UC Davis
UC Davis Previously Published Works

Title

An item response theory analysis of the Executive Interview and development of the EXIT8:
A Project FRONTIER Study.

Permalink

https://escholarship.org/uc/item/7r7581w8

Journal

Journal of clinical and experimental neuropsychology, 37(3)

ISSN

1380-3395

Authors

Jahn, Danielle R
Dressel, Jeffrey A
Gavett, Brandon E
et al.

Publication Date

2015

DOI

10.1080/13803395.2014.1002757

Peer reviewed

An Item Response Theory Analysis of the Executive Interview and Development of the EXIT8: A Project FRONTIER Study

Danielle R. Jahn, Ph.D.,

Department of Psychological Sciences, Texas Tech University, MS 2051 Psychology Building, Lubbock, TX 79409, (806) 742-3711, danielle.jahn@ttu.edu

Jeffrey A. Dressel, Ph.D.,

Human Solutions, Inc., 600 Maryland Avenue SW, Suite 800E, Washington, DC 20024, (202) 479-2057, jeffdressel@gmail.com

Brandon E. Gavett, Ph.D., and

Psychology Department, University of Colorado at Colorado Springs, 1420 Austin Bluffs Parkway, Colorado Springs, CO 80918, (719) 255-4135, bgavett@uccs.edu

Sid E. O'Bryant, Ph.D.

University of North Texas Health Sciences Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107, (817) 735-2961, sid.o'bryant@unthsc.edu

Abstract

Introduction—The EXIT25 is an effective measure of executive dysfunction, but may be inefficient due to the time it takes to complete 25 interview-based items. The current study aimed to examine psychometric properties of the EXIT25, with a specific focus on determining if a briefer version of the measure could comprehensively assess executive dysfunction.

Method—The current study applied a graded response model (a type of item response theory model for polytomous categorical data) to identify items that were most closely related to the underlying construct of executive functioning and best discriminated between varying levels of executive functioning. Participants were 660 adults ages 40 to 96 living in West Texas, who were recruited through an ongoing epidemiological study of rural health and aging, called Project FRONTIER. The EXIT25 was the primary measure examined. Participants also completed the Trail Making Test and Controlled Oral Word Association Test, among other measures, to examine the convergent validity of a brief form of the EXIT25.

Results—Eight items were identified that provided the majority of the information about the underlying construct of executive functioning; total scores on these items were associated with total scores on other measures of executive functioning and were able to differentiate between cognitively healthy, mildly cognitively impaired, and demented participants. In addition, cutoff scores were recommended based on sensitivity and specificity of scores.

Address correspondence (including requests for reprints) to: Sid E. O'Bryant, University of North Texas Health Sciences Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107; sid.o'bryant@unthsc.edu; Phone: (817) 735-2961; Fax: (817) 735-0611..

Conclusion—A brief, eight-item version of the EXIT25 may be an effective and efficient screening for executive dysfunction among older adults.

Keywords

Item response theory; Executive functioning; Aging; Brief assessment; Cognition

Executive functioning (EF) is an important consideration in aging, as the aging process can affect components of EF (Phillips & Henry, 2008). In addition, disorders associated with aging (e.g., dementia due to Alzheimer's disease, frontotemporal dementia) are associated with executive dysfunction (Elliott, 2003). EF is generally defined as a set of multifaceted cognitive processes that combine subprocesses such as problem solving, planning, reasoning, and self-monitoring to achieve goals or desired outcomes (Elliott, 2003; Phillips & Henry, 2008). Cognitive flexibility and coordination of complex processes are critical components of EF (Elliott, 2003; Phillips & Henry, 2008). Neuroimaging research has linked EF to the frontal lobe and prefrontal cortex; this link may explain the executive dysfunction found in disorders like frontal lobe dementia, as this area of the brain loses volume and becomes damaged (Elliott, 2003). Even in normal aging, there can be changes in brain structure and function (de Lucena Ferretti et al., 2010; Sonnen et al., 2011), which makes it critical to identify potential problems in EF among cognitively healthy adults and older adults, as well as those with cognitive disorders.

Poor EF has been linked to variety of negative outcomes in older adults. For example, executive dysfunction has been linked to functional impairment (Pereira, Yassuda, Oliveira, & Forlenza, 2008), problem gambling (Von Hippel et al., 2009), poorer glycemic control (among those with diabetes; Nguyen et al., 2010), and poorer medication adherence (among those who are HIV-positive; Ettenhoffer et al., 2009). In addition, EF has been identified as a key mediating variable in relations between changes in brain function (i.e., reduced lateralization of tasks) and structure (i.e., white matter hyperintensities) and memory (Angel, Fay, Bouazzaoui, & Isingrini, 2011; Parks et al., 2011). Because many adults and older adults experience changes in EF as they age, and because EF has been linked to negative outcomes in older adults, it is important for health professionals to have effective methods to measure EF as adults age.

One commonly utilized measure of EF is the Executive Interview (EXIT25; Royall, Mahurin, & Gray, 1992). The EXIT25 is a 25-item interview designed to screen for a number of deficits related to various executive dysfunctions (Royall et al., 1992), whereas other EF measures only assess one construct (e.g., response inhibition). The EXIT25 includes assessment of frontal release signs associated with severe cognitive impairment (e.g., grasp reflex, utilization behavior, snout reflex), as well as less severe executive dysfunction, such as design and verbal fluency, verbal set-switching, distraction, interference, novel motor programming, disinhibition (i.e., go-no go task), and perseveration (Royall et al., 1992). The EXIT25 has been shown to be a valid and reliable measure of EF (Moorhouse, Gorman, & Rockwood, 2009; Royall et al., 1992). EXIT25 scores have also been linked to frontal lesions beyond the effects of overall cognitive impairment (Royall, Rauch, Roman, Cordes, & Polk, 2001), and discriminate between older adults with dementia

and older adult controls after controlling for overall cognitive functioning (Stokholm, Vogel, Gade, & Waldemar, 2005). Collectively, these results provide evidence that the EXIT25 is a valid assessment of executive dysfunction specifically.

Goals in the creation of the EXIT25 were to provide an interview that all health professionals could use, not only those with extensive psychological training, and to create a brief but effective assessment of executive dysfunction. In the original validation study, the authors reported that completion of all 25 items took approximately 10 minutes (Royall et al., 1992). However, other authors have reported that the EXIT25 takes nearly 15 minutes to complete (Moorehouse et al., 2009). Because U.S. primary care physicians have approximately 18 minutes for a routine visit with a patient (Konrad et al., 2010), utilizing the EXIT25 as a screening measure may be an inefficient use of health professionals' time. Thus, creating brief measures to effectively assess critical components of aging, such as EF, is essential.

