



Published in final edited form as:

*Mov Disord.* 2016 January ; 31(1): 126–134. doi:10.1002/mds.26339.

## Development and Initial Testing of the Penn Parkinson's Daily Activities Questionnaire

Laura Brennan, PhD<sup>1,2</sup>, Andrew Siderowf, MD, MSCE<sup>3</sup>, Jonathan D. Rubright, PhD<sup>4</sup>, Jacqueline Rick, PhD<sup>5</sup>, Nabila Dahodwala, MD<sup>5</sup>, John E. Duda, MD<sup>2,5</sup>, Howard Hurtig, MD<sup>5</sup>, Matthew Stern, MD<sup>5</sup>, Sharon X. Xie, PhD<sup>6</sup>, Lior Rennert, MS<sup>6</sup>, Jason Karlawish, MD<sup>7</sup>, Judy A. Shea, PhD<sup>8</sup>, John Q. Trojanowski, MD, PhD<sup>5</sup>, and Daniel Weintraub, MD<sup>1,2,5,\*</sup>

<sup>1</sup>Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA

<sup>2</sup>Philadelphia Veterans Affairs Medical Center, Philadelphia, Pennsylvania, USA

<sup>3</sup>Avid Radiopharmaceuticals, Philadelphia, Pennsylvania, USA

<sup>4</sup>American Institute of Certified Public Accountants, Ewing, New Jersey, USA

<sup>5</sup>Department of Neurology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA

<sup>6</sup>Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA

<sup>7</sup>Departments of Medical Ethics and Medicine, PENN-CMU Roybal Center on Behavioral Economics and Health, the LDI Center for Health Incentives, and Alzheimer's Disease Center

<sup>8</sup>Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA

### Abstract

**Objective**—The aim of this work was to describe the development and psychometric analysis of the Penn Parkinson's Daily Activities Questionnaire. The questionnaire is an item response theory-based tool for rating cognitive instrumental activities of daily living in PD.

**Methods**—Candidate items for the Penn Parkinson's Daily Activities Questionnaire were developed through literature review and focus groups of patients and knowledgeable informants. Item selection and calibration of item-response theory parameters were performed using responses from a cohort of PD patients and knowledgeable informants (n = 388). In independent cohorts of PD patients and knowledgeable informants, assessments of test-retest reliability (n = 50), and construct validity (n = 68) of the questionnaire were subsequently performed. Construct validity

\*Correspondence to: Dr. Daniel Weintraub, MD, Department of Psychiatry, University of Pennsylvania School of Medicine, 3615 Chestnut Street, #330, Philadelphia, PA 19104, USA; daniel.weintraub@uphs.upenn.edu.  
Dr. Rubright completed this work while at University of Delaware School of Education, Newark, Delaware, USA.

Relevant conflicts of interest/financial disclosures: Nothing to report.

Full financial disclosures and author roles may be found in the online version of this article.

Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

was assessed by correlating questionnaire scores with measures of motor function, cognition, an existing activities of daily living measure, and directly observed daily function.

**Results**—Fifty items were retained in the final questionnaire item bank. Items were excluded owing to redundancy, difficult reading level, and when item-response theory parameters could not be calculated. Test-retest reliability was high (intraclass correlation coefficient = 0.97;  $P < 0.001$ ). The questionnaire correlated strongly with cognition ( $r = 0.68$ ;  $P < 0.001$ ) and directly observed daily function ( $r = 0.87$ ;  $P < 0.001$ ), but not with motor impairment ( $r = 0.08$ ;  $P = 0.53$ ). The questionnaire score accurately discriminated between PD patients with and without dementia (receiver operating characteristic curve = 0.91; 95% confidence interval: 0.85–0.97).

**Conclusions**—The Penn Parkinson's Daily Activities Questionnaire shows strong evidence of reliability and validity. Item response theory-based psychometric analysis suggests that this questionnaire can discriminate across a range of daily functions.

### Keywords

Parkinson's disease; instrumental activities of daily living; cognition; item response theory

---

Cognitive impairment in Parkinson's disease (PD) and PD dementia (PDD) is a major source of disability,<sup>1,2</sup> caregiver burden,<sup>3</sup> and mortality.<sup>4</sup> Even in nondemented PD patients, impaired cognitive test performance has an impact on ability to perform activities of daily living (ADLs).<sup>5-8</sup> Cognitive impairment in PD is also increasingly recognized as a potential therapeutic target, and treatment benefit should include improvement in function.

Cognitive impairment in PD of mild-to-moderate severity predominantly affects ability to perform instrumental ADLs (IADLs; e.g., managing money and shopping) rather than basic ADLs (e.g., bathing and dressing). Improved IADL measurement facilitates testing of new cognitive enhancing treatments in PD. Several existing scales used to measure IADLs in PD clinical trials were developed for use in Alzheimer's disease (AD),<sup>9</sup> but do not take into account the specific features of PD, including motor symptoms and impairments in multiple cognitive domains.<sup>10</sup>

In instrument development, item response theory (IRT) has potential advantages over classic test theory. These include producing consistent reliability across a broad range of impairments and the ability to effectively utilize computerized adaptive testing (CAT; a strategy whereby subsequent items are chosen based on response to previous items), which can reduce the time and burden required to obtain adequate measurement precision.<sup>11,12</sup>

We describe the development and initial psychometric evaluation of the Penn Parkinson's Daily Activities Questionnaire (PDAQ), an IRT-based measure of IADL function in PD. Evidence of reliability and validity were obtained from independent cohorts. To assess test-retest reliability and construct validity, the PDAQ was administered with established measures of cognitive abilities and motor function, a validated questionnaire of ADLs for dementia patients, and a standard performance-based assessment of directly observed tasks of daily function.

