The Graphic Pattern Generation Test in Healthy Aging and Alzheimer's Disease: Psychometric Properties and Normative Data

Preeti Sunderaraman[a], **Elisaveta Sokolov**[b], **Sarah Cines**[c], **Elizabeth Sullo**[d], **Aidan Orly**[c], **Bianca Lerer**[c], **Jason Karlawish**[d], **Edward Huey**[e], and **Stephanie Cosentino**[f]

[a] Department of Psychology, Drexel University, Philadelphia, Pennsylvania

[b] Department of Medicine, Imperial College London Charing Cross Hospital, London, United Kingdom

[c] Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Medical Center, New York, New York

[d] Departments of Medicine and Medical Ethics, Alzheimer's Disease Center, Institute on Aging, and Center for Health Incentives, University of Pennsylvania, Philadelphia, Pennsylvania

[e] Gertrude H. Sergievsky Center, and Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Medical Center, New York, New York

[f] Taub Institute for Research on Alzheimer's Disease and the Aging Brain, and Gertrude H. Sergievsky Center, Columbia University Medical Center, NewYork, NewYork

Abstract

Design fluency tests, commonly used in both clinical and research contexts to evaluate nonverbal concept generation, have the potential to offer useful information in the differentiation of healthy versus pathological aging. While normative data for older adults are available for multiple timed versions of this test, similar data have been unavailable for a previously published untimed task, the Graphic Pattern Generation Task (GPG). Time constraints common to almost all of the available design fluency tests may cloud interpretation of higher level executive abilities, for example in individuals with slow processing speed. The current study examined the psychometric properties of the GPG and presents normative data in a sample of 167 healthy older adults (OAs) and 110 individuals diagnosed with Alzheimer's disease (AD). Results suggest that a brief version of the GPG can be administered reliably, and that this short form has high test-retest and inter-rater reliability. Number of perseverations was higher in individuals with AD as compared to OAs. A cut-off score of 4 or more perseverations showed a moderate degree of sensitivity (76%) and specificity (37%) in distinguishing individuals with AD and OAs. Finally, perseverations were associated with nonmemory indices, underscoring the nonverbal nature of this error in OAs and individuals with AD.

Address correspondence to: Stephanie Cosentino, Ph.D. Columbia University Medical Center, 630 West 168th Street, P&S Box 16, New York, NY 10032. sc2460@cumc.columbia.edu.

INTRODUCTION

Early deficits in executive functioning characterize a host of neurodegenerative diseases, including Alzheimer's disease (AD; Mez et al., 2012), frontotemporal dementia (FTD; Bozeat, Gregory, Ralph, & Hodges, 2000), Lewy Body dementia (Calderon et al., 2001), and Parkinson's disease (Lewis, Dove, Robbins, Barker, & Owen, 2003). Comprehensive assessment of both verbal and visuospatial executive abilities is important for accurate diagnosis and characterization of cognitive profiles. Design fluency tasks, traditionally considered measures of executive functioning, require individuals to generate a series of unique designs within the confines of specific task parameters (e.g., using a certain number of lines, staying within the allocated design framework). Number of overall designs produced has been associated with the integrity of bilateral prefrontal and parietal regions, as well as the right temporal cortex, striatum, and thalamus, whereas repetitions (or perseverations) on the task appear to have greater selectivity for bilateral orbitofrontal regions (Possin et al., 2012).

Multiple design fluency tasks have been developed over time, generally with the goal of improving the psychometric properties of the prevailing tests and/or improving their sensitivity to mild frontal deficits. The original Design Fluency test (Jones-Gotman & Milner, 1977), developed as a nonverbal analogue to the verbal fluency test, was one of the first neuropsychological tests to demonstrate sensitivity to right frontal lobe pathology. However the relatively weak psychometric properties associated with this test, namely high subjectivity in scoring and low test-retest reliability (for e.g., Goebel, Fischer, Ferstl, & Mehdorn, 2009; Strauss, Sherman, & Spreen, 2006), led to its replacement by other design fluency tests such as the Five-Point test (Regard, Strauss, & Knapp, 1982), Ruff's Figural Fluency Test (RFFT; Ruff, Light, & Evans, 1987), the Graphic Pattern Generation test (GPG; Glosser, Goodglass, & Biber, 1989), the Design Fluency subtest of the Delis-Kaplan Executive Function System battery (DKEFS; Delis, Kaplan, & Kramer, 2001), and more recently by the NEPSY Design Fluency Test (Possin et al., 2013). Among the existing tests, one of the most widely used is the RFFT, a test that has demonstrated increased sensitivity to general frontal functioning as compared to the NEPSY or the D-KEFS fluency tests (Strauss et al., 2006), as well as high inter-rater and test-retest reliability (Strauss et al., 2006). However, like almost all other design fluency tasks it is timed, yielding scores that are in part based on processing speed (Glosser et al., 1989).

Sensitivity to processing speed is a task characteristic that may or may not be beneficial depending on the specific patient and purpose of the evaluation. For example, a person with a history of brain trauma or stroke might have difficulty with initiation or have slowed cognitive processing, both of which would affect performance on a timed task. Similarly, speed based design fluency tasks may not yield useful information regarding executive deficits in patients with movement disorders affecting upper limb dexterity, such as those with Parkinson's disease (Lee et al., 1997) or, in AD patients with diminished motor and processing speed (Kluger et al., 1997; Possin et al., 2013). Thus, it might be argued that the speeded nature of almost all of the available design fluency tests may cloud interpretation of higher level executive abilities such as concept generation and maintenance of mental set, especially for those with movement disorders or for individuals with slow processing speed

(Keil & Kaszniak, 2002). Hence, in order to more purely measure executive functioning in certain populations, it might be advantageous to use a test without time constraints.