In line with this goal, researchers have applied a Rasch model to analyze the EXIT25 (Larson & Heinemann, 2010). They demonstrated that the test could be reduced from 25 to 14 items (i.e., the Quick EXIT) while maintaining internal consistency and convergent validity with other cognitive function indices. However, Rasch analyses compare difficulties across items on a measure assuming that all items are identically related to the underlying construct (Thissen & Steinberg, 2009), which is improbable for a measure such as the EXIT25, and the primary goal of Rasch analyses is to identify a model that relies solely on the difficulty of the item and the ability of each person in the sample to determine the items that best identify differences in respondents' ability on a measure (Thissen & Steinberg, 2009). Perhaps most importantly, Rasch models assume that specific objectivity is required for accurate measurement, meaning that only the differences between respondents and items influence any observation (Thissen & Steinberg, 2009). Because it is unlikely that the criteria of specific objectivity could be met and also unlikely that all EXIT25 items are identically related to EF, the current analyses employ a graded response model to determine the best items on the EXIT25 (Samejima, 1969). This model is an item response theory model not in the Rasch family that was developed for categorical polytomous data (Samejima, 1969). The application of this model will either provide converging evidence to support the Quick EXIT (Larson & Heinemann, 2010) or may suggest an alternative brief interview format for the EXIT25.

The current study aimed to examine psychometric properties of the EXIT25 (Royall et al., 1992), with a specific focus on applying a graded response model (Samejima, 1969) to the items in the interview to determine if a briefer version of the measure can comprehensively assess executive dysfunction in rural-dwelling adults and older adults.

Methods

Participants and Procedures

Participants were 660 rural-dwelling adults ages 40 to 96 living in West Texas. All participants were recruited through Project FRONTIER (Facing Rural Obstacles to healthcare Now Through Intervention, Education & Research). Project FRONTIER is an

ongoing epidemiological longitudinal study that examines a number of factors related to health and aging in rural settings. Any individual living within the rural catchment area who is age 40 or older is eligible for inclusion in the study. This project utilizes community-based participatory research strategies to recruit participants and manage aspects of the study. More information about the recruitment procedures is available in previous publications (O'Bryant, Schrimsher, Johnson, & Zhang, 2011a). The participants recruited into the study generally resemble the makeup of the county in terms of age, gender, race, and educational level (O'Bryant, Edwards, Menon, Gong, & Barber, 2011). The community, as well as the study cohort, is primarily non-Hispanic White, with a significant Hispanic representation as well. Most are female, high school graduates, and middle-aged; please see O'Bryant et al. (2011b) for additional details. The measures utilized in the current study were administered as part of a comprehensive examination that includes a physical exam by a health professional, bloodwork, neuropsychological battery, and history interview. The neuropsychological battery was completed by a research assistant who was blind to any cognitive diagnoses of participants. All study-related procedures for Project FRONTIER have been approved by the Texas Tech University Health Sciences Center Institutional Review Board.

All Project FRONTIER participants who provided complete EXIT25 data, and had been determined to be cognitively healthy, or were diagnosed with dementia or mild cognitive impairment (MCI) were included in this study. The only exclusion criteria were incomplete EXIT25 data or other cognitive diagnosis (i.e., cognitive impairment no dementia or age associated cognitive impairment). Of the total sample, 560 participants had complete data that was utilized in the IRT analyses; many of the participants were missing only one or two items on the EXIT25, but were excluded from this analysis if any data were missing. We conducted a series of *t*-tests to examine potential differences between participants who were included in analyses (i.e., those with complete EXIT25 data) and those who were excluded (i.e., those with missing EXIT25 data). The only significant difference was in level of education ($t = -3.12$, $p = .002$), with excluded participants reporting lower education ($M = 9.32$, $SD = 4.76$) than participants with complete data ($M = 10.92$, $SD = 4.26$). However, the mean for both groups reflected that some high school was completed. The full sample was utilized for the validity analyses (i.e., ANOVAs, regressions, and ROC curve analyses).

Diagnosis

Cognitive diagnoses were established through weekly consensus review meetings, which consisted of a physician, neuropsychologist, and psychologist. The full cognitive testing battery included the Mini Mental State Examination (MMSE), Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) form A, American National Adult Reading Test (AMNART), Brief Smell Identification Test (BSIT), Trail Making Test, FAS, Animal naming, EXIT25, CLOX, Boston Naming Test, and grip strength. Mental health was assessed using the 30-item Geriatric Depression Scale (GDS), Beck Anxiety Inventory, and Alcohol Use Disorders Identification Test (AUDIT). A structured interview was conducted with an informant (in-person or via telephone) to assess activities of daily living (e.g. managing checkbook, driving, shopping, household chores, cooking), which included the completion of the Clinical Dementia Rating scale (CDR). Medical examinations included a

review of systems, neurological exam, Hachinski Ischemic Index Scale as well as clinical labs (i.e., lipid panel, Complete Blood Count, HbA1c, thyroid levels, comprehensive metabolic panel, fasting glucose, B12 and gamma-glutamyl transferase). All information was taken into account when assigning diagnoses of possible/probable Alzheimer's disease (based on criteria McKhann et al., 1984 criteria), vascular dementia (based on Román et al., 1993 criteria), "other" dementia (i.e. do not meet criteria for AD or VaD), mild cognitive impairment (based on Winblad et al., 2004 criteria), cognitive impairment no dementia (CIND; included life-long impairments such as mental retardation rather than recently developed cognitive loss), age associated cognitive impairment (AACI, cognitively normal with complaints of impairment), or cognitively normal. In this study, participants were included if they were diagnosed with dementia ($n = 15$) or MCI ($n = 109$), or were identified as cognitively healthy ($n = 420$). Within the dementia diagnostic category, seven participants were diagnosed with probable dementia due to Alzheimer's disease, two were diagnosed with possible dementia due to Alzheimer's disease, three were diagnosed with mixed-type dementia, and three were diagnosed with dementia not otherwise specified.

Measures

Executive Interview (EXIT25; Royall et al., 1992). A brief explanation of the structure of the EXIT25 was provided earlier. The EXIT25 has been shown to have excellent internal consistency reliability, interrater reliability, and convergent validity (Royall et al., 1992). In the original validation sample, a cutoff score of 15 (on a scale of 0 to 50, with increasing scores indicating greater executive dysfunction) was recommended as a way to separate older adults with disruptive behaviors related to poor EF from those with appropriate behaviors.

Trail Making Test (TMT; Reitan, 1955). The TMT is a task measuring cognitive flexibility in which the participant draws a line through marked circles on a page, either through consecutive numbers (TMT-A) or switching between consecutive numbers and letters (e.g., 1-A-2-B; TMT-B). Differences in the amount of time participants require to complete parts A and B can be indicative of difficulty with response inhibition and executive control. This measure was included to assess convergent validity of the brief form of the EXIT25. This measure was chosen because it is a well-established measure of EF (Arbuthnott & Frank, 2000) and does not overlap with the tasks on the EXIT25, as the EXIT25 does not include a written set-switching task. Therefore, it ensures that associations identified are not simply the result of task overlap, and instead truly represent conceptual associations between various measures of EF.

