Published in final edited form as:

Am J Phys Med Rehabil. 2008 October ; 87(10): 842–852. doi:10.1097/PHM.0b013e318186b7ca.

Adaptive Short Forms for Outpatient Rehabilitation Outcome Assessment

Alan M. Jette, Ph.D., MPH, P.T., FAPTA¹ [Director], Stephen M. Haley, Ph.D., P.T.² [Associate Director], Pengsheng Ni, MD, MPH.³ [Research Assistant Professor], and Richard Moed, MPA⁴ [President]

¹Health and Disability Research Institute, Boston University School of Public Health

²Health and Disability Research Institute, Boston University School of Public Health

³Health Policy and Management, Boston University School of Public Health

⁴CRE Care, LLC, Boston, MA

Abstract

Objective—To develop outpatient adaptive short forms (ASFs) for the Activity Measure for Post-Acute Care (AM-PAC) item bank for use in outpatient therapy settings.

Design—A convenience sample of 11,809 adults with spine, lower extremity, upper extremity and miscellaneous orthopedic impairments who received outpatient rehabilitation in one of 127 outpatient rehabilitation clinics in the US. We identified optimal items for use in developing outpatient ASFs based on the *Basic Mobility* and *Daily Activities* domains of the AM-PAC item bank. Patient scores were derived from the AM-PAC computerized adaptive testing (CAT) program. Items were selected for inclusion on the ASFs based on functional content, range of item coverage, measurement precision, item exposure rate, and data collection burden.

Results—Two outpatient ASFs were developed: 1) an 18-item Basic Mobility ASF and 2) a 15-item Daily Activities ASF, derived from the same item bank used to develop the AM-PAC-CAT. Both ASFs achieved acceptable psychometric properties.

Conclusions—In outpatient PAC settings where CAT outcome applications are currently not feasible, IRT-derived ASFs provide the efficient capability to monitor patients' functional outcomes. The development of ASF functional outcome instruments linked by a common, calibrated item bank has the potential to create a bridge to outcome monitoring across PAC settings and can facilitate the eventual transformation from ASFs to CAT applications easier and more acceptable to the rehabilitation community.

Keywords

Outcomes Assessment; Rehabilitation; Item Response Theory; Physical Functioning

Correspondence: Alan Jette, Health and Disability Research Institute, Boston University School of Public Health, Boston University Medical Center, 580 Harrison Avenue, 4th Floor, Boston, MA 02118-2639; ajette@bu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

INTRODUCTION

The Centers for Medicare and Medicaid Services (CMS) has acknowledged that steps need to be taken toward developing a post-acute care (PAC) system in the United States that provides payment and assures quality of care for an overall episode of PAC, rather than for each individual component within the continuum of care.^{1,2} As an essential step toward accomplishing this policy objective, assessment methods are needed to collect and compare relevant outcomes across various sites where PAC is provided. In this paper we will focus on functional outcomes, since they have been recommended for inclusion in PAC outcome monitoring systems.³ We define functional outcomes very broadly to include an individual's ability to carry out basic and instrumental activities of daily living (ADLs and IADLs) as well as their participation in advanced functional activities.

There are numerous well-respected, traditional functional outcome instruments in widespread use within PAC settings.⁴⁻⁸ Most traditional functional outcome instruments used today were developed using classical test theory methods, which employ a majority of items in the middle range of the functional scale due to the limitations of the number of items that are practical to administer in a busy clinical setting.⁹ All patients are expected to answer every item, often presenting to patients a series of questions that are either too easy, too difficult, irrelevant, or redundant.¹⁰ Traditional functional outcome instruments lack measurement precision at some levels of the outcome being measured, creating ceiling and floor effects when one instrument is used across inpatient as well as outpatient care settings.^{9, 11-13} Since distinct functional outcome instruments have been developed for each PAC setting and, because each instrument uses a different set of items, rating scales and scores, and different administration and scoring rules, cross instrument data incompatibility renders the ability to track relevant outcomes across different instruments difficult to accomplish with precision.^{14, 15}

Contemporary measurement methods such as Item Response Theory (IRT) and Computerized Adaptive Testing (CAT) methods provide a promising means to achieve eventually the long-term goal of longitudinal monitoring of functional status (as well as other relevant outcomes) across an episode of PAC by developing outcome instruments that are psychometrically adequate, comprehensive and precise to monitor function across a wide range of ability while being practical for widespread application in clinical settings.¹⁶⁻²¹ However, CAT methods require providers who conduct each functional assessment to have some type of point-of-contact computing power which is currently beyond the capability of some care settings. An attractive interim strategy, in such cases where CAT technology is not feasible, might be to develop and implement Adaptive Short Form instruments (ASFs) for different settings that are linked by a common underlying metric.²²