## Materials and Methods

### Overview of PDAQ Development

**Phase I**—First, a comprehensive list of IADLs was assembled by reviewing existing literature on IADL scales in the public domain, particularly those from the Neuro-QoL banks,<sup>13</sup> and conducting focus groups with PD patients and their knowledgeable informants (KIs). KIs were required to have daily contact with the PD patient to qualify for participation. Four focus groups (5–10 participants each) comprised of nondemented PD patients (1 meeting) and KIs, relatives, or close friends (three meetings) were held to review existing items and identify IADLs not addressed by existing scales. Participants were recruited from the University of Pennsylvania's Parkinson's Disease and Movement Disorders Center (PD&MDC) and local support groups. An internal advisory board consisting of local investigators (A.S., J.S., J.K., D.W., S.S., J.R., J.R., L.B.) and an external advisory board (P.C., J.G., B.V., L.S., C.N.) consisting of nationally recognized experts in PD, cognitive impairment, and scale development were assembled to review items and provide guidance throughout the development process.

Next, we created a preliminary bank of 128 items formatted to be consistent with Neuro-QoL item banks for proxy response. Each item consisted of a description of an IADL and a KI rating of difficulty the PD patient has in performing that IADL on the following scale: “none,” “a little,” “somewhat,” “a lot,” “cannot do.” Initial item reduction removed 30 items based on expert consensus.

**Phase II**—The remaining 98 items were administered to a large cohort of KIs able to provide proxy responses about functional status. KIs, rather than patients, were chosen as the primary respondents to optimize accurate assessment of patients across a broad range of cognitive ability and facilitate instrument use in longitudinal studies. Patients and KIs were recruited from the PD&MDC, two community-based neurology practices in the Philadelphia region (n = 239), and by postings on the University of Pennsylvania Morris K. Udall Parkinson's Disease Research Center website and the National Parkinson Foundation website inviting participation by Web-based responding (n = 149). In addition to responding to the items, KIs gave a rating of disease severity on a 5-point scale that followed the levels of the H & Y scale.<sup>14</sup> KIs also answered yes/no questions regarding the presence of depression and cognitive impairment in PD patients. If depression or cognitive impairment was present, KIs rated severity on a 3-point scale (mild, moderate, and severe). KIs followed the same procedure to self-report their own level of depression (yes/no; if present, mild, moderate, or severe).

Based on initial statistics from KI responses, the item bank was reduced to 50 items, which were calibrated using IRT methods to estimate parameters for each item. Parameters estimated were slopes (item discrimination) and thresholds (locations between response categories on the IADL ability scale). From these item parameters, we calculated each item's information (measurement precision) and each respondent's theta (ability to perform an IADL).

## PDAQ Reliability and Validity

We created four cohorts of patients with the following characteristics: (1) diagnosis of idiopathic PD<sup>15</sup>; (2) cognitive diagnoses from normal cognition to demented; (3) those who had not participated in the item generation/calibration phase; and (4) those who had a KI available to complete the PDAQ. PDAQ responses were obtained by paper administration. The study was approved by the University of Pennsylvania Institutional Review Board, and informed consent was obtained from all participants. Neuropsychological assessments, motor examinations, and direct observation of daily functioning of PD patients were all performed while in the levodopa *on* state.

**Neuropsychological Assessment**—Trained research staff administered the Mattis Dementia Rating Scale 2 (MDRS-2)<sup>16</sup> as an assessment of global cognition.

**Assessment of Dementia**—Cognitive diagnoses were made during regular, established consensus conferences conducted by neurologists and psychiatrists with expertise in PD cognition at the Penn Udall Center. Diagnoses of dementia<sup>17</sup> and mild cognitive impairment<sup>18</sup> were based on International Parkinson and Movement Disorder Society (MDS) criteria.

**Motor Examination**—Motor subscale (part III) of the UPDRS<sup>19</sup> and H & Y<sup>14</sup> staging were conducted by trained and certified research staff. Motor examinations were conducted within 2 months of cognitive testing.

**ADLs**—ADLs were also evaluated with the Alzheimer's Disease Cooperative Study Activities of Daily Living Inventory (ADCS-ADL),<sup>9</sup> which is widely used in AD and also in PD.<sup>20</sup> The ADCS-ADL was also completed by a KI, either during the research visit or by telephone within 2 months. The ADCS-ADL consists of 23 ADL and IADL items (range, 0–78), with higher scores indicating better function.

**Direct Observation of Daily Function**—PD patients' ability to perform ADLs was directly observed using the Direct Assessment of Functional Status (DAFS) scale.<sup>21</sup> The DAFS, a structured, performance-based assessment of daily function using standard activities and props (e.g., telephone, checkbook, grocery boxes and pill box), was administered by trained research staff. The seven activities are time orientation, communication, transportation, finances, shopping, grooming and eating, and medication management. Total scores range from 0 to 56, with higher scores indicating better performance. Scores from this measure have been shown to have excellent test-retest reliability along with evidence of construct validity relative to other measures of functional status in the elderly.<sup>21,22</sup>

## Statistical Analysis

IRT parameters were estimated using Samejima's homogeneous Graded Response Model (GRM).<sup>23</sup> The GRM is a cumulative categories approach to polytomous IRT modeling. Polytomous modeling refers to items with multiple levels, such as the Likert scale responses used here. The threshold parameter of each of these scoring functions refers to the latent trait

score needed to have a 0.5 probability of endorsing a particular response option or higher. This model was chosen because it is theoretically appropriate for items with ordered thresholds representing increasing quantities of a given attribute (i.e., IADL ability). The process of item generation and reduction is shown in Figure 1.