Controlled Oral Word Association Test (FAS; Strauss, Sherman, & Spreen, 2006). FAS is a test of verbal fluency, which is a component of executive functioning. This test requires participants to generate as many words as possible that begin with a specified letter (i.e., F, A, and S), with a time limit of 60 seconds per letter. This measure was also utilized to assess convergent validity of the brief EXIT25. This measure was chosen because it assesses an important component of executive functioning, and it overlaps but expands upon a task in the EXIT25. This allowed us to assess for overlap between the measures and, in conjunction

with the TMT, assess for whether this overlap was due to task similarity or truly assessing EF as it is conceptualized (i.e., as a set of complex processes).

Mini Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975). The MMSE is a validated interview that assesses cognitive impairment on a 30-point scale. MMSE items assess orientation, registration, attention, calculation, recall, and language to provide a measure of overall cognitive functioning. This measure was utilized to assess discriminant validity of the brief EXIT25.

Data Analysis

To ensure that the correct item response theory model was employed, an exploratory factor analysis was conducted. If the EXIT25 has multiple factors, it is important to conduct an analysis that assesses how items relate to each component of EF, whereas finding a unidimensional structure would suggest that all items could be examined as they relate to one single underlying construct of EF. An unweighted least squares extraction with direct oblimin rotation was run in IBM SPSS Statistics 19.0 (SPSS Inc., 2010), and eigenvalues and the scree plot were examined to assess factors; additionally, a parallel analysis (O'Connor, 2000) was also used to examine the number of factors to retain.

For the IRT analysis, Samejima's (1969) graded response model (GRM) was applied to the data using the ltm package (Rizopoulos, 2006) for R (R Development Core Team, 2006). Item Response Category Characteristic Curves (ICCs) were then created for each of the EXIT25 items. The ICC is a graph of the likelihood of a response being endorsed across the entire range of ability levels (theta). These graphs can be used to look for desirable distinct shifts between response options (i.e., visible peaks for each response option that are not overlapped by the other response curves). Discrimination estimates were also used to identify items that are most central to the underlying construct and distinguish between varying levels of the construct. Higher discrimination indicates that the item distinguishes between respondents with varying levels of the underlying construct (in this case, EF), whereas discrimination close to zero indicate that the item does not differentiate among levels of the construct. Because items with high discrimination estimates provide more information, they are considered better items.

After identifying the best items based on ICCs and discrimination scores, the Item Information Curves (IICs) were created to examine for assessment of the full range of the underlying construct. Overlap among the IICs indicates that, while the items are each good at discriminating between varying levels of the underlying construct, they provide similar information about the underlying construct. Items that significantly overlap with others can be removed for efficiency while still ensuring that the full range of the construct is measured.

Once the optimal items were identified, they were validated as effective measures of EF. We validated these items in a number of ways using IBM SPSS Statistics 19.0 (SPSS Inc., 2010). First, we conducted a one-way analysis of variance (ANOVA) to test whether total scores on the identified items differentiated between cognitively healthy participants and those with cognitive impairment (i.e., MCI and dementia).

Second, we conducted regression analyses with the sum of subsets of the EXIT25 items as the predictor variable (i.e., the selected items in the first block, the non-selected items in the second block), and TMT scores and FAS scores as the criterion measures of EF. Total number of words generated across categories was used for FAS. For the TMT, we used a difference score, in which the number of seconds to complete Part A was subtracted from number of seconds for Part B (B - A). Because this variable evidenced significant kurtosis, we transformed it using a log transformation. We also controlled for age in these analyses, as age is associated with changes in cognition. Each of the analyses examining relations between the EXIT items and other measures of EF was conducted using hierarchical linear regression.

To ensure that the items identified by the IRT analysis had significant incremental validity in predicting other measures of executive functioning beyond the effects of overall cognitive functioning, two additional hierarchical linear regression analyses were conducted, with FAS scores and TMT scores as the criterion variables. In the first step of each analysis, age was controlled and MMSE scores were entered; scores on the IRT-selected EXIT25 items were entered in the second step to examine whether they were a significant predictor of other measures of executive functioning after controlling for variance accounted for by overall cognitive functioning.

An internal consistency reliability analysis was also conducted to assess differences in internal reliability between the set of selected items and the 25 items that comprise the EXIT25; additionally, the correlation between the final set of selected items and the whole EXIT25 was examined. Finally, with a reduced number of items, it is critical to identify a cutoff score that discriminates between those with and without significant executive dysfunction. Therefore, we conducted two receiver operating characteristic (ROC) curve analyses with a total score on the selected items as the test variable, and a state variable of MCI versus no diagnosis in the first analysis, and dementia versus no diagnosis in the second analysis (see Figure 5). ROC curves seek to maximize the sensitivity and specificity of scores. We also used FAS scores and TMT scores in ROC curve analyses to compare their ability to distinguish cognitive diagnoses with the selected EXIT25 items' ability to distinguish cognitive diagnoses.

Results

Participants included 448 women (68.1% of sample). Mean age of the sample was 61.25 (*SD* = 12.59) and average level of education was 10.68 years (*SD* = 4.37). Average IQ, as measured by the American Version of the National Adult Reading Test, range from 84 to 129 (*M* = 107.59, *SD* = 10.38). The breakdown of race in the sample was: 91.5% White, 4.1% African-American, 2.7% American Indian/Alaskan Native, 2.4% Other, and 0.5% Asian. Participants were able to endorse more than one race, so percentages total more than 100%. Approximately 45% of the sample identified as Hispanic. Of those who reported information on employment based on Hollingshead codes, 43.8% were unemployed or retired, 7.4% were farm or service workers, 13.3% were other unskilled workers, 11.0% were semi-skilled workers, 6.1% were craftsmen or other workers, 5.2% were in clerical/ sales or were small farm owners, 3.6% were semiprofessionals or technicians, 3.1% were

managers, larger farm owners, small business owners, or other professionals, and 6.5% were administrators, higher level professionals, or medium-size business owners.

See Table 1 for demographic characteristics of the sample. The exploratory factor analysis indicated that the EXIT25 is a unidimensional measure. An examination of the scree plot indicated a substantial decrease in the eigenvalues of factors after the first factor. The eigenvalue of the first factor was 3.81 and it accounted for 15.24% of the variance in EXIT25 scores. Eigenvalues for the next seven factors ranged from 1.08 to 1.69, and accounted for no more than 6.76% of the variance. The ratio of the first eigenvalue to the second was 2.25, which exceeds the less stringent criterion of 2 (Lord, 1980) but fails to meet more stringent criteria (i.e., 4; Reeve et al., 2007). Additionally, the parallel analysis suggested that eight factors should be retained, as eigenvalues for eight factors exceeded the 95th percentile random data eigenvalue. However, O'Connor (2000) noted that parallel analysis for factor analysis purposes (as opposed to principal components analysis purposes) can indicate that more factors should be retained than are pragmatically appropriate, and other procedures should be used to remove trivial factors. We examined the factor loadings for the eight suggested factors and found that factor eight had no items that strongly loaded onto it (i.e., factor loading above 0.3). Factors three through seven had only one item that loaded onto each of them. Thus, all of these factors were discarded. When examining a two-factor model, the first factor had 12 items that loaded onto it and the second factor had three items that loaded onto it, two of which also cross-loaded onto the first factor. Therefore, since only one item loaded solely on the second factor, we also discarded this factor. While, the parallel analysis and Kaiser's (1960) rule regarding eigenvalues greater than one both suggested that eight factors be retained, these factors are uninterpretable and impractical. Collectively, the Scree plot, ratio of eigenvalues, and factor loadings suggested that one factor is most appropriate; as a result, one factor was retained.