A key characteristic of ASFs is that the functional items incorporated into a setting specific instrument are drawn from the same underlying item bank and are selected using IRT methods. This approach provides the basis for linking different setting specific assessment instruments together along a common scale for an outcome dimension, in this case, functional outcomes. ASFs allow scores derived from each instrument to be compared, avoiding the problems of combining scores from two instruments that have not been co-calibrated on the same metric using IRT methods.^{19, 23} An item bank is a collection of items that represent a range of performance or difficulty levels for a particular outcome dimension²⁴⁻²⁶ and are developed by equating outcome items from different sources so that they can be meaningfully linked together on a common underlying metric. Once the structure and ordering of items is determined, items can then be selected to create a setting specific ASF based on a number of criteria, including, comprehensiveness of content,

functional ability of the patient group, item precision, and practical considerations such as respondent burden and administration cost.

In this paper, we describe the use of calibrated item banking to develop two ASF instruments created specifically for the outpatient rehabilitation setting. To accomplish this goal, we utilized two physical functioning item banks that have been previously shown to be distinct and unidimensional in PAC patients and recently expanded for outpatient rehabilitation applications, which we collectively refer to as the Activity Measure for Post-acute Care (AM-82 PAC).^{27–29} In this analysis we used four criteria for selecting items from the calibrated AM-PAC item banks for the outpatient ASFs: (1) content considerations and item exposure rate, (2) range of patients' ability on the AM-PAC content domains of **Basic Mobility** and **Daily Activities**, (3) measurement precision using test information functions, and practical considerations of length. Below, we describe the content, psychometric properties, and floor and ceiling effects of the ASFs, as well as estimates of how the ASFs items match person abilities from individuals in an adult outpatient rehabilitation practice.

METHODS

AM-PAC Item Bank

The AM-PAC item banks were initially developed from a sample of 1,041 patients with an age of 18-years or greater who were receiving post-acute care services in inpatient rehabilitation, ambulatory care centers, skilled nursing facilities and home care agencies. Details of the full sampling plan have been published elsewhere.³⁰ The original post-acute care sample included three major patient groups: 1) 33.2% neurological (e.g., stroke, multiple sclerosis, Parkinson's disease, brain injury, spinal cord injury, neuropathy); 2) 28.4% musculoskeletal (e.g., fractures, joint replacements, orthopedic surgery, joint or muscular pain); and 3) 38.4% medically complex (e.g., debility resulting from illness, cardiopulmonary conditions, or post-surgical recovery).

A core set of 124 Basic Mobility items and 65 Daily Activity items were chosen for inclusion in the original AM-PAC item bank based on a comprehensive review of items from existing instruments, the ICF framework, a review by ten measurement and content experts, and suggestions solicited from several focus groups of individuals with disabilities. Based on calibration work of the post-acute care sample (N= 1,041) items were checked for fit to the model and any items with significant differential item functioning were removed.³⁰ Items were phrased, "How much difficulty do you currently have (without help from another person or device) with the following activities...?" A polytomous response choice included "none," "a little," "a lot," and "cannot do." We framed the activity questions in a general fashion without specific attribution to health, medical conditions, or disabling factors. The original AM-PAC item pool was expanded by adding and calibrating 7 Basic Mobility and 23 Daily Activity items to the original item bank from data collected on a calibration sample of 11,809 outpatient rehabilitation patients who were receiving outpatient rehabilitation services in clinics operated by Select Physical Therapy, a division of clinics owned and operated by Select Medical Corporation.³¹

In order to link the new items to the existing item banks, we used an online calibration method for the development of the item calibrations and an item fit procedure using a Pearson chi-square test based on the posterior distribution.^{32–34} These methods are more fully explained in Haley et al.³⁵

Subjects

We developed ‘patient scores’ from a convenience sample of 11,809 persons who completed the AM-PAC-CAT on admission and discharge (23,618 scores) while receiving services in an ambulatory outpatient rehabilitation practice. These data came from routine administration of the AM-PAC-CAT of patients receiving outpatient physical therapy services in 127 outpatient clinics across 12 states that were operated by Select Physical Therapy. See Table 1 for a summary of the outpatient sample characteristics.