Test-retest reliability regarding KI's responses to the PDAQ was calculated using the intraclass correlation coefficient (ICC).<sup>24</sup> ICC analysis was based on the one-way random-effects model for absolute agreement on single measurements. The association between PDAQ scores and measures of cognition and motor function was assessed using linear regression and partial correlation analysis. Age, gender, and education were included as covariates in these models. Partial correlation coefficients were calculated to estimate the independent correlation between the PDAQ (or ADCS-ADL) and cognitive and motor function.

Association between directly observed ADL function and the PDAQ and ADCS-ADL was assessed using regression and partial correlation, adjusting for age, gender, education, and motor symptoms. Receiver operating characteristic (ROC) analysis was used to measure the ability of the PDAQ to distinguish between subjects with and without dementia. All analyses were conducted at a two-sided alpha = 0.05 significance level, without adjustment for multiple comparisons. Analyses were carried out using STATA (version 10; StataCorp LP, College Station, TX) or SPSS software (version 22; SPSS, Inc., Chicago, IL).

## Results

### Characteristics of PDAQ Development Response Sample

A cohort of 388 PD patients and their KIs were in the response sample. KIs completed the 50-item PDAQ. Clinical characteristics of PD patients that were rated and the relationship of the KIs to the patients are shown in Table 1.

### Factor Structure

Principal components analysis (PCA) was carried out after sampling adequacy (Keiser-Meyer-Olkin statistic = 0.99) and sphericity ( $P < 0.001$ ) were checked. The PCA showed one dominant factor for the 50 items in the final bank. The eigenvalue for that factor was 35.0, which accounted for 70% of the variance between subjects. The second factor had an eigenvalue of 1.6 (3% of the variance), and the third factor had an eigenvalue of 1.2 (2% of the variance). After oblique rotation, the correlation between factors 1 and 2 was 0.81, between factors 1 and 3 was 0.76, and between factors 2 and 3 was 0.74. Thus, the correlation among these factors and the large percentage of variance explained by a single factor support the unidimensionality of the PDAQ battery. Additionally, poly-DIMTEST ( $T = 0.83$ ;  $P = 0.2$ ) confirmed the essential unidimensionality of the data, indicating measurement of a single latent trait (IADL function) and satisfying the IRT assumption of unidimensionality when utilizing polytomous data.

## Reading Level and Item Characteristics

Average reading level for all items based on the Lexile system ([www.lexile.com](http://www.lexile.com)) was 596, consistent with a third-grade level. IRT methods include estimation of several parameters for each item, including *theta* (or ability) and information. In IRT, *theta* represents ability, both the amount of a trait possessed by an examinee and how much IADL function is required to complete a given task. It is measured as a *z* score (based on the number of standard deviations [SDs] from the cohort mean). Threshold parameters for individual IADL items indicate the location of each on the *theta* (ability) scale (i.e., how difficult a given IADL is relative to other ones). More extreme threshold scores indicate items are either very easy or very difficult in general. For example, the threshold for the transition from *none* to *a little* difficulty for the item “. . .remembering a list of 4 or 5 errands without writing it down” was a *z* score of 1.3, indicating that a high level of ability is required to perform this task. The threshold for the transition from *a lot* of difficulty to *cannot do* for the item “. . .identifying the rooms in your home” was a *z* score of -3.3, indicating that even individuals with a low level of ability were still able to complete this task.

We measured the information provided by each item and the entire scale. In IRT, the Fisher information function describes measurement precision along the ability scale and is defined as the inverse of the squared standard error. In the present study, the standard error represents the measurement precision for test takers of varying levels of IADL ability. The test information function is an additive combination of the information provided by each individual item. The location of, and maximum value for, the information provided by each item are given in Supporting Table 1. The amount of maximum information can be used to identify items with more discriminating power, and the quadrature point listed notes where along the ability scale each item has the most discriminating power.

The conditional standard error and test information functions for the scale as a whole are shown in Figure 2. Figure 2 shows that the battery has low levels of standard error (i.e., high levels of test information) across a broad range of abilities, indicating that the PDAQ demonstrates strong discriminative power across the spectrum of IADL functioning.

## Relationship Between Item Bank Scores and KI Ratings of Patient Symptom Severity and KI Self-Reported Depression

To assess preliminary construct validity, we tested for differences in IRT-derived PDAQ scores (*theta*) in patients at different levels of cognitive, affective, and motor impairment. In bivariate analysis, there was a strong relationship between cognitive status and PDAQ score; mean *theta* scores for the four cognitive impairment levels were 1.24 (no impairment; SD = 0.78), 0.65 (mild; SD = 0.79), -0.43 (moderate; SD = 0.79), and -1.65 (severe; SD = 0.97) (all  $P < 0.001$ ). There was also a relationship between disease severity and PDAQ score; mean *theta* scores for the five motor levels were 1.23 (I; SD = 0.82), 0.81 (II; SD = 0.87), 0.34 (III; SD = 0.98), -0.55 (IV; SD = 0.86), and -1.33 (V; SD = 1.24) (all  $P < 0.001$ ). However, the association between *theta* and cognitive status was substantially stronger than the association between *theta* and disease severity; the adjusted partial correlation coefficient for the association between *theta* and cognitive severity was nearly twice as high ( $r = 0.60$ )

as the adjusted partial correlation between theta and disease severity ( $r = 0.31$ ). Neither patient nor KI depression was associated with theta scores.

### Characteristics of PDAQ Reliability and Validity Cohorts

The independent test-retest cohort consisted of 50 PD patients and their KI, and the independent validation cohort consisted of 68 PD subjects and their KI (Table 1). Of the validation group, 23 also participated in the test-retest analysis.