Therefore, a graded response model (GRM; Samejima, 1969) was applied to the data. As a first step, goodness of fit was assessed for constrained and unconstrained models. In an item response theory analysis, a constrained model assumes that discrimination parameters across all items are equal (i.e., a Rasch-family analysis), whereas an unconstrained model allows the discrimination parameter to vary across items. A likelihood ratio test revealed that the unconstrained GRM was preferable to the constrained model. Bayesian Information Criterion (BIC) for the unconstrained model was 14558.60; for the constrained model, BIC was 14679.90 ($p < .001$).

Using the unconstrained model, ICCs were used as a visual representation of the response options for each item, and discrimination estimates were also examined. After identifying the best items based on ICCs and discrimination estimates, the IICs were examined for overlap, to screen for items that could be removed for efficiency of administration. See Figure 1 for ICCs for included items, Figure 2 for ICCs for excluded items, Figure 3 for IICs for included items, Figure 4 for the test information function for the full EXIT25, and Figure 5 for the test information function for included items. Using these criteria, eight items were identified as the best assessors of the underlying construct of EF and best discriminators between levels of EF. Additionally, these items assessed a full range of the latent trait without significant overlap. Table 2 lists discrimination estimates and extremity parameters

for each item; all items chosen had a discrimination estimate above 1.40. The eight items selected, which we refer to as the EXIT8, yielded an average discrimination estimate of 1.64, whereas the seventeen remaining items yielded an average discrimination estimate of . 70. (Figures 1 and 2 include all ICCs.) As a note to readers, the extremity parameters, while theoretically similar to difficulty parameters, are not directly interpretable as difficulty parameters.

A preliminary evaluation of the information provided by the eight items selected (i.e., Number-letter task, Word fluency, Design fluency, Anomalous sentence repetition, Go/no go task, Luria hand sequence I, Luria hand sequence II, and Serial order reversal task) was 21.11. Information in all 25 items was 36.70. Therefore, 57.4% of the information in the full EXIT25 can be gathered using the selected eight items. Conversely, in the 17 items we discarded, only 42.5% of the total information was obtained.

The results of the ANOVA were significant, $F(2, 524) = 84.96$, $p < .001$, suggesting that participants who had been diagnosed with different categories of cognitive disorders yielded different EXIT8 scores. In addition, the effect size was large (Partial $\eta^2 = .245$). Furthermore, post-hoc Tukey analyses revealed significant differences in the expected directions in EXIT8 scores between cognitive healthy participants and those with MCI ($p < .001$), as well as those with dementia ($p < .001$). In addition, there were significant differences between participants diagnosed with MCI and dementia ($p = .003$). Details of these comparisons are presented in Figure 6.

In the regression analyses, EXIT8 scores were a significant predictor of FAS scores after controlling for age (see Table 3). Importantly, the non-selected EXIT items did not add significant predictive power beyond the EXIT8 items. When TMT scores, which were log-transformed as they were non-normally distributed, were entered as the criterion variable, EXIT8 scores were again a significant predictor, after controlling for age. However, the non-selected items were also a significant predictor after controlling for variance associated with EXIT8 scores (see Table 3).

When examining the incremental validity of EXIT8 scores beyond effects of overall cognitive functioning (i.e., MMSE scores) and age on other measures of EF, results also supported the EXIT8 as a strong measure of EF (see Table 4). MMSE scores were a significant predictor of FAS scores, and EXIT8 scores were also a significant predictor after controlling for MMSE scores. The incremental validity of EXIT8 scores was also examined with log-transformed TMT scores entered as the criterion variable. Significant results were found for both MMSE scores, though EXIT8 scores only approached significance in this analysis.

In the internal consistency reliability analysis, Cronbach's alpha was expected to be relatively high, as all items measure EF and therefore should correlate with each other, but was not expected to be excellent as items on the EXIT25 assess different aspects of EF and thus would not be expected to be perfectly reliable with one another. For the full EXIT25, Cronbach's alpha was .64. For the EXIT8, Cronbach's alpha as .74. This indicates that reliability between items on the scale increased with fewer items, suggesting that some

excluded items may not have been internally consistent with other items. The bivariate correlation between the EXIT25 and the EXIT8 was .82, *p* < .001.

The area under the first ROC curve, which examined EXIT8 scores predicting MCI diagnosis, was .81 and suggests that EXIT8 scores predict MCI significantly better than chance alone (asymptotic significance < .001). Both sensitivity (.81) and specificity (.65) are maximized at a score of 2.50 on the EXIT8. At this score, 80.8% of participants with a cognitive disorder are correctly identified as having mild cognitive impairments, and 65.2% of participants without this diagnosis would be correctly identified as cognitively healthy. See Figure 7 for this ROC curve. When FAS total scores were used as a predictor for MCI diagnosis, the area under the curve was only .69, though it was still better than chance (asymptotic significance < .001). Similarly, when TMT scores were used, the prediction was better than chance (asymptotic significance < .001) but was only .69.

The second set of ROC curve analyses yielded similar results. The first, which used EXIT8 scores to predict dementia diagnosis, identified a cutoff score of 2.50 maximizing sensitivity and specificity. The area under this curve was .92 (asymptotic significance < .001), indicating that EXIT8 scores predict dementia diagnosis better than chance alone. Sensitivity at this score was 1.00, indicating all participants with dementia are correctly identified, and specificity was also good (.65). At this cutoff score, 65.2% of cognitive healthy older adults would be identified as such. See Figure 7 for the ROC curve. When FAS scores were used, the area under the curve was smaller, at .72 (asymptotic significance = .01), and when TMT scores were used, the area under the curve was .69 (asymptotic significance = .02).

Discussion

Our results indicate that eight items from the EXIT25 drive most of the information in scores on the measure, are most closely related to the underlying construct of EF, and best discriminate between varying levels of EF. These eight items, which we refer to as the EXIT8, are correlated with other measures of EF, even beyond the effects of general cognition and age, and can differentiate between cognitively healthy participants and those with cognitive impairment. These eight items primary measure set-switching, mental flexibility, fluency, novel motor programming, and response inhibition. Collectively, these results indicate that the EXIT8 is an effective brief measure of EF. A cutoff score of 2.50 is recommended to differentiate between cognitively healthy adults and those with either MCI or dementia. The EXIT8 appears to be better at differentiating these groups than other measures of EF.