Data Collection Procedures—Subjects (N=11,809) completed the self-report AM-PAC-CAT on a tablet computer provided to them in the waiting room prior to their therapy visit. Each person in the sample completed an admission and discharge AM-PAC-CAT, so the number of AM-PAC-CAT administrations was 23,618. Computer-adaptive testing (CAT) employs an algorithm [12] that selects items from the AM-PAC item banks directly tailored to the person, and shortens or lengthens the test to achieve either the desired precision or a pre-assigned item stopping rule. We set the stop-rule at 7-items to best meet the needs of the clinic operation. In a CAT program, score estimates are affected by an optimal choice of test items that are selected based on proximity to the person’s functional level.

An office staff member was available to the subjects during the administration process to answer any questions. All 11,809 subjects included in this analysis completed both admission and discharge AM-PAC-CATs. Subject demographic information, surgical status, and major impairment were all available from administrative data collected routinely by each outpatient clinic, and combined with the AM-PAC-CAT scores. Spine impairments included impairments of the cervical, thoracic, or lumbosacral region of the spine. Upper-extremity impairments included conditions of the shoulder, elbow, hand, or wrist. Lower-extremity impairments were conditions of the hip, knee, foot, or ankle. Other conditions included neurological, medical and unspecified major impairments.

Analyses

To determine the optimal items for the Basic Mobility and Daily Activity outpatient ASFs, we considered 4 criteria: *functional content*; *range of coverage* of the functional items, including floor and ceiling effects; *measurement precision*; and when available, *item exposure rates* from the AM-PAC-CAT.

Functional content—Content considerations were used to make sure that there were sufficiently varied items in the ASFs, such that the ASFs were not dominated by a single type of functional item (such as walking skills).

Range of coverage—Content range coverage assessed how well the AM-PAC item bank captured the range of physical functioning experienced by these outpatients in each of the AM-PAC physical functioning domains that were obtained from CAT-based scores. Content range was described in the common metric of the AM-PAC-CAT and ASF scores, which have a mean of 50 and SD of 10. To gain an estimate of content coverage of each short form, we compared the range from the lowest to the highest item-response estimates of the short forms to the complete item pools. The item-response estimates are based on the expected value at each item-response point under the latent scale. The expected value was calculated by the sum of the category values times the corresponding probabilities. For example, for the category 2, the range of this category under the latent scale will be the value corresponding to the expected value between 1.5 to 2.5.³⁶ We also examined potential ceiling effects (i.e., the point at which subjects received the highest score) and floor effects (i.e., the point at which subjects received the lowest possible score).

Measurement precision—Item information functions are related to the location and shape of the item characteristic curve, which describes the probabilities of responding to particular response options of an item.³⁷ Using item information functions, we identified whether a given item was precise at any level of the activity domain, as higher item information means more precision. Item information values were continuous over the ability scale and therefore summed and displayed graphically to form *test* information functions (TIFs).³⁸ In order to determine the best item set for each ASFs using the TIFs, we added and removed items iteratively so that a subset of items was identified that most closely fit the levels of precision needed to match the person ability scores. The location on the scale where the test information curve peaked indicated the portion of the scale best measured by the scale for that sample. When the test information was peaked at or around the same range on the scale as the patients' peak of ability distribution, the instrument was assumed to "fit" the population being measured. We developed a final TIF for the 18-item Basic Mobility ASF and the 15-item Daily Activities ASF that best fit the patient scores, taking into account the other criteria as mentioned above.

Item exposure—The item exposure rate (IER),³⁹ identified which AM-PAC items were administered more often in the CAT application. IER was defined as the ratio of the total number of times an item was administered over the total number of test occasions in a CAT study. The IER is influenced by the difficulty and discrimination of items, the distribution of ability of the patients, what other similar items are in the item bank. Items with the best discrimination and item information typically have high item exposure rates. For some of the newly calibrated items, IER was not available.

RESULTS

Basic Mobility Adaptive Short Form

Because of the wide range of content needed to adequately measure basic mobility in elderly and non-elderly adults, the Basic Mobility ASF included 18 items, which we felt represented a reasonable balance between quality of measurement and data collection burden in the clinic. Content included transfers, walking skills, bending and carrying, and running, and strenuous activities. (Table 2) The full range of the 18-items was between 16.18 and 95.78 (mean 50; SD 10) compared with the Basic Mobility item bank range which was 0 to 95.78. Two percent of this sample was at the ceiling of the Basic Mobility scale at discharge from their episode of care with no members of the sample at the floor of the scale. We selected items for this form so that items matched as closely as possible where persons were scoring. The TIF plot peaked in the scoring range where most people were located. As noted in Figure 1, ± 2 SD of the outpatient sample scores on Basic Mobility corresponded to 92.8% of the area under TIF. Only 381 patients (3% of the sample) had misfitting data on this ASF.

