is discussed in Section 15.11.

Another important quality of an outcome measure is how sensitive the measure is as an indicator of change. Here, we must ask ourselves if small, discrete changes in ADL ability can be measured (i.e., detected) with the assessment tool. Sensitivity is, in large part, reflected by the size of the standard error of measurement (*SE*) (see Section 15.11.5).

Another issue that can affect sensitivity is related to floor and ceiling effects. Ceiling effects occur when more able persons obtain maximum scores on the assessment. For example, with a sample of persons with multiple sclerosis, Månsson and Lexell (2004) found substantial ceiling effects when using the Functional Independence Measure™ (FIM™) with persons who were independent (or nearly independent) in the performance of PADL tasks. However, when the AMPS was utilized with this same group of persons, ceiling effects were not encountered as the tasks available within the AMPS offered a wider range of task challenge and included both PADL and IADL tasks.

When considering ceiling effects, it is important to keep in mind that a group of persons who obtain maximum scores on an ADL assessment may actually differ in their levels of ADL ability. Moreover, persons who receive maximum scores may actually have an ADL disability. In such cases, the ADL assessment results cannot be used to document their problems of ADL task performance nor can they be used to differentiate among the ADL abilities of the more able persons (i.e., more or less able persons may obtain the same maximum scores, despite observed differences in their ability). Similar problems are encountered with persons who are very frail/disabled who obtain the lowest possible score on an assessment (floor effects).

Ultimately, a problem with sensitivity is often vocalized by clinicians in terms of their frustration or dissatisfaction when assessment tools fail to produce scores or measures that reflect the improvements that they have observed (e.g., occupational therapists observe improved ADL ability, yet the scores or measures between two test sessions did not change in a meaningful or significant manner). ***This problem of limited sensitivity is minimized when an occupational therapist uses the AMPS*** because he or she observes and scores the smallest observable units of ADL task performance (i.e., ADL motor and ADL process skills). Through scoring the person's quality of performance on each of the 16 ADL motor and 20 ADL process skills, across two separate task observations, a total of 64 items are scored. Thus, very small changes in quality of ADL task performance can be detected and measured. Moreover, the AMPS includes 116 ADL tasks that vary in challenge, from very easy to very hard. ***Provided the occupational therapist ensures that the person observed performs ADL tasks that are chosen, relevant, and offer some degree of a challenge, the AMPS provides a measurement system that is very sensitive to small changes in ADL ability***.

In previous sections of this chapter we have discussed the sensitivity of the AMPS ability measures, with regard to detecting changes in ADL ability for persons of varying ages and levels of disability (see Sections 15.6, 15.7, 15.9 and 15.10). In the sections that follow, we will discuss research supporting the sensitivity of the AMPS ability measures when used to detect changes in ADL ability following intervention.

### 15.13.2 Some Words of Caution In Relation to Interpreting the Results of Research Where the AMPS Was Used As an Outcome Measure

Prior to discussing research studies in which the ADL motor and ADL process ability measures were utilized as outcome measures, we want to offer a few points of caution to consumers of AMPS-related research. All research must be read with a critical and informed perspective. In the case of the AMPS, three special precautions are in order:

- ***The reader should be very cautious when interpreting the results of studies in which raw AMPS scores have been summed and compared***. Raw ordinal AMPS scores (i.e., skill item scores of 1, 2, 3, or 4) can be utilized by occupational therapists to document and describe a person's skill strengths and challenges; however, raw scores are not valid for comparing performance from one assessment session to another or between two different persons (Chapter 9, Section 9.1.2). The only valid means of comparing AMPS evaluation results across time or between persons is through the comparison of linear ADL motor and/or ADL process ability measures. Such measures take into consideration the difficulty of the skill items, challenge of the tasks performed, and the severity of the rater who scored the person's ADL task performance.

- ***The reader must remain alert to the fact that many authors continue to erroneously equate ADL motor ability and/or ADL motor items with physical body functions and ADL process ability and/or ADL process items with cognitive body functions***. As we have discussed in Chapter 1, Section 1.1 and Chapter 8, Section 8.4.3, the ADL motor and ADL process skills are observable actions of performance, not internal body functions (e.g., strength, coordination, attention, memory). While we recognize that impairments of body functions can impact the quality of a person's ADL task performance, body functions and quality ADL task performance are two different constructs. See Section 15.12 for further discussion of the relation between the AMPS and tests of body functions.