When comparing the items identified by our analyses to those from previous analyses (Larson & Heinemann, 2010), we found that the items identified in the EXIT8 are all contained in the 14-item set known as the Quick EXIT (Larson & Heinemann, 2010). This provides more evidence that these items are best for discriminating between varying levels of executive dysfunction, as two different types of analyses conducted in independent samples have indicated that these are the optimal items. Given that health professionals have limited time with patients (Konrad et al., 2010), the EXIT8 may be an ideal screening tool

for executive dysfunction. We estimate that the EXIT8 would take no more than five minutes to administer and score, making it a viable tool in busy medical practices. Health professionals could easily screen for executive dysfunction and refer out for further evaluation if patients score above the recommended cutoff score.

Previous research indicates that EF is associated with dementia (Elliott, 2003), and the findings of the current study lend support to this link, as participants with dementia scored more poorly than participants with MCI or cognitively healthy participants on the EXIT8. Studies also suggest that EF is comprised of various cognitive processes (Elliott, 2003; Phillips & Henry, 2008), though our results suggest that these processes should be examined as a unidimensional construct. Finally, the EXIT8 may provide an ideal brief screening of EF for medical practices, which is critical because poor EF has been associated with poorer outcomes in various domains (Pereira et al., 2008; Von Hippel et al., 2009), and annual screening may be beneficial in reducing poor outcomes for patients.

The primary limitations in this study were sample characteristics. While the sample was large and diverse in terms of age, it was from a geographically limited area and was primarily composed of white and Hispanic participants. The mean education was low, and there were significant differences in education between included and excluded participants, limiting the generalizability of our results to adults and older adults of other education levels. In addition, there were relatively few older adults with cognitive diagnoses, meaning that participants were generally cognitively healthy; the group diagnosed with dementia was small. With a higher base rate of cognitive diagnoses, it is possible that our items would better classify adults with various cognitive diagnoses. Given this limitation, these results should be replicated in samples with more cognitive dysfunction, varying races, and a range of education levels, as well as from other areas to ensure that our results generalize to adults and older adults outside our sample. The overlap with items identified previously (Larson & Heinemann, 2010) begins to address this limitation, but more research is needed. Additionally, as executive functioning includes a variety of components, results may be slightly different if other executive functioning measures were used to validate the EXIT8. Therefore, future research should include other executive assessments to examine the validity of the EXIT8.

Conclusion

Applying a graded response model (Samejima, 1969) to the EXIT25 (Royall et al., 1992) identified eight items that were optimally related to the underlying construct of EF and discriminated between levels of EF well. The EXIT8 may be an effective and efficient measure of EF that can be used in a variety of settings without undue burden on health professionals conducting the interview or older adults being screened, though additional validation of the measure is needed.

Acknowledgments

References

Angel L, Fay S, Bouazzaoui B, Isingrini M. Two hemispheres for better memory in old age: Role of executive functioning. Journal of Cognitive Neuroscience. 2011; 23:3767–3777. doi:10.1162/jocn_a_00104. [PubMed: 21812559]

Arbuthnott K, Frank J. Trail Making Test, Part B as a measure of executive control: Validation using a set-switching paradigm. Journal of Clinical and Experimental Neuropsychology. 2000; 22:518–528. doi:10.1076/1380-3395(200008)22:4;1-0;FT518. [PubMed: 10923061]

de Lucena Ferretti RE, Jacob-Filho W, Grinberg LT, Paraizo Leite RE, Farfel JM, Suemoto CK, Nitrini R. Morphometric brain changes during aging: Results from a Brazilian necropsy sample. Dementia & Neuropsychologia. 2010; 4:332–337. Retrieved from http://www.demneuropsy.com.br/conteudo.asp?pag=1.

Elliott R. Executive functions and their disorders: Imaging in clinical neuroscience. British Medical Bulletin. 2003; 65:49–59. doi:10.1093/bmb/65.1.49. [PubMed: 12697616]

Ettenhofer ML, Hinkin CH, Castellon SA, Durvasula R, Ullman J, Lam M, Foley J. Aging, neurocognition, and medication adherence in HIV infection. American Journal of Geriatric Psychiatry. 2009; 17:281–290. doi:10.1097/JGP.0b013e31819431bd. [PubMed: 19307857]

Folstein MF, Folstein SE, McHugh PR. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research. 1975; 12:189–198. doi:10.1016/0022-3956(75)90026-6. [PubMed: 1202204]

Kaiser HF. The application of electronic computers to factor analysis. Educational and Psychological Measurement. 1960; 20:141–151. doi:10.1177/001316446002000116.

Konrad TR, Link CL, Shackelton RJ, Marceau LD, von dem Knesebeck O, Siegrist J, McKinlay JB. It's about time: Physicians' perceptions of time constraints in primary care medical practice in three national healthcare systems. Medical Care. 2010; 48:95–100. Retrieved from http://www.lww-medicalcare.com. [PubMed: 20057331]

Larson EB, Heinemann AW. Rasch analysis of the Executive Function Interview (The EXIT-25) and introduction of an abridged version (The Quick EXIT). Archives of Physical Medicine and Rehabilitation. 2010; 91:389–394. doi:10.1016/j.apmr.2009.11.015. [PubMed: 20298829]

Lord, FM. Applications of item response theory to practical testing problems. Lawrence Erlbaum; Hillsdale, NJ: 1980.

McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology. 1984; 34:939–944. doi:10.1212/WNL.34.7.939. [PubMed: 6610841]

Moorhouse P, Gorman M, Rockwood K. Comparison of EXIT-25 and the Frontal Assessment Battery for evaluation of executive dysfunction in patients attending a memory clinic. Dementia and Geriatric Cognitive Disorders. 2009; 27:424–428. doi:10.1159/000212755. [PubMed: 19372680]

Nguyen HT, Grzywacz JG, Arcury TA, Chapman C, Kirk JK, Ip EH, Quandt SA. Linking glycemic control and executive function in rural older adults with diabetes mellitus. Journal of the American Geriatrics Society. 2010; 58:1123–1127. doi:10.1111/j.1532-5415.2010.02857.x. [PubMed: 20722846]

O'Bryant SE, Edwards M, Menon CV, Gong G, Barber R. Long-term low-level arsenic exposure is associated with poorer neuropsychological functioning: A Project FRONTIER study. International Journal of Environmental Research and Public Health. 2011b; 8:861–874. doi:10.3390/ijerph8030861. [PubMed: 21556183]

O'Bryant, SE.; Schrimsher, GW.; Johnson, LA.; Zhang, Y. The impact of sociodemographic factors on neuropsychological test performance in a rural sample: A Project FRONTIER study.. Journal of Rural Community Psychology, E14(1). 2011a. Retrieved from http://www.marshall.edu/jrcp/VE%2014%20N%201/JRCP%20Obryant%2014.1%20read y.pdf

O'Connor BP. SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. Behavior Research Methods, Instrumentation, and Computers. 2000; 32:396–402. doi:10.3758/BF03200807.