Daily Activity Adaptive Short Form

The Daily Activity ASF included 15 items. (Table 3) Content included fine manipulation skills, relatively difficulty ADLs such as tying shoes and cutting toenails, household IADLs, and lifting heavy objects. The full range of the 15-items is between 27.35 and 100 (Daily Activity item bank range is 23.26 to 100). Six percent of this sample was at the ceiling of the Daily Activity scale at discharge from their episode of care while no members of the sample were at the floor of the scale. The TIF plot peaks at about the scoring range where most people are scoring. (Figure 2) As noted in Figure 2, ± 2 SD of the outpatient sample scores on Daily Activity corresponds to 82.1% of the area under TIF. For example, in Figure 2, note that the peaks of the information function curves are approximately at the midpoint of the sample score distributions for Basic Mobility. In the Daily Activity scale, only 270 patients (2% of the sample) had misfitting data.

The Basic Mobility and Daily Activity domains, while distinct, were correlated in this outpatient sample at 0.47 at admission and 0.59 at discharge from their episode of care. Both ASFs are provided in the appendix to this manuscript.

DISCUSSION

An essential step toward accomplishing the policy objective of developing a PAC system in the United States that assures quality for an overall episode of PAC, rather than for each individual component within the continuum of care, are assessment systems that can be used to monitor relevant outcomes achieved across various sites where PAC is provided. A persistent barrier to fulfilling this PAC policy mandate has been the inability to achieve a standardized, functional outcome assessment approach that can provide appropriate information on outcomes and quality of care that can be applied over time, and across different settings where PAC services are provided.¹ Unfortunately, traditional functional outcome instruments used within PAC in the United States were developed independently for each PAC setting and, consequently, yield distinct and setting specific metrics that are not easily compared across settings.^{14,40,41}

In this outpatient sample, we found that the 18-item Basic Mobility ASF had generally good psychometric properties, however, we did include 3 items (bed mobility, chair transfer, and stand from a low surface) that had relatively poor fit since those items were either felt to have essential content or represent a critical placement along the continuum. Similarly, we included 4 items (wash windows indoors, move a sofa to clean under, lift 25 pounds, lift 50 pounds) in the Daily Activity ASF that had some misfit problems, but the items matched the sample scores, and clinically, there were important content items to include in an overall outpatient assessment.

The results from this study demonstrate that general IRT methods can be an effective methodology for building ASFs tailored to one care setting (in this case the outpatient rehabilitation setting) that when linked to other ASFs developed from the same calibrated item bank can be used to track functional outcomes across other PAC settings. These outpatient ASFs build on prior work we have done where different ASFs designed for institutionalized patients and community-based patients were linked to track functional progress throughout an episode of PAC.¹³ Using ASFs designed for different settings increases measurement efficiency and effectiveness compared with a single traditional measurement instrument applied to everyone. Respondents need only respond to one form (a subset of the functional items that best targets their level of function based on the setting in which they are receiving PAC) yet scores derived from an ASF can be linked to scores derived from other outcome instruments developed from the same underlying item bank and applied in different care settings.¹³

Ultimately, we believe that CAT technology will provide the best means to develop outcome measures that are practical and psychometrically adequate for use across PAC settings. However, CAT methods require providers who conduct each functional assessment to have some type of point-of-contact computing power which is currently beyond the capability of some care settings. As constructed, ASFs provide an initial step toward improving measurement across PAC settings while not increasing response burden in settings where CAT outcome applications are neither feasible nor likely to be accepted for some time. IRT-derived ASFs, such as those developed in this study, make it possible to monitor patient functional outcomes across care settings using different ASFs developed from the same functional item bank.

Similar item banking and IRT methods can be used to develop CAT applications.^{14, 42} CAT methodology uses a computer interface for the person (or a computerized interview clinician report) that is tailored to the unique ability level of that person. CAT applications require: (1) a large set of items in any one functional area (item pools), (2) items that consistently scale along a dimension of low to high functional proficiency, and (3) rules guiding starting, stopping and scoring procedures. The CAT application can achieve good efficiency *without* the loss in individual score precision that is seen with the ASFs.

Although there are many methods and analytic techniques available to develop short form version of outcome instruments, such as linear regression, factor analysis, inter-item correlations, item total correlations, comparing items to external criteria and using patients' ratings of importance, we employed a strategy of linking items together in a common item bank and using IRT analyses to select the most relevant items for each care setting. Unlike traditional functional assessment short forms like the Functional Independence Measure, ASFs derived from a common item bank have the distinct and critical advantage of providing a common metric that can be employed to track how a patient's function changes across an entire episode of care. In this way, it is conceivable for clinical settings to use different formats of the AM-PAC (ASF or CAT) and be able to compare patient scores within and across settings. This provides maximum flexibility when designing and implementing a patient monitoring system that can be applied across an entire episode of care.