### Test-Retest Reliability

The PDAQ was administered to the same KI twice to evaluate test-retest reliability, with an interval between administrations ranging from 23 to 185 days. The ICC was 0.97 ( $P < 0.001$ ). The length of time between administrations did not affect agreement ( $P = 0.33$ ). The ICC remained high (0.94;  $P < 0.001$ ), even after excluding subjects with extreme PDAQ scores (i.e., less than  $-2$  or greater than  $2$ ).

### Correlation with Motor and Cognitive Function

The PDAQ score showed strong correlation with total MDRS-2 total score ( $r = 0.71$ ;  $P < 0.001$ ), even when adjusting for motor function ( $r = 0.68$ ;  $P < 0.001$ ).

The PDAQ was more specific for cognitive impairment than the ADCS-ADL scale. Although the PDAQ correlated with total UPDRS score in bivariate analysis ( $r = 0.43$ ;  $P < 0.001$ ), the association was no longer significant after adjustment for total MDRS-2 ( $r = 0.08$ ;  $P = 0.53$ ). By comparison, the ADCS-ADL correlated even more strongly with motor performance ( $r = 0.64$ ;  $P < 0.001$ ) than the PDAQ, and the association between ADCS-ADL and UPDRS part III remained significant after adjustment for total MDRS-2 score ( $r = 0.44$ ;  $P < 0.001$ ). In a multivariable model including age, sex, education, UPDRS part III, and MDRS-2, MDRS-2 ( $r = -0.72$ ;  $P < 0.001$ ), but not UPDRS ( $r = -0.13$ ;  $P = 0.32$ ), was significantly associated with PDAQ score. In a similar model substituting ADCS-ADL for PDAQ, both MDRS-2 ( $r = 0.57$ ;  $P < 0.001$ ) and UPDRS ( $r = -0.41$ ;  $P < 0.001$ ) were significantly associated with ADCS-ADL score.

### Correlation With Directly Observed Daily Function

The PDAQ correlated strongly with directly observed daily function (i.e., DAFS score;  $r = 0.78$ ;  $P < 0.001$ ; Fig. 3). In a model that controlled for UPDRS motor score, the PDAQ retained its association with the DAFS score ( $r = 0.79$ ;  $P < 0.001$ ).

### Discrimination Between Demented and Nondemented Subjects

Twenty-two participants in this study were classified as cognitively intact, 21 classified as mild cognitive impairment (MCI), and 36 as demented (11 of these 36 demented subjects could not complete the MDRS-2 and were therefore not included in above correlation analyses). The mean PDAQ theta group scores were: 1.5 ( $\pm 0.72$ ) for normal cognition;  $-0.57$  ( $\pm 0.65$ ) for MCI; and  $-0.80$  ( $\pm 1.17$ ) for dementia. The optimal cutoff between demented and nondemented subjects was  $-0.16$  (sensitivity 78%, specificity 80%). The optimal cutoff between intact and MCI/dementia was  $-0.83$  (sensitivity, 81%; specificity, 91%). ROC

analysis for discrimination between intact, MCI, and demented subjects is shown in Figure 4. Box 1 presents theta scores for four representative PD patients (i.e., no cognitive impairment, MCI, mild dementia, and severe dementia), along with their respective scores on the UPDRS, ADCS-ADL, MDRS-2, and DAFS, as an aid to illustrate and interpret the clinical characteristics that would be anticipated in a PD patient with a given theta score.

## Discussion

The results of this study establish parameters of the PDAQ item bank and provide evidence for the reliability and construct validity of the PDAQ in PD. The 50 items that comprise the PDAQ cover IADLs across a range of abilities. We demonstrate strong construct validity relative to existing ADL scales and directly observed ADL function in PD patients across a range of cognitive abilities. We also show significant test-retest reliability. In addition, the PDAQ has the theoretical properties of IRT-based administration, including comparability of responses across a broad range of functions and the potential for efficient, computer-based administration (CAT).

The items that comprise the PDAQ were initially conceptualized as belonging to distinct domains of IADL function (e.g., communication, decision making, and financial planning). Items from each of these categories were retained so that the scale would be more intuitive for potential users, including physicians, patients, and caregivers. However, both exploratory and confirmatory factor analysis showed that all items could be conceptualized as belonging to a single dominant factor, and thus the scale is suitable for CAT administration. CAT administration can be implemented on platforms such as the NIH-PROMIS Assessment Center Web-based survey tool ([www.assessmentcenter.net](http://www.assessmentcenter.net)).

Importantly, the PDAQ score, after adjustment for cognitive function, was not affected by motor performance, in contrast to the ADCS-ADL scale, which showed a strong correlation with the motor scores. Thus, the PDAQ is well suited to PD studies that seek to separate the impact of cognition from motor function, including treatment studies that may improve both motor performance and cognition.

The PDAQ should be completed by a KI rather than the patient in studies that include patients with dementia. We found a high degree of correlation between functional status as reported by KIs, mainly spouse caregivers, and directly observed daily function. Assessment by KIs has several attributes that make it preferable to assessment by patients with significant cognitive impairment. That is, moderately and severely demented patients are not able to provide reliable estimates of constructs such as ADL function and health-related quality of life,<sup>25,26</sup> and greater cognitive impairment has been associated with overestimation of functional capacity in PD patients.<sup>27</sup> Use of a single rating perspective is an overriding consideration, particularly in longitudinal studies, where a high percentage of patients transition to dementia. Future studies can compare patient- and KI-rated PDAQ scores in nondemented patients.