- ***The reader should remain critical of intervention studies that report only ADL motor ability or ADL process ability, not both***. This precaution is likely closely related to that above, as many authors erroneously assume that interventions aimed at improving or compensating for diminished physical body functions only influence ADL motor ability and that interventions aimed at improving or compensating for diminished cognitive body functions only influence ADL process ability. In such instances, authors may only report changes in either ADL motor ability or ADL process ability. Since the AMPS is utilized to measure quality of ADL task performance, not body functions, it is important to recognize that many intervention strategies, regardless of their intended impact on body functions, may result in changes in both ADL motor and ADL process ability. For example, teaching a person to use a walker tray to transport task objects as a means of compensating for physical body function impairments (e.g., decreased balance, hemiplegia) may result in increased efficiency of the ADL task performance.

### 15.13.3 Results of AMPS-related Outcome Research

Recognizing that we cannot summarize the results of all of the published AMPS-related outcome studies, we have selected a few different types of studies which highlight how the AMPS has been utilized in health-related research. These studies reflect a variety of methodological approaches, and hence, varying degrees of overall quality of the study and confidence one can have in the research findings. While there are differing beliefs with regards to what constitutes "excellent" research, there is a general trend that systematic reviews of the literature and randomized controlled trials (RCTs) constitute the highest levels of evidence, while case studies and expert opinions generally reflect the lowest levels of evidence (e.g., Evans, 2003; Straus, Richardson, Glasziou, & Haynes, 2005). In the text that follows, we summarize a few research studies, starting with a single case study design and working up to randomized controlled trials. While the latter likely provides the strongest form of evidence, many studies of lower quality that cross-validate one another can ultimately also be a source of strong evidence (Asplund, Jonsson, & Britton, 2002). Each researcher must carefully balance the constraints of "the realities of practice" and the "idealistic" research designs. Moreover, ***each occupational therapist, in the context of evidence-based practice, has the possibility to gather evidence that the methods he or she uses are effective***. Such evidence can begin with the use of simple research designs that are clinically realistic.

**Single Case Study Design — Pretest – Posttest**

While single case studies may not offer the most powerful level of evidence, such studies do provide some evidence to support clinical decision-making. Simple pretest – posttest single case studies offer a feasible and "clinically accessible" means of systematically gathering evidence — in fact occupational therapists can treat each client as a single case study to evaluate the effectiveness of the intervention provided. As an example, Kuiken et al. (2007) utilized the AMPS as an outcome measure in their investigation of the efficacy of ***targeted sensory reinnervation*** (TSR) surgery on a woman with a left ***arm amputation*** at the humeral neck. After completion of the TSR surgery and training with her newly innervated myoelctric prosthesis, her ADL motor ability improved from 0.30 to 1.98 logits and her ADL process ability improved from 0.90 to 1.98 logits, likely representing both clinically and statistically significant improvements on both scales.

**Posttest Only Design — Comparative Analysis**

In a comparative analysis of non-equivalent groups, Parks, Rasch, Mansky, and Oakley (2009) evaluated the ***late effects of pediatric sarcoma therapy***. Parks and her colleagues found that a group of persons ($n = 32$) who were an average of 17 years post cancer treatment, had significantly lower ADL motor and ADL process ability measures than their healthy age and gender matched peers. In fact, the study participants had mean ADL motor ability measures that were statistically similar to healthy persons who were 10 years their senior and mean ADL process ability measures that were statistically similar to groups of healthy persons who were 10 and 20 years older. These results support an overall trend of early functional decline for persons who have undergone treatment for pediatric sarcoma, and "suggest that the influences of treatment late effects on performance of daily life activities among pediatric sarcoma survivors are more widespread than reported, even among many survivors with no noticeable physical impairment" (Parks et al., 2009, p. 5).

**One Group Pretest-posttest Design — No Control Group**

Next, we explore pretest-posttest studies where a control group was not utilized. In other words, the researchers evaluated the outcomes of a group of persons who had received an intervention, but the participants' outcomes were not compared to a group of persons who did not receive the intervention. While this methodological approach provides stronger evidence than does a case study, this level of evidence is still on the lower end of the scale, as there is insufficient evidence to verify if time or other factors influenced the outcomes. Despite such shortfalls, this is a relatively common approach in health-related research and program evaluation, especially when there are ethical considerations related to withholding intervention from a potential control group, and/or there is a high participant drop out rate among potential participants assigned to the control group when they realize that they will not receive the intervention.