We found the approach of matching Test Information Functions (TIFs) to ability of the target population in each setting to be very useful in selecting items to include in each outpatient ASF. We performed this process for the entire sample and for the sub-sample of adults aged 65+ to be sure the ASFs were adequate for use with older patients as well as with non-elderly members of the sample. This process allowed us to identify items by matching item difficulty with person scores as well as by examining precision provided by a set of selected items. Although item difficulty and precision are related criteria, they are not always the same, particularly with polytomous items such as those used in the AM-PAC. Our ability to match test information function and ability level for each target group in the sample was due to the large number of functional items in the AM-PAC item bank.

In earlier work, we used this same analytic approach in building ASFs designed for both inpatient and community assessment of function for a wide range of post-acute care patients.⁴³ We were successful in developing ASFs that provided a broad range of coverage for institutional and community-based care settings with very little change in ceiling and floor effects going from the full item banks to the ASFs. The ASFs were subsequently used successfully in a number of long-term outcome studies to track the functional progress of patients over time and to understand predictors of functional recovery.⁴⁴⁻⁴⁷

There are several limitations to the study that should be noted. Although one advantage of this approach is that regardless of the items that are selected for a particular ASF, each ASF is scored on a similar metric, allowing for scoring comparability across ASFs within the entire item bank. One decided disadvantage of the ASFs over using a CAT or an entire item bank is some loss of precision around any one individual score. The precision loss between the ASFs and a CAT depends upon the number of items in the ASFs and in the CAT, and the location of a person on the functional continuum. For example, in prior work,²⁸ we found nearly a 2-fold increase in standard error in the Basic Mobility and Daily Activity scales when comparing a 10-item ASF to a 10-item CAT. The loss of precision of the ASFs at the extreme score ranges is even greater. In data simulations on a subset of this sample, we found that the average CAT SEs were always less than in comparable length short forms by at least a factor of 1.3. Nonetheless, the ASF approach provides a useful compromise

between content comprehensiveness, precision and feasibility for implementation in care settings.

The reader should also understand that the items in the ASFs developed in this study and their calibrations are optimized for a sample of outpatient rehabilitation patients and may not be appropriate for other populations and settings. Future research needs to be conducted to develop and evaluate additional ASFs from the same AM-PAC item banks for other patient populations and settings.

It is possible that some of the item calibrations may not be generalizable over time even though this was not the case in this study. For those items with at least 100 subjects at both admission and discharge, we did test the differential item functioning (DIF) based on logistic regression modeling of the item response probability based on different independent variables. There were 3 possible models, the first included only the functional ability score; the second modeled the functional ability score and the time variable (admission to discharge); the third included functional ability, the time variable and the interaction term between each other.. The R-square change between the third model and first model was <0.035 indicating no item showed time DIF in either AM-PAC ASF scale in this sample.

It is also appropriate to point out that the AM-PAC and the derivative ASFs were developed as composite outcome indicators in two specific functional domains: Basic Mobility and Daily Activity. Their use to assess patient progress in these domains of function are not meant to replace clinical treatment planning that may target a single functional skill or a more limited or different set of functional skills. These ASFs assess two carefully delineated dimensions of physical function believed to be of importance in monitoring PAC outcomes.

In summary, this analysis illustrated that two psychometrically adequate outpatient rehabilitation ASFs could be developed from the AM-PAC item bank: an 18-item Basic Mobility ASF and a 15-item Daily Activities ASF which yielded an outcome metric that is common to the one provided by the AM-PAC-CAT. In PAC settings where CAT outcome applications are neither feasible nor likely to be accepted at the present time, ASFs make it possible to track patient functional outcomes with an instrument that can be linked to other instruments developed from the same item bank. The future linkage of different ASFs developed for different PAC settings from the same item banks will serve as a transitional step to facilitate the eventual transformation to episode wide methods for assessing outcomes such as functional ability.