Parks CM, Iosif A-M, Farias S, Reed B, Mungas D, DeCarli C. Executive function mediates effects of white matter hyperintensities on episodic memory. Neuropsychologia. 2011; 49:2817–2824. doi: 10.1016/j.neuropsychologia.2011.06.003. [PubMed: 21689669]

Pereira FS, Yassuda MS, Oliveira AM, Forlenza OV. Executive dysfunction correlates with impaired functional status in older adults with varying degrees of cognitive impairment. International Psychogeriatrics. 2008; 20:1104–1115. doi:10.1017/S1041610208007631. [PubMed: 18752698]

Phillips, LH.; Henry, JD. Adult aging and executive functioning.. In: Anderson, V.; Jacobs, R.; Anderson, PJ., editors. Executive functions and the frontal lobes: A lifespan perspective. Taylor & Francis; Philadelphia, PA: 2008. p. 57-79.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2006.

Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, PROMIS Cooperative Group. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Medical Care. 2007; 45:S22–31. doi:10.1097/01.mlr.0000250483.85507.04. [PubMed: 17443115]

Reitan RM. The relation of the Trail Making Test to organic brain damage. Journal of Consulting Psychology. 1955; 19:393–394. doi:10.1037/h0044509. [PubMed: 13263471]

Román GC, Tatemichi TK, Erkinjuntti T, Cummings JL, Masdeu JC, Garcia JH, Scheinberg P. Vascular dementia. Diagnostic criteria for research studies: Report of the NINDS-AIREN International Workshop. Neurology. 1993; 43:250–260. doi:10.1212/WNL.43.2.250. [PubMed: 8094895]

Royall DR, Mahurin RK, Gray KF. Bedside assessment of executive cognitive impairment: The Executive Interview. Journal of the American Geriatrics Society. 1992; 40:1221–1226. Retrieved from http://www.wiley.com/bw/journal.asp?ref=0002-8614. [PubMed: 1447438]

Royall DR, Rauch R, Roman GC, Cordes JA, Polk MJ. Frontal MRI findings associated with impairment on the Executive Interview (EXIT25). Experimental Aging Research. 2001; 27:293–308. doi:10.1080/03610730109342350. [PubMed: 11681194]

Samejima, F. Estimation of latent ability using a response pattern of graded scores. Psychometrika Monograph Supplement, 17. 1969. Retrieved from http://www.psychometrika.org/journal/online/MN17.pdf

Sonnen JA, Cruz KS, Hemmy LS, Woltjer R, Leverenz JB, Montine KS, Montine TJ. Ecology of the aging human brain. Archives of Neurology. 2011; 68:1049–1056. doi:10.1001/archneurol.2011.157. [PubMed: 21825242]

SPSS Inc.. IBM SPSS Statistics Base 19. 2010. Retrieved from http://www.sussex.ac.uk/its/pdfs/SPSS_Base_19.pdf

Stokholm J, Vogel A, Gade A, Waldemar G. The Executive Interview as a screening test for executive dysfunction in patients with mild dementia. Journal of the American Geriatrics Society. 2005; 53:1577–1581. doi:10.1111/j.1532-5415.2005.53470.x. [PubMed: 16137290]

Strauss, E.; Sherman, EMS.; Spreen, O. A compendium of neuropsychological tests: Administration, norms, and commentary. 3rd ed.. Oxford University Press; Oxford, England: 2006.

Thissen, D.; Steinberg, L. Item response theory.. In: Millsap, RE.; Maydeu-Olivares, A., editors. The SAGE handbook of quantitative methods in psychology. SAGE; Thousand Oaks, CA: 2009. p. 148-177.

Von Hippel W, Ng L, Abbot L, Caldwell S, Gill G, Powell K. Executive functioning and gambling: Performance on the Trail Making Test is associated with gambling problems in older adult gamblers. Aging, Neuropsychology, and Cognition. 2009; 16:654–670. doi:10.1080/13825580902871018.