A pilot ***pharmacological study*** conducted by Oakley, Khin, Parks, Bauer, and Sunderland (2002) provides one such example. In this study, the researchers investigated ADL changes in a group of ***elderly persons with post-bereavement depression*** ($n$ = 10). The preliminary study was a double-blind parallel trial of two antidepressant medications, Sertraline and Nortriptyline. Pre-treatment and post treatment ADL motor and ADL process ability measures were analyzed and also compared to those of healthy, age-matched peers. There was no control group of persons with post-bereavement depression who did not receive pharmacological intervention. The results reveal that prior to treatment, the group of persons with depression had significantly lower ADL motor and ADL process ability measures than did the group of well, age-matched peers. Thus, prior to treatment, the group of persons with bereavement induced depression demonstrated increased clumsiness or physical effort, inefficiency, decreased safety, and/or decreased independence when performing familiar ADL tasks. Furthermore, the majority of the persons fell within or below the "risk zone" (i.e., ± 0.30 logit of the ADL motor and/or ADL process cutoff measures), indicating the potential need for assistance to live in the community. In response to the medication regime, there was a significant, positive change in both mean ADL motor and ADL process ability ($p < 0.010$ and 0.001, respectively). Additionally, after treatment, the group of persons with depression did not differ in ADL motor or ADL process ability when compared to their healthy peers, and all of the study participants had ADL process ability measures that were above the risk zone. This study provided evidence that pharmacological intervention had a positive impact on everyday abilities, and that the AMPS may be useful as an outcome measure in future studies of this nature.

In a more recent study, Chard, Liu, and Mulholland (2009) investigated the efficacy of providing verbal cues and/or environmental modifications for women residing in an assisted living facility who were diagnosed with ***Alzheimer's disease or a dementia-related disorder*** ($n$ = 5). After baseline AMPS evaluations had been conducted, the participants' caregivers were trained in how to provide verbal prompts, verbal reinforcement, and environmental modifications as a means of supporting and engaging the participants in the performance of ADL tasks. Each caregiver utilized these methods over the course of 2 weeks when the participants were performing ADL tasks. After the 2 week intervention phase, a posttest AMPS evaluation for each participant was conducted. Following caregiver training and 2 weeks of implementing the strategies, ADL process ability significantly improved for all five of the participants, one participant demonstrated a significantly higher ADL motor ability measure, and one participant demonstrated a clinically meaningful improvement in ADL motor ability. Thus, with a small sample of older adults with dementia, each of whom had baseline ADL process ability measures below or near 0.0 logit, caregiver strategies focused on verbal cuing, verbal reinforcement, and environmental modification had a positive and significant impact on the quality of the participants' ADL performance. The importance of this evidence cannot be overlooked — occupational therapists have the potential to ***employ occupation-based compensatory strategies*** to promote continued engagement in daily life occupations with persons who have progressive cognitive and neurological impairments.

In another study, with a sample of persons with ***acquired brain injury*** ($n$ = 36), Wæhrens and Fisher (2007) investigated changes in the quality of ADL task performance following interdisciplinary rehabilitation. Results of paired $t$ tests indicated that mean ADL motor and ADL process ability measures significantly improved following rehabilitation ($p < 0.001$). Additionally, the effect size statistics ($d$) were 0.5 and 0.6, respectively for ADL motor and ADL process ability, indicating medium effects. When analyzing the data for clinically meaningful changes in ADL ability (i.e., a change in ability of at least 0.30 logit), 80.5% of the participants had improved ADL motor and/or ADL process ability measures following rehabilitation. Due to the ethical implications of not providing rehabilitation services to a group of persons, this study did not include a control group. However, the clients within this study were referred to the rehabilitation unit once the spontaneous recovery process had slowed and the potential for subsequent improvements in client outcomes were minimal. Thus, it was likely that this particular group of persons had ADL abilities that had reached a plateau prior to the implementation of the rehabilitation services and any changes in their ADL performance were likely the result of the rehabilitation services. These results indicate that following intensive, interdisciplinary rehabilitation, the quality of ADL task performance of the participants improved.

Kinnman, Andersson, Wetterquist, Kinnman, and Andersson (2000) utilized the AMPS as one of their outcome measures within their investigation of how ***cooling suits*** impact persons with ***multiple sclerosis*** ($n$ = 8). In this study, seven of the eight participants were able to complete the AMPS portion of the project, and all seven persons had ADL motor ability measures that were clinically higher following the use of the cooling suit. In contrast, two persons had higher ADL process ability measures following the use of the cooling suit. The researchers found that not only did the cooling suit result in a reduction of symptoms (e.g. balance, dexterity, spasticity, strength), it also had a positive impact on the quality of the ADL task performance.

Lastly, with a sample of adolescents with ***cerebral palsy*** ($n$ = 9), Bonnier, Eliasson, and Krumlinde-Sundholm (2006) documented the impact of ***constraint-induced movement therapy*** on hand and arm function and ADL task performance. In this study, Bonnier and colleagues found that overall, the participants demonstrated improved hand function with regard to dexterity, coordination, precision, and manipulative abilities, with many of these improved body functions remaining 5 months post treatment. When evaluated with the AMPS, ADL process ability did not improve significantly following treatment and ADL motor ability improved significantly for only two of the nine participants; these improvements were not maintained 5 months post treatment. Thus, while lasting improvements in hand function were documented, such improvements did not translate into improved ADL task performance.