Acknowledgments

Disclosures:

Supported in part by the National Institute on Disability and Rehabilitation Research (Rehabilitation Research and Training Center on Measuring Outcomes (H133B990005) and R01 HD43568 from the National Institute of Child Health and Human Development (NICHD) and the Agency for Healthcare Research and Quality (AHRQ). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIDRR, NICHD, or AHRQ. This manuscript was also supported in part by an Independent Scientist Award (K02 HD45354-01) to Dr. Haley. Drs. Haley and Jette and Mr. Moed have stock interest in CRE Care, LLC which distributes the Activity Measure for Post-Acute Care (AM-PAC) Products. Select Medical Corporation purchased the Outpatient Rehabilitation Division of HealthSouth Corporation on May 1, 2007 and the individual clinics that participated in this study are now known as “*Select Physical Therapy*.” We would like to thank all of the Select Physical Therapy clinical sites who participated in our study by providing the data used to develop the ASFs.

REFERENCES

1. Johnson, M.; Holthaus, D.; Harvell, J.; Coleman, E.; Eilertsen, T.; Kramer, A. A Medicare post-acute care quality measurement: final report. In: Report prepared for the Office of the Asst Sec for Planning and Eval. , editor. University of Colorado Health Sciences Center; 2001.
2. US House Ways and Means Subcommittee on Health, Hearing on Post Acute Care; 2005 Jun 16. Testimony of Herb Kuhn, Director, Center for Medicare Management, Centers for Medicare Medicaid Services.
3. National Committee on Vital and Health Statistics (US). Classifying and Reporting Functional Status: National Committee on Vital and Health Statistics. 2002
4. Harvey R, Jellinek HM. Functional performance assessment: a program approach. Arch Phys Med Rehabil. 1981; 62:456–461. [PubMed: 7283689]
5. Shaughnessy, P.; Crisler, KS.; Schlenker, RE. Center for Health Services and Policy Research. Denver, CO: 1997. Medicare's OASIS: Standardized Outcome and Assessment Information Set for Home Health Care - OASIS B.
6. Ware J, Sherbourne C. The MOS 36-item short form health survey (SF-36): conceptual framework and item selection. Med Care. 1992; 30:473–483. [PubMed: 1593914]
7. Guide for the Uniform Data Set for Medical Rehabilitation (including the FIMTM instrument), Version 5.1. Buffalo, NY: State University of New York at Buffalo; 1997.
8. Morris, JN.; Murphy, K.; Nonemaker, S. Version 2.0. Washington: American Health Care Association; 1995. Long-term Resident Care Assessment User's Manual.
9. Jette AM, Haley SM, Ni P. Comparison of functional status tools used in post-acute care. Health Care Financ Rev. 2003; 24:13–24. [PubMed: 12894632]
10. McHorney C. Health status assessment methods for adults: Past accomplishments and future challenges. Annu Rev Public Health. 1999; 20:309–335. [PubMed: 10352861]
11. McHorney C. Generic health measurement: past accomplishments and a measurement paradigm for the 21st century. Annals of Internal Med. 1997; 127(8):743–750.
12. McHorney C, Cohen A. Equating health status measures with Item Response Theory: illustrations with functional status items. Med Care. 2000; 38:II43–II59. [PubMed: 10982089]
13. Coster W, Haley S, Jette A. Measuring patient-reported outcomes after discharge from inpatient rehabilitation settings. J Rehabil Med. 2006; 38:237–242. 2006. [PubMed: 16801206]
14. Buchanan JL, Andres PL, Haley SM, Paddock S, Young DC, Zaslavsky A. An assessment tool translation study. Health Care Financ Rev. 2003; 24:45–60. [PubMed: 12894634]
15. Jette AM, Haley SM. Achieving Uniformity in Functional Status Monitoring across PAC Settings. Centers for Medicare and Medicaid. 2006
16. Weiss D. Improving measurement quality and efficiency with adaptive testing. Applied Psychological Measurement. 1982; 6:473–492.
17. Velozo CA, Kielhofner G, Lai JS. The use of Rasch analysis to produce scale-free measurement of functional ability. Am J Occup Ther. 1999; 53:83–90. [PubMed: 9926224]
18. Wainer, H. Computerized Adaptive Testing: A Primer. New Jersey: Lawrence Erlbaum Associates; 2000.
19. Hays R, Morales L, Reise S. Item response theory and health outcomes measurement in the 21st century. Med Care. 2000; 38:II-28–II-42. [PubMed: 10982088]
20. Cella D, Chang C-H. A discussion of Item Response Theory and its applications in health status assessment. Med Care. 2000; 38:II66–II72. [PubMed: 10982091]
21. Hambleton RK. Emergence of item response modeling in instrument development and data analysis. Med Care. 2000; 38:II60–II65. [PubMed: 10982090]
22. Cella D, Gershon R, Lai JS, Choi S. The future of outcome measurement: item banking, tailored short-forms, and computerized adaptive assessment. Qual Life Res. 2007 (published on-line March 13).
23. Hambleton, RK. Applications of Item Response Theory to Improve Health Outcomes Assessment: Developing Item Banks, Linking Instruments, and Computer-Adaptive Testing. In: Lipscomb, J.;