The PDAQ bank should be viewed in the context of the over 75 different ADL instruments that have been published in the medical literature.<sup>28</sup> However, a smaller number have been



applied to PD,<sup>29,30</sup> and of these scales most (e.g., Schwab and England ADL index<sup>31</sup> and ADL section of the UPDRS<sup>32</sup>) emphasize activities that are dependent on motor function. The ADCS-ADL scale was chosen as a comparator for our study because it is the “gold standard” for AD and has also been used to identify the effects of cognitive enhancing treatments in PDD and dementia with Lewy bodies.<sup>20,33</sup>

The Pill Questionnaire was included as a measure of daily function in the MDS guidelines for the diagnosis of PDD<sup>17</sup> and has since been studied across the spectrum of cognition in PD.<sup>34,35</sup> Whereas the Pill Questionnaire is a valid screening tool, it may not be appropriate as a continuous measure of ADL ability for uses such as assessing change in ADL function over time. The Cognitive Functional Rating Scale (PD-CFRS)<sup>30,36</sup> is a 12-item questionnaire specifically designed to be sensitive to mild effects of cognitive decline in PD; it is administered to a KI in interview format. The initial validation study of the PD-CFRS provided evidence for reliability and discriminant validity across stages of cognitive impairment. The use of IRT methodology, a larger number of items, and validation in a larger sample differentiate the PDAQ from the PD-CFRS.

The PDAQ has a number of potential uses. It was designed with the intent to be used as an outcome measure in clinical trials of treatments that could affect both motor and cognitive performance in PD, but where the impact of cognition was of particular interest. Our study supports this use, but definitive studies showing that the PDAQ is sensitive to treatment effects are still needed. The PDAQ has other potential applications, including observational studies in which classification of patients into demented and nondemented groups is based partly on assessment of functional status.

This study should be considered in light of several limitations. The PDAQ item bank shares content with several publicly available scales, notably the Neuro-QoL bank; however, many ADL scales have similar content, and the PDAQ does include items intended to specifically examine cognitive deficits reported in PD. Additionally, it has not been validated for patient reporting. Although the MDRS-2 has been extensively studied in PD, it does not specifically target the cognitive profile, including problems with planning and organization, most typical of PD and may also be susceptible to ceiling effects. These limitations may have impacted the observed relationship between ADL function and global cognition in this study. Additionally, all clinical assessments were performed in the *L*-dopa *on* state, and associations may have been impacted if PD patients were assessed in the *off* state. Another limitation is that a portion of respondents used for psychometric testing in the PDAQ instrument development cohort included Web respondents, among whom the clinical diagnosis of PD was not verified. In addition, this cohort consisted of a high proportion of patients with mild PD symptoms. However, parameters from item response models are independent of study sample. Nonetheless, the internal consistency reliability of the items should be confirmed in other cohorts.

In summary, the PDAQ is a newly developed instrument for assessment of cognitive IADLs in patients with PD. This study provides evidence supporting the test-retest reliability of PDAQ scores, along with construct validity. Future studies are needed to replicate these

results and assess responsiveness to changes both in cognition over time and to therapeutic interventions.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The authors thank the external advisory board members who assisted in the PDAQ item bank development, including Peter Como, PhD (U.S. Food and Drug Administration, Silver Spring, MD), Jennifer Goldman, MD, MS (Department of Neurological Sciences, Section of Parkinson Disease and Movement Disorders, Rush University Medical Center, Chicago, IL), Barbara G. Vickery, MD, MPH (UCLA Department of Neurology and VA Greater Los Angeles Health Care System, Los Angeles, CA), Lisa M. Shulman, MD (Department of Neurology, University of Maryland School of Medicine, Baltimore, MD), and Cindy J. Nowinski, MD, PhD (Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL).

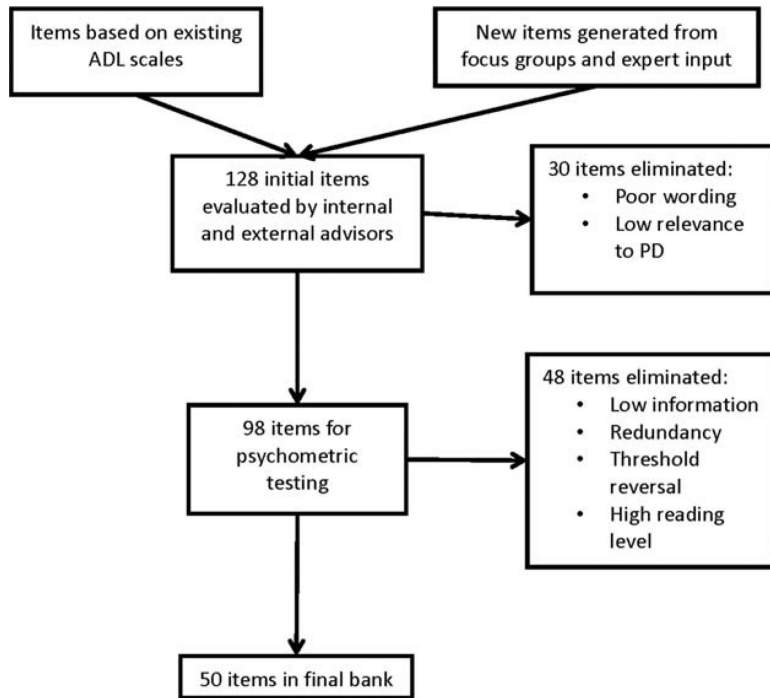
**Funding agencies:** This study was funded by a Morris K. Udall Parkinson's Disease Research Center of Excellence grant from the National Institute of Neurological Disorders and Stroke (NS-053488) and by SAP4100027296, a health research grant awarded by the Department of Health of the Commonwealth of Pennsylvania from the Tobacco Master Settlement Agreement under Act 2001-77.