Winblad B, Palmer K, Kivipelto M, Jelic V, Fratiglioni L, Wahlund L-O, Petersen RC. Mild cognitive impairment–beyond controversies, towards a consensus: Report of the International Working Group on Mild Cognitive Impairment. Journal of Internal Medicine. 2004; 256:240–246. doi:10.1111/j.1365-2796.2004.01380.x. [PubMed: 15324367]
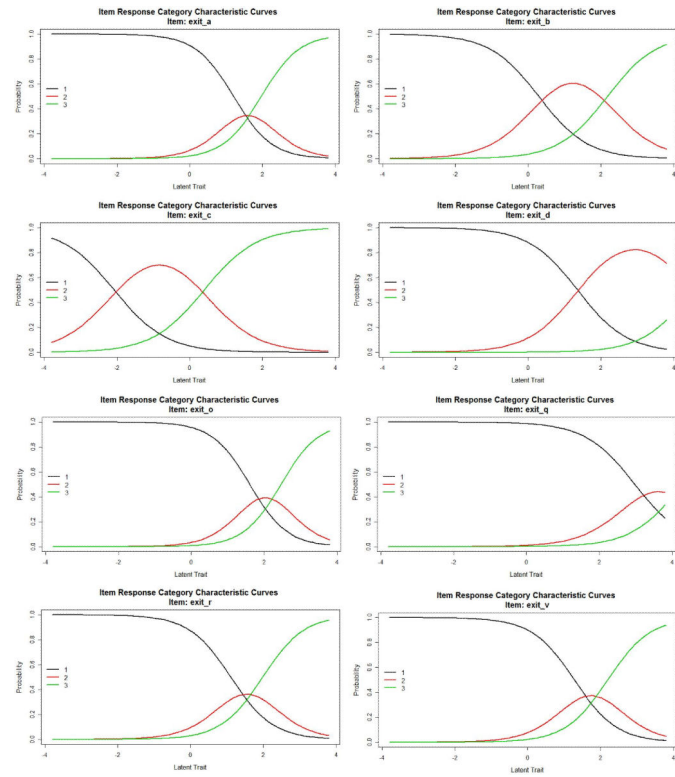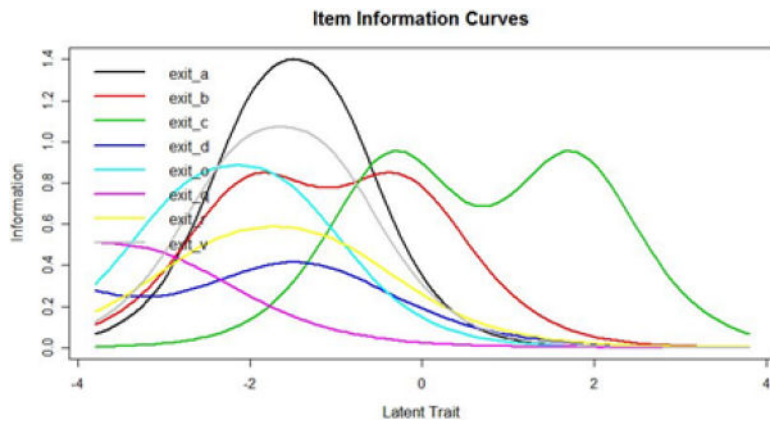
Figure 1.
Item Characteristic Curves for eight Executive Interview items (EXIT8) with distinct
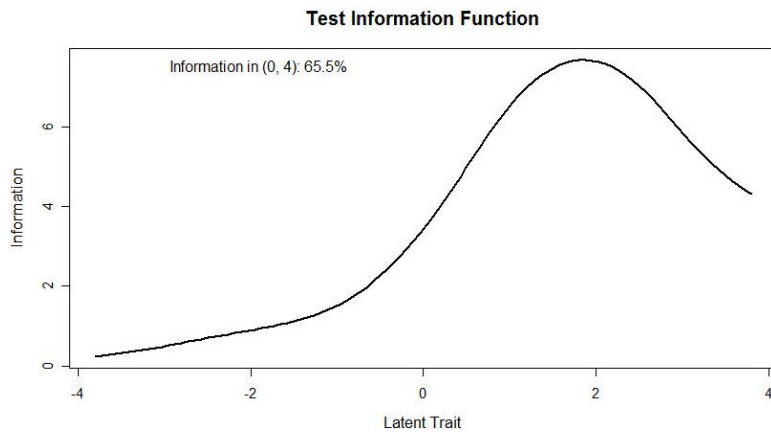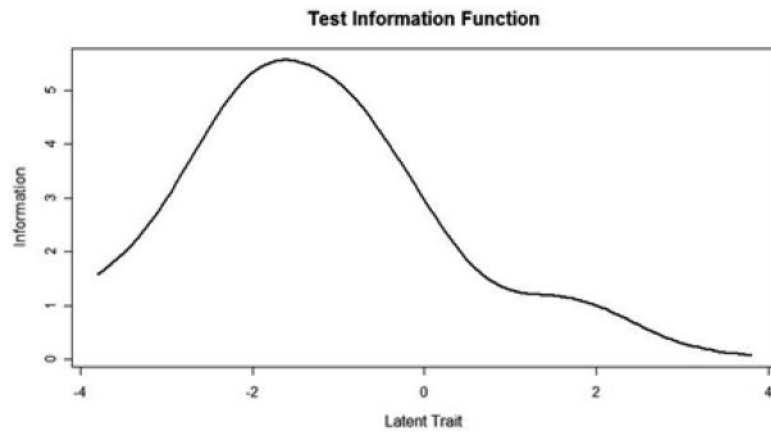response options shifts.
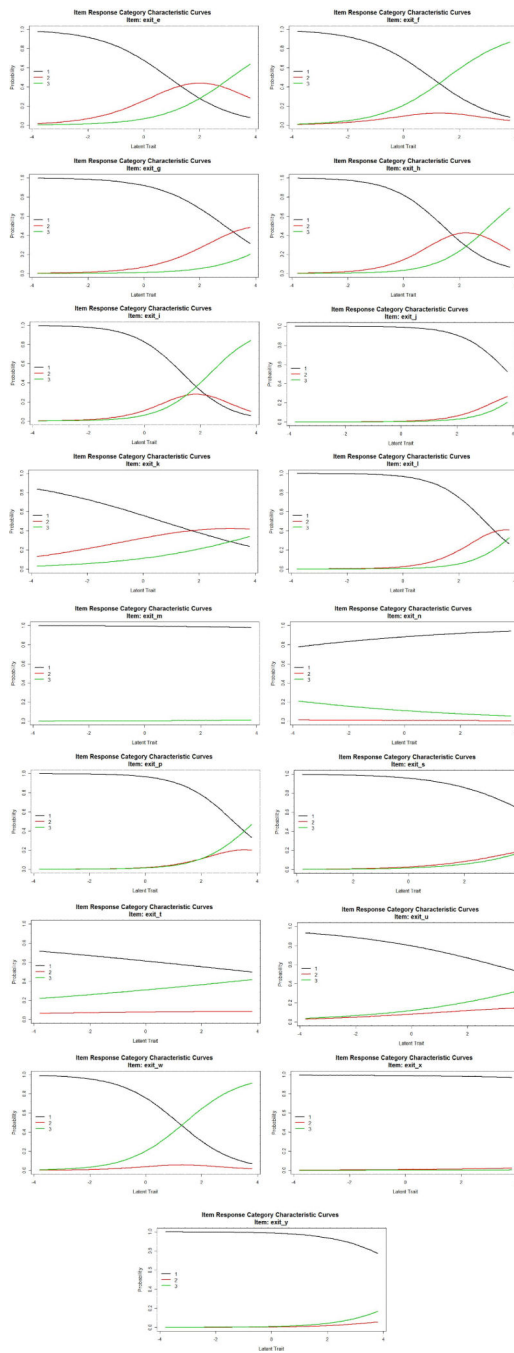
Figure 2.
Item Characteristic Curves for 17 Executive Interview items with less distinct response options shifts.
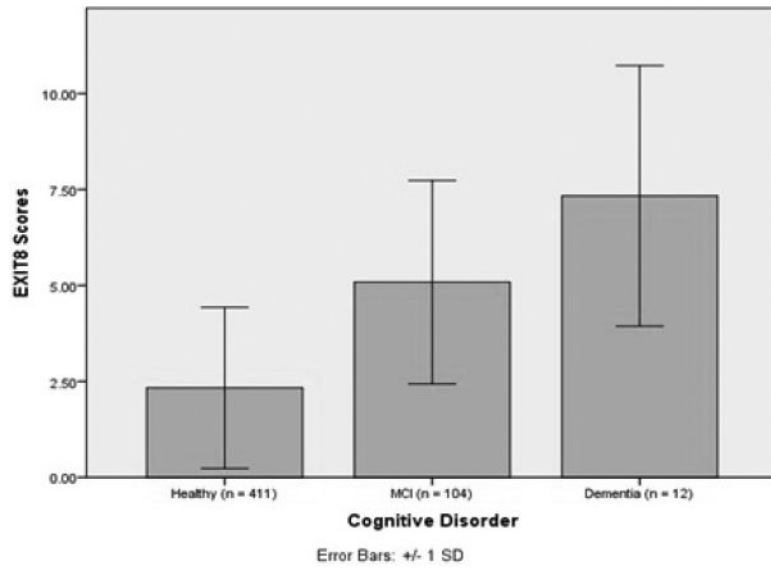
Figure 3.
Item Information Curves for the eight selected Executive Interview (EXIT8) items.

Figure 4.
Test information function for the full Executive Interview.

Figure 5.
Test information function for the eight selected Executive Interview (EXIT8) items.