- Gotay, CC.; Snyder, C., editors. Outcomes Assessment in Cancer. Cambridge, UK: Cambridge University Press; 2005.
24. Hambleton, R.; Swainathan, H. Item Banking. In: Hambleton, R.; Swainathan, H., editors. Item Response Theory: Principles and Applications. Boston, MA: Kluwer Nijhoff Publishing; 1985.
 25. Revicki DA, Cella DF. Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. *QualLife Res.* 1997; 6:595–600.
 26. Bode RK, Lai JS, Cella D, Heinemann AW. Issues in the development of an item bank. *Arch Phys Med Rehabil.* 2003; 84:S52–S60. [PubMed: 12692772]
 27. Haley SM, Coster WJ, Andres PL, Ludlow LH, Bond T, Sinclair SJ, Jette AM. Activity Outcome Measurement for Post-acute Care. *Medical Care.* 2004; 42(1 Suppl):I-49–I-69. [PubMed: 14707755]
 28. Haley SM, Andres PL, Coster WJ, Kosniski M, Ni PS, Jette A. Short-form activity measure for post-acute care (AM-PAC). *Arch Phys Med Rehabil.* 2004; 85:649–660. [PubMed: 15083443]
 29. Coster WJ, Haley SM, Andres PL, Ludlow LH, Bond T. Refining the conceptual basis for rehabilitation outcome measurement: personal care and instrumental activities domain. *Medical Care.* 2004; 42(Suppl. 1):I-62–I-72. [PubMed: 14707756]
 30. Haley SM, Ni P, Hambleton RK, Slavin MD, Jette AM. Computer adaptive testing improves accuracy and precision of scores over random item selection in a physical functioning item bank. *J Clin Epidemiol.* 2006; 59:1174–1182. [PubMed: 17027428]
 31. Jette A, Haley S, Tao W, Ni P, Meyers D, Zurek M. Prospective evaluation of the AM-PAC-CAT in outpatient rehabilitation settings. *Phys Ther.* 2007; 87:385–398. [PubMed: 17311888]
 32. Stone CA. Monte-Carlo based null distribution for an alternative fit statistic. *Journal of Educational Measurement.* 2000; 37:58–75.
 33. Stone CA. Empirical power and Type I error rates for a goodness-of-fit statistic based on posterior expectations and resampling-based inference. *Educational and Psychological Measurement.* 2003; 63:566–583.
 34. Stone CA, Zhang B. Comparing three new approaches for assessing goodness-of-fit in IRT models. *Journal of Educational Measurement.* 2003; 4:331–352.
 35. Haley SM, et al. Replenishing a Computerized Adaptive Test (CAT) of Patient Reported Outcomes. Under review. 2008
 36. van der Linden, WJ.; Hambleton, RK. Handbook of modern item response theory. New York: Springer. Mcleod. LD; 1997.
 37. Murnki E. Information functions of the generalized partial credit model. *Applied Psychological Measurement.* 1993; 17:351–363.
 38. Dodd BG, Koch WR. Effects of variations in item step values on item and test information in the partial credit model. *Appl Psychol Meas.* 1987; 11:371–384.
 39. Revuelta J, Ximenez MC, Olea J. Psychometric and psychological effects of item selection and review on computerized testing. *Educational and Psychological Measurement.* 2003; 63:791–808.
 40. Fisher W. Physical disability construct convergence across instruments: towards a universal metric. *J Outcome Meas.* 1997; 1:87–113. [PubMed: 9661716]
 41. Smith R, Taylor P. Equating rehabilitation outcome scales: developing common metrics. *J Appl Meas.* 2004; 5:229–242. [PubMed: 15243171]
 42. Siebens H, Andres P, Ni P, Coster W, Haley S. Measuring physical function in patients with complex medical and postsurgical conditions: a computer adaptive approach. *Am J Phys Med Rehabil.* 2005; 84:741–748. [PubMed: 16205429]
 43. Haley SM, Coster WJ, Andres PL, Kosinski M, Ni PS. Score comparability of short-forms and computerized adaptive testing: simulation study with the Activity Measure for Post-Acute Care (AM-PAC). *Arch Phys Med Rehabil.* 2004; 85:661–666. [PubMed: 15083444]
 44. Jette AM, Keysor J, Coster W, Ni P, Haley SM. Beyond function: Predicting participation in a rehabilitation cohort. *Arch Phys Med Rehabil.* 2005; 86:2087–2094.
 45. Keysor J, Jette A, Coster W, Bettger J, Haley S. Association of Environmental Factors With Levels of Home and Community Participation in an Adult Rehabilitation Cohort. *Arch Phys Med Rehabil.* 2006; 87:1566. [PubMed: 17141635]