## References

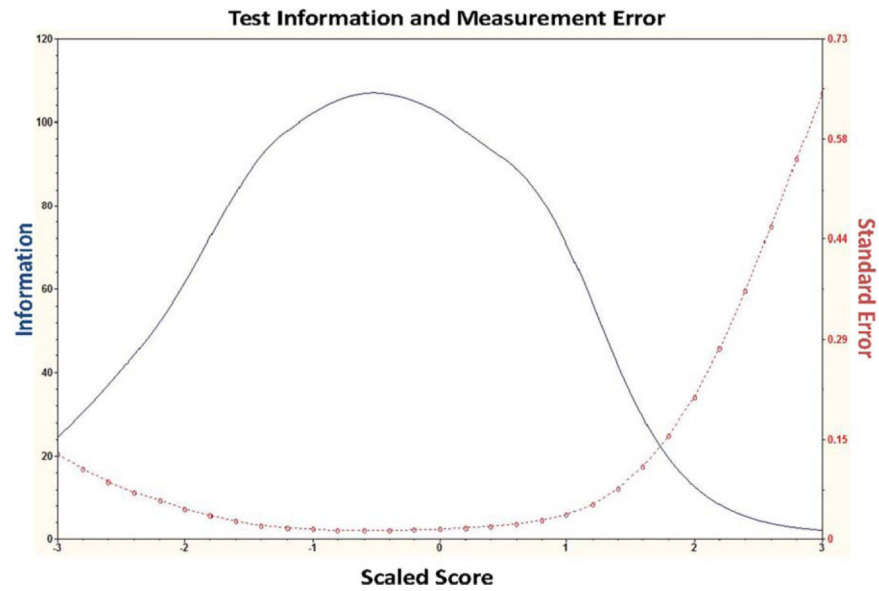
1. Bronnick K, Ehrt U, Emre M, et al. Attentional deficits affect activities of daily living in dementia-associated with Parkinson's disease. *J Neurol Neurosurg Psychiatry*. 2006; 77:1136–1142. [PubMed: 16801351]
2. Cahn D, Sullivan E, Shear P, Pfefferbaum A, Heit G, Silverberg G. Differential contributions of cognitive and motor component processes to physical and instrumental activities of daily living in Parkinson's disease. *Arch Clin Neuropsychol*. 1998; 13:575–583. [PubMed: 14590618]
3. Aarsland D, Larsen JP, Karlsen K, Lim G, Tandberg E. Mental symptoms in Parkinson's disease and important contributors to caregiver distress. *Int J Geriatr Psychiatry*. 1999; 14:866–874.
4. Buter TC, van der Hour A, Matthews F, Larsen JP, Brayne C, Aarsland D. Dementia and survival in Parkinson disease A 12-year population study. *Neurology*. 2008; 70:1017–1022. [PubMed: 18362281]
5. Martin RC, Triebel KL, Kennedy RE, et al. Impaired financial abilities in Parkinson's disease patients with mild cognitive impairment and dementia. *Parkinsonism Relat Disord*. 2013; 19:986–990. [PubMed: 23899743]
6. Pirogovsky E, Martinez-Hannon M, Schiehser DM, et al. Predictors of performance-based measures of instrumental activities of daily living in nondemented patients with Parkinson's disease. *J Clin Exp Neuropsychol*. 2013; 35:926–933. [PubMed: 24074137]
7. Pirogovsky E, Schiehser DM, Obtera KM, et al. Instrumental activities of daily living are impaired in Parkinson's disease patients with mild cognitive impairment. *Neuropsychology*. 2014; 28:229–237. [PubMed: 24417192]
8. Rosenthal E, Brennan L, Xie S, et al. Association between cognition and function in patients with Parkinson disease with and without dementia. *Mov Disord*. 2010; 25:1170–1176. [PubMed: 20310053]
9. Galasko D, Bennett D, Sano M, et al. An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. The Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord*. 1997; 11(Suppl 2):S33–S39. [PubMed: 9236950]
10. Starkstein SE, Sabe L, Petracca G, et al. Neuropsychological and psychiatric differences between Alzheimer's disease and Parkinson's disease with dementia. *J Neurol Neurosurg Psychiatry*. 1996; 61:381–387. [PubMed: 8890777]
11. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care*. 2000; 38:II28–II42. [PubMed: 10982088]