Figure 6.
Means for eight selected Executive Interview (EXIT8) scores by cognitive diagnosis.
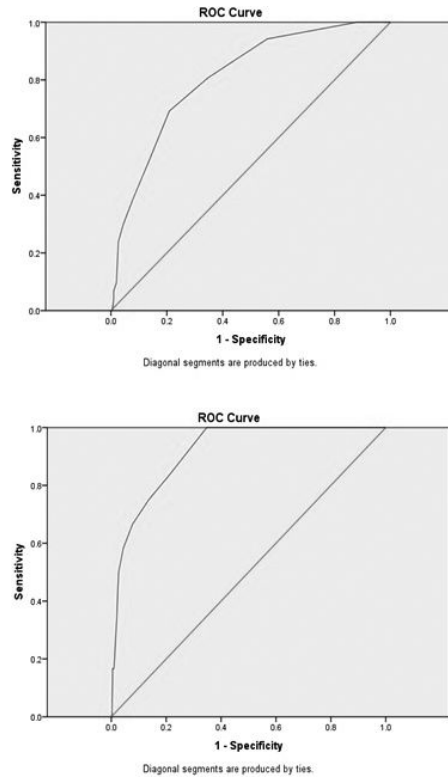
Figure 7.
Receiver operating characteristic curves for the eight selected executive interview (EXIT8) predicting participants with mild cognitive impairment versus no cognitive diagnosis, and predicting participants with dementia versus no cognitive diagnosis.

Table 1

Demographic characteristics

	Total Sample	Cognitively Healthy	Mild Cognitive Impairment	Dementia
Gender (Female)	448 (68.1%)	295 (70.2%)	71 (65.1%)	6 (40.0%)
Age	61.25 (12.59)	59.01 (11.83)	67.34 (12.24)	74.27 (11.80)
Education	10.68 (4.37)	11.25 (4.27)	9.40 (4.15)	9.47 (4.47)
EXIT25	7.44 (4.77)	5.99 (3.84)	11.74 (4.78)	14.75 (5.26)
EXIT8	3.03 (2.55)	2.35 (2.12)	5.06 (2.65)	7.33 (3.39)
FAS	27.85 (12.34)	30.17 (12.07)	22.29 (10.67)	21.57 (11.29)
TMT	74.19 (64.23)	63.85 (56.37)	101.50 (73.88)	126.62 (114.57)
MMSE	27.55 (2.80)	28.32 (2.05)	25.58 (3.07)	21.93 (5.21)

Note: Mean is presented, with standard deviation in parentheses for all variables except gender. For gender, the number of females in each sample is presented, with percentage in parentheses. EXIT25: Executive Interview total scores; EXIT8: Total scores on selected Executive Interview items; FAS: Total number of words on Controlled Oral Word Association Test; TMT: Difference in completion time for Trail Making Test B minus Trail Making Test A; MMSE: Mini Mental State Exam total scores.

Table 2

Discrimination estimates and Extremity Parameters for Executive Interview items

#	Item	Discrimination Estimate	Lower Extremity Parameter	Upper Extremity Parameter
1	**Number-Letter Task**	**1.909**	**1.206**	**1.964**
2	**Word Fluency**	**1.488**	**0.292**	**2.177**
3	**Design Fluency**	**1.403**	**−2.077**	**0.396**
4	**Anomalous Sentence Rep.**	**1.482**	**1.358**	**4.507**
5	Thematic Perception	0.842	0.881	3.126
6	Memory/Distraction Task	0.848	0.970	1.566
7	Interference Task	0.852	2.891	5.413
8	Automatic Behavior I	1.087	1.398	3.079
9	Automatic Behavior II	1.156	1.355	2.353
10	Grasp Reflex	1.218	3.885	4.900
11	Social Habit I	0.369	0.665	5.573
12	Motor Impersistence	1.179	2.930	4.409
13	Snout Reflex	0.326	15.480	17.615
14	Finger-Nose-Finger Task	−0.202	−10.012	−10.343
15	**Go/No-Go Task**	**1.934**	**1.606**	**2.462**
16	Echopraxia	1.088	3.155	3.912
17	**Luria Hand Sequence I**	**1.467**	**2.976**	**4.268**
18	**Luria Hand Sequence II**	**1.737**	**1.106**	**1.983**
19	Grip Task	0.638	4.828	6.390
20	Echopraxia II	0.123	3.727	6.522
21	Complex Command Task	0.328	4.161	6.029
22	**Serial Order Reversal Task**	**1.715**	**1.281**	**2.197**
23	Counting Task I	0.982	1.168	1.396
24	Utilization Behavior	0.268	17.214	23.948
25	Imitation Behavior	0.830	5.281	5.723

Note: While extremity parameters are theoretically similar to difficulty parameters, these numbers cannot be directly interpreted like difficulty parameters.

Table 3

Linear regression results: Regressing executive functioning measures onto Executive Interview items to determine predictive power of selected items

Predictor	*b*	Standard Error	*t*	*p*
Criterion: FAS Total Scores				
Intercept	32.296	2.733	11.816	.000
Age	−.071	.044	−1.626	.105
Intercept	31.852	2.349	13.558	.000
Age	.062	.039	1.596	.111
EXIT8	−2.551	.190	−13.441	.000
Intercept	31.932	2.357	13.550	.000
Age	.065	.039	1.648	.100
EXIT8	−2.509	.208	−12.066	.000
Other EXIT items	−.084	.171	−.492	.623
Criterion: TMT Difference Scores (log transformed)				
Intercept	1.4785	.075	19.648	.000
Age	.004	.001	3.550	.000
Intercept	1.495	.074	20.294	.000
Age	.003	.001	2.086	.038
EXIT8	.032	.007	4.699	.000
Intercept	1.464	.073	20.194	.000
Age	.002	.001	1.651	.099
EXIT8	.018	.007	2.454	.014
Other EXIT items	.024	.005	4.539	.000

Note:

FAS: Controlled Oral Word Association Test; EXIT8: Total score on the eight selected items from the Executive Interview; Other EXIT items: Nonselected 17 items from the Executive Interview; TMT: Trail Making Test difference scores

Table 4

Linear regression results: Regressing executive functioning measures onto EXIT8, controlling for overall cognitive functioning to determine predictive power of selected items

Predictor	*B*	Standard Error	*t*	*p*
Criterion: FAS Total Scores				
Intercept	−33.880	5.964	−5.680	.000
Age	.021	.038	.550	.583
MMSE	2.192	.183	11.988	.000
Intercept	−1.202	7.304	−.165	.869
Age	.069	.037	1.853	.064
MMSE	1.094	.233	4.702	.000
EXIT8	−1.748	.244	−7.154	.000
Criterion: TMT Difference Scores (log transformed)				
Intercept	2.599	.203	12.826	.000
Age	.003	.001	2.722	.007
MMSE	−.038	.006	−5.963	.000
Intercept	2.362	.240	9.828	.000
Age	.003	.001	2.240	.026
MMSE	−.030	.008	−3.826	.000
EXIT8	.015	.008	1.827	.068

Note:

FAS: Controlled Oral Word Association Test; EXIT8: Total score on the eight selected items from the Executive Interview; TMT: Trail Making Test difference score; MMSE: Mini Mental State Exam total score