46. Coster WJ, Haley SM, Jette AM. Measuring patient-reported outcomes after discharge from inpatient rehabilitation settings. *J Rehabil Med.* 2006; 38:237–242. [PubMed: 16801206]
47. Coster W, Haley SM, Jette A, Tao W, Siebens H. Predictors of basic and instrumental activities of daily living performance in persons receiving rehabilitation services. *Arch Phys Med Rehabil.* 2007; 88(7):928–935. [PubMed: 17601476]

Figure 1.
Test Information Function for the Basic Mobility ASF

Figure 2.
Test Information Function for the Daily Activity ASF

Table 1

Characteristics of the Study Sample

Sample size	11809
Mean Age	50(18.21)
Gender (Female %)	59.12%
% Post Surgery	29.31%
Primary Impairment	
SPINE	29.85%
UPPER	22.57%
LOWER	25.75%
OTHER	21.83%

Table 2

Content, Range of Coverage, and Item Exposure of the Basic Mobility ASF

Basic Mobility SF Item Content	Mean Calibration	Range of Expected Values	Model Fit(chi-square, p)	% Exposure Rate in CAT
1. Bed mobility	40.36	16.18~65.93	22.58, 0.09	0.03%
2. Car transfer	49.21	34.09~67.92	3.98, 0.784	0.04%
3. Clean up floor spills	49.41	30.11~69.91	10.93, 0.617	_*
4. Chair transfer	50.61	40.06~61.95	26.23, 0.051	22%
5. Walk in home	51.9	46.03~59.96	11.23, 0.884	19%
6. Light housework	53.19	48.02~59.96	9.61, 0.567	9%
7. Bending	54.69	44.04~67.92	23.78, 0.416	100%
8. Stand from low surface	56.18	42.05~71.9	24.78, 0.074	77%
9. Walk on an uneven surface	56.78	50.01~65.93	11.18, 0.799	16%
10. Walk several blocks	57.37	52~63.94	13.33, 0.714	46%
11. Flight of stairs	60.26	53.99~67.92	11.22, 0.796	29%
12. Walking up and down inclines	60.56	53.99~67.92	10.36, 0.736	40%
13. Walk one mile	62.95	55.98~69.91	2.04, 1	65%
14. Carrying something while climbing stairs	63.14	55.98~71.9	11.12, 0.802	24%
15. Run a short distance	66.03	59.96~69.91	16.94, 0.322	32%
16. Making sharp turns when running fast	68.42	61.95~75.88	5.85, 0.884	19%
17. Run for 5 minutes on even surfaces	70.61	63.94~77.87	9.66, 0.562	30%
18. Strenuous activities	78.77	63.94~95.78	13.00, 0.293	13%

* New item, not in CAT item pool

Table 3

Content, Range of Coverage, and Item Exposure of the Daily Activity ASF

Daily Activity SF Item Content	Mean Calibration	Range of Expected Values	Model Fit (chi-square(df), p)	% Exposure Rate in CAT
1. Tie shoes	45.38	27.38~62.36	20.60, 0.15	100%
2. Sew a button	49.39	39.72~58.24	11.01, 0.61	74%
3. Pound a nail with hammer	48.98	37.66~62.36	12.11, 0.356	65%
4. Open jar	47.34	31.49~64.41	10.49, 0.84	61%
5. Replace or tighten small parts	46.41	37.66~54.13	5.03, 0.93	30%
6. Remove packaging	44.66	33.55~56.18	11.12, 0.519	13%
7. Cut toenails	57.62	27.38~87.05	17.37, 0.43	18%
8. Hang wash on a clothes line	48.36	29.43~68.53	27.31, 0.27	—*
9. Wash indoor windows	48.26	29.43~68.53	38, 0.06	—*
10. Move a sofa to clean under	57.31	29.43~84.99	31.68, 0.09	—*
11. Use a manual screwdriver	47.03	39.72~54.13	13.85, 0.733	44%
12. Lifting 25 pounds	57.01	27.38~86.02	57.06, 0.05	—*
13. Lifting 100 pounds	72.23	37.66~100	57.04, 0.02	—*
14. Five push-ups	64.72	37.66~64.41	52.14, 0.14	—*
15. Manage clothing behind back	51.14	31.49~72.64	26.95, 0.53	—*

* New item, not in CAT item pool