12. Lord, FM. Applications of Item Response Theory to Practical Testing Problems. Routledge; New York: 1980. Available at: [http:// books.google.com/books?hl5en&lr5&id5\\_qu8YGkdTvwC&pgis51](http://books.google.com/books?hl5en&lr5&id5_qu8YGkdTvwC&pgis51). [21 October 2014]
13. Gershon R, Lai J, Bode R, Choi S, Moy C, Bleck T. Neuro-QOL: quality of life item banks for adults with neurological disorders: item development and calibrations based upon clinical and general population testing. *Qual Life Res.* 2012; 21:475–486. [PubMed: 21874314]
14. Hoehn M, Yahr M. Parkinsonism: onset, progression and mortality. *Neurology.* 1967; 17:427–442. [PubMed: 6067254]
15. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry.* 1992; 55:181–184. [PubMed: 1564476]
16. Mattis, S. Dementia Rating Scale. Psychological Assessment Resources; Odessa, FL: 1988.
17. Dubois B, Burn D, Goetz C, et al. Diagnostic procedures for Parkinson's disease dementia: recommendations from the Movement Disorder Society Task Force. *Mov Disord.* 2007; 22:2314–2324. [PubMed: 18098298]
18. Litvan I, Goldman JG, Tröster AI, et al. Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines. *Mov Disord.* 2012; 27:349–356. [PubMed: 22275317]
19. Fahn, S.; Elton, R. Unified Parkinson's Disease Rating Scale.. In: Fahn, S.; Marsden, C.; Calne, D.; Goldstein, M., editors. *Recent Developments in Parkinson's Disease.* Macmillan Health Care Information; Florham Park, NJ: 1987. p. 153-164.
20. Emre M, Aarsland D, Albanese A, et al. Rivastigmine for dementia associated with Parkinson's disease. *N Engl J Med.* 2004; 351:2509–2518. [PubMed: 15590953]
21. Loewenstein DA, Amigo E, Duara R, et al. A new scale for the assessment of functional status in Alzheimer's disease and related disorders. *J Gerontol.* 1989; 44:P114–P121. [PubMed: 2738312]
22. Loewenstein, DA.; Bates, CB. Manual for Administration and Scoring. Neuropsychological Laboratories and the Wien Center for Alzheimer's Disease and Memory Disorders, Mount Sinai Medical Center; Miami, FL: 2006. The Direct Assessment of Functional Status Revised (DAFS-R)..
23. Samejima. Graded Response Model.. In: Kempf-Leonard, K., editor. *Encyclopedia of Social Measurement.* Academic; New York: 2004.
24. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas.* 1973; 33:613–619.
25. Coucill W, Bryan S, Bentham P, Buckley A, Laight A. EQ-5D in patients with dementia: an investigation of inter-rater agreement. *Med Care.* 2001; 39:760–771. [PubMed: 11468496]
26. Boyer F, Novella JL, Morrone I, Jolly D, Blanchard F. Agreement between dementia patient report and proxy reports using the Nottingham Health Profile. *Int J Geriatr Psychiatry.* 2004; 19:1026–1034. [PubMed: 15481077]
27. Shulman LM, Pretzer-Aboff I, Anderson KE, et al. Subjective report versus objective measurement of activities of daily living in Parkinson's disease. *Mov Disord.* 2006; 21:794–799. [PubMed: 16482533]
28. McHorney CA, Cohen AS. Equating health status measures with item response theory: illustrations with functional status items. *Med Care.* 2000; 38(9 Suppl):II43–II59. [PubMed: 10982089]
29. Ramaker C, Marinus J, Stiggelbout AM, Van Hilten BJ. Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. *Mov Disord.* 2002; 17:867–876. [PubMed: 12360535]
30. Marras C, Tröster AI, Kulisevsky J, Stebbins GT. The tools of the trade: a state of the art “How to Assess Cognition” in the patient with Parkinson's disease. *Mov Disord.* 2014; 29:584–596. [PubMed: 24757108]
31. Schwab RS, England AC, Poskanzer DC, Young RR. Amantadine in the treatment of Parkinson's disease. *JAMA.* 1969; 208:1168–1170. [PubMed: 5818715]
32. Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov Disord.* 2008; 23:2129–2170. [PubMed: 19025984]

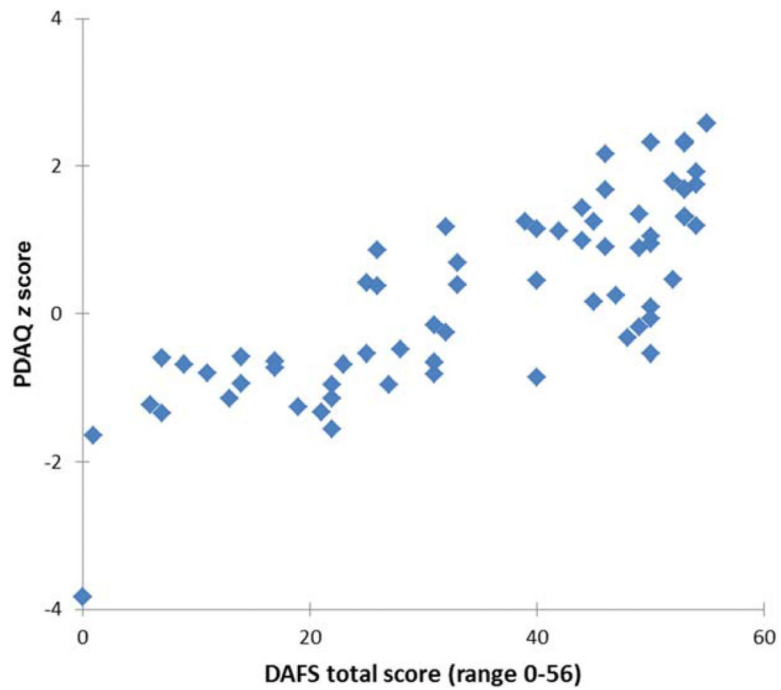
33. McKeith I, Del Ser T, Spano P, et al. Efficacy of rivastigmine in dementia with Lewy bodies: a randomised, double-blind, placebo-controlled international study. *Lancet*. 2000; 356:2031–2036. [PubMed: 11145488]
34. Lee WJ, Chang YY, Lin JJ, et al. Comparison of activities of daily living impairments in Parkinson's disease patients as defined by the Pill Questionnaire and assessments by neurologists. *J Neurol Neurosurg Psychiatry*. 2013;969–973. [PubMed: 24218526]
35. Martinez-Martin P. Dementia in Parkinson's disease: usefulness of the pill questionnaire. *Mov Disord*. 2013; 28:1832–1837. [PubMed: 24123241]
36. Kulisevsky J, Fernandez de Bobadilla R, Pagonabarraga J, et al. Measuring functional impact of cognitive impairment: validation of the Parkinson's disease cognitive functional rating scale. *Parkinsonism Relat Disord*. 2013; 19:812–817. [PubMed: 23773412]