**Pretest-posttest Design — Participants Serving As Their Own Control**

As stated above, due to the implications of withholding treatment from a group of persons, it can be challenging to design a study with a control group for comparison. One approach to this methodological barrier is to allow each participant to serve as his or her own control. When using this design, researchers assess the participants multiple times (at least twice) prior to introducing intervention. Then, the participants are tested again after the intervention is provided. While clinically less feasible, and increasing the risk for practice effects, the study design can be enhanced by testing each person before intervention until such time that the participant demonstrates stable measures. As there is some form of control, there is greater confidence that any changes in the outcome measure are a result of the intervention and not a result of time or extraneous environmental factors. As such, this approach provides a higher level of evidence than the various designs described above.

One such example is a study conducted by Fisher, Atler, and Potts (2007). In this study, the researchers used the AMPS as an outcome measure within their investigation of the efficacy of implementing true top-down occupational therapy services, guided by the Occupational Therapy Intervention Process Model (OTIPM) (Fisher, 2009), with a group of ***community dwelling, frail older adults*** (*n* = 8). During the first phase of the study, prior to implementing occupational therapy interventions, each participant was assessed twice with the AMPS. This phase of the research established the baseline stability of each participant's ADL task performance (i.e., none of the participants' ADL motor or ADL process ability measures were significantly different between these two assessment sessions). Following the intervention phase, each participant was reevaluated using the AMPS. Data analyses revealed that the mean group ADL motor ability significantly improved, but ADL process ability remained unchanged. The authors concluded that compensatory strategies that the client's had incorporated into their ADL task performance during the intervention phase resulted in decreased physical effort when completing the ADL tasks. Improvements in ADL task performance were cross-validated from a qualitative perspective through examining the documented goals and progress reports. More specifically, following the intervention phase of the study, 77% of the goals with documented outcomes showed some improvement in occupational performance. These findings provided initial evidence that a client-centered, top-down reasoning and evaluation process results in improved occupational performance.

**Pretest-posttest Randomized Controlled Trial**

The highest level of evidence is produced when researchers employ a randomized controlled trial design, where participants are randomly assigned into a treatment group or a control group. Through randomized assignment, there is greater evidence of group equivalency and fewer threats to the validity of the results (Gliner & Morgan, 2000). Graff and colleagues (2006) conducted a randomized controlled trial to evaluate the efficacy of a community-based occupational therapy program for persons who have ***dementia*** and their caregivers (*n* = 135). Persons in the treatment group received 10 sessions of community-based occupational therapy (over 5 weeks), which incorporated the use of ***compensatory strategies, environmental modifications,***

***and caregiver training.*** Persons in the control group did not receive occupational therapy services. The authors concluded that community-based occupational therapy improved the daily functioning of persons with dementia and diminished the burden of care on their primary caregivers, with such functional improvements remaining stable 7 weeks after the completion of the intervention. More specifically, when compared to the control group, the group of persons who received the occupational therapy services had significantly higher mean ADL process ability measures following the intervention phase; ADL motor ability was not reported. This study provided evidence, that despite diminished cognitive body functions, persons with dementia are capable of increasing the quality of their daily task performance through occupation-based compensatory strategies. Through utilizing the AMPS as one of their outcome measures, the researchers were able to document this important change in occupational performance and the efficacy of occupational therapy services.

In another study, Goverover, Johnston, Tolglia, and Deluca (2007) conducted a randomized control trial study to investigate the effect of ***self-awareness training*** with persons who have an ***acquired brain injury*** ($n$ = 20). Overall, the persons who were in the intervention group and received the self-awareness training had significantly higher mean ADL process ability measures than did those in the control group who had received "traditional" interventions (e.g., corrective feedback). Mean ADL motor ability was not significantly different between the two groups. This study provides one piece of evidence that interventions, aimed at enhancing self-awareness, had a greater impact on ADL task performance than more traditional interventions (e.g., corrective feedback) for persons who have an acquired brain injury.

**The studies summarized above provide evidence that:**

- **The AMPS can be utilized as a sensitive outcome measure in health-related research.**
- **There is a need to design and implement studies that include occupation-based outcome measures (vs. measures of body functions).**
- **The AMPS offers an objective, reliable, valid, and sensitive means of measuring changes in the quality of ADL task performance that can be used to enhance the quality of health-related research.**