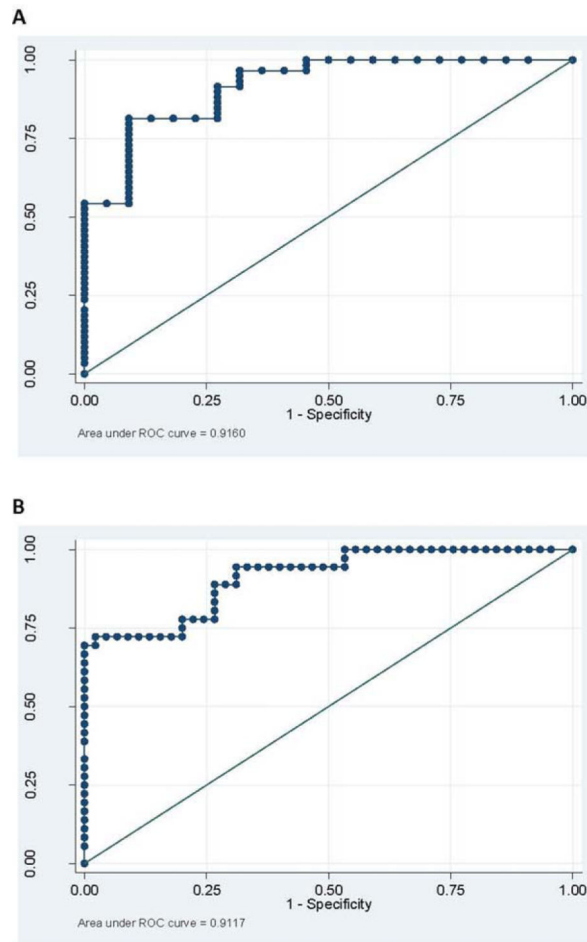
**FIG. 1.**  
Item generation and reduction.



**FIG. 2.** Item information: test information function (solid line) and conditional standard error curve (dotted line). These curves indicate a high degree of ability to discriminate between different levels of ADL ability across a broad range of function. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]



**FIG. 3.** Scatterplot showing relationship between DAFS total scores and PDAQ scores. This figure shows a high correlation between the PDAQ ability score and directly observed daily function ( $r = 0.78$ ;  $P < 0.001$ ). [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]



**FIG. 4.** ROC curves for the distinction between intact subjects and those with either dementia or MCI (A) and for the distinction between intact/MCI subjects and those with dementia (B). The area under the ROC curve for the distinction between intact and MCI/dementia is 0.92 (95% confidence interval [CI]: 0.85–0.98), and the area under the ROC curve for the distinction between intact/MCI and dementia is 0.91 (95% CI: 0.85–0.97). [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]



**TABLE 1**

Characteristics of participants in development, reliability, and validity cohorts

	Scale Development Cohort, n = 388 Mean (SD)	Reliability Cohort, n = 50 Mean (SD)	Validity Cohort, n = 68 Mean (SD)
Age	71.70 (9.90) <sup>a</sup>	73.88 (8.11)	74.19 (8.67)
Gender, % male	50.6	76.0	78.0
Disease duration, years	3.17 (3.67)	9.57 (5.64)	7.90 (5.40)
Education, years	n/a	14.80 (3.07)	15.38 (2.82)
MDRS-2 total raw score	n/a	127.58 (16.21)	120.91 (19.67)
UPDRS, part III	n/a	34.68 (15.07)	34.40 (16.06)
KI Relationship, n = 357 (%)			
Spouse	202 (56.6)		
Child	105 (29.4)		
Other	50 (14.0)		
KI rating of patient motor severity, n = 375 (%)			
I	55 (14.7)		
II	43 (11.5)		
III	174 (46.4)		
IV	68 (18.1)		
V	35 (9.3)		
KI rating of patient cognitive severity, n = 369 (%)			
None	69 (18.7)		
Mild	139 (37.7)		
Moderate	124 (33.6)		
Severe	37 (10.0)		
KI rating of patient depression, n = 388 (%)			
None	121 (32.2)		
Mild	102 (26.3)		
Moderate	123 (31.7)		
Severe	42 (10.8)		
KI self-reported depression, n = 383 (%)			
None	160 (41.8)		
Mild	109 (28.5)		
Moderate	95 (24.8)		
Severe	19 (5.0)		

n/a, not applicable.

<sup>a</sup> Overall, 149 patients in the scale development cohort who participated by Web survey did not have data available regarding date of birth/age. Age is reported for remaining 239 subjects.

**BOX 1**

Interpreting Individual PDAQ theta (ability) scores - In order to provide context regarding interpretation of ability scores generated by the IRT model for representative patients, thetas are provided along with measures of motor, cognitive, and functional impairment.

<b>PD patient's cognitive status</b>	<b>Theta</b>	<b>MDRS-2</b>	<b>UPDRS-III</b>	<b>ADCS-ADL</b>	<b>DAFS</b>
Patient 1 (intact)	1.68	138	23	78	53
Patient 2 (MCI)	0.47	134	22	77	52
Patient 3 (mild dementia)	-0.54	113	28	49	25
Patient 4 (severe dementia)	-1.65	101	37	17	1

Higher scores on the UPDRS-III indicate greater motor impairment. Lower scores on the MDRS-2, ADCS-ADL, and DAFS indicate greater cognitive and functional impairment, respectively.