The GPG (Glosser & Goodglass, 1990; Glosser et al., 1989) is the only non-speed based test used to assess design fluency. It was developed as a modification of the earlier design fluency tests with the added aim of simplifying the task for the neurologically impaired patient by reducing the complexity of motor responses. Glosser and Goodglass (1990) found the test to be highly sensitive to right frontal lobe pathology in a group of aphasics. Specifically, they found that people with right frontal pathology had significantly more perseverations compared to those with left frontal or right posterior pathology. Moreover, they found this test to be correlated not only with measures of visuoconstruction, but with measures of cognitive flexibility and visuospatial sequencing; highlighting the executive components of the task. The authors further argued that because the GPG did not correlate with language measures, this test might be particularly useful for examining executive functioning in patients with language disturbances. The lengthy nature of this task, however, has limited the extent to which it might be used in clinical or research evaluations.

Individuals with AD and other dementias often present with significant language impairment and/or deficits in motor and processing speed that can cloud assessment of executive abilities when measured with verbal or timed tasks (Kluger et al., 1997; Mez et al., 2012). While existing work has demonstrated impaired performance on the GPG in patients with AD (Budson et al., 2002), information regarding level of performance across different aspects of the task (e.g., perseverations, rule violations, number of unique designs) in AD as well as in healthy older adults has not been available. Moreover, the original version of this test can be quite lengthy and not feasible for use in clinical or research protocols assessing individuals with dementia. In this study, we administered the GPG to 110 individuals with mild to moderate AD and 167 healthy older adults to examine its psychometric properties such as inter-rater, and test-retestreliability. Moreover, we investigated performance differences across the two groups on a brief version of the GPG, and present normative data for this revised task in both healthy older adults and individuals with AD. Finally, we examine the neuropsychological correlates of design fluency in each of these groups.

We hypothesized that number of perseverations on design fluency would be negatively correlated with average verbal fluency given the similar nature of the tasks and their mutual reliance on prefrontal networks important for concept generation (Baldo, Shimamura, Delis, Kramer, & Kaplan, 2001). Furthermore, because the number of unique designs produced on this task is typically inversely related to perseverations, we hypothesized a positive relationship between number of unique designs and average verbal fluency.

METHODS

Participants

The participants were healthy controls and patients with mild to moderate AD recruited at two centers. 110 patients (48 from Columbia University Medical Center (CUMC) and 62 from the University of Pennsylvania's Penn Memory Center) with mild to moderate AD, defined as a score of 17 or greater on the Mini-Mental State Examination were recruited.

Individuals with mild AD were recruited through the CUMC Department of Neurology Memory Disorders Clinic and the University of Pennsylvania's Penn Memory Center, respectively. At both the sites, diagnoses of AD were made according to the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINDS-ADRDA) criteria. All participants provided informed consent and were given monetary compensation (reimbursed $30.00 at CUMC and $ 40.00 at Penn Memory Center). Capacity to consent was initially determined by the referring physician. In addition, the examiner obtaining the consent ensured that the individuals understood the risks and benefits of participating in this study. The Institutional Review Board approved of the study at both the medical centers and all individuals provided informed consent prior to participation.

In addition, altogether 167 healthy older adults (OAs; 68 from CUMC and 99 from University of Pennsylvania) were recruited. At CUMC, OAs were recruited from three sources: the healthy control database available through the CUMC-ADRC, local senior centers, and market mailing procedures. For the Penn Memory Center, patients' family members who had previously agreed to be contacted for research studies were sent a formal letter describing the study. A research assistant then called the contact person, and explained the study in detail. Controls were thoroughly screened by interview to exclude individuals with neurologic, psychiatric, or severe medical disorders. Participants were considered eligible for the study if they were age 55 or above, and scored at least 27 on the Mini Mental State Examination) (MMSE; Folstein, Folstein, & McHugh, 1975; Weintraub & Mesulam, 1985).

Only a subset of the participants from CUMC completed both rows of the design fluency task due to time constraints, and only a subset of them agreed to return for repeat testing. An attempt was made to include neuropsychological data. However, because some neuropsychological tests were not administered at the Penn Memory Center, only data from CUMC was available for certain measures (see Results section). The number of participants in each analysis is indicated in the relevant results table.

Procedures

Participants were seen for neuropsychological testing over the course of two to three, 2-hour test sessions within two weeks. 137 participants (34 from CUMC, and 103 from Penn Memory Center) were re-tested with the GPG one year after the initial testing session. Thirty-five participants from CUMC (20 OAs and 15 ADs) completed the original version of the task consisting of two rows of stimuli and were thus used to determine reliability of a brief version of this test. Tests that were double scored for 88 participants (53 OAs and 35 ADs) were used to determine inter-rater reliability. The initial and double scoring were completed by research assistants with either a Bachelor of Arts (BA) or a Masters of Arts (MA) degree.

Measures

Graphic Pattern Generation (GPG)—The original GPG test is characterized by two separate rows of stimuli. Each row consists of 20 identical dot arrays. The arrays are

composed of five dots, and are different for each row (Figure 1). The test requires participants to generate as many novel designs as they can for each row (administered separately). They are required to do so using exactly four lines to join the dots in each array. The examiner reads the following instructions for the practice trial which consists of ten arrays: "Here is a set of dots (Point to first array of 5 dots.) This set is repeated several times across the page. I want to see how many different designs you can draw using four lines within each set of dots. You may connect these dots in any order you choose by drawing four lines between the dots." The examiner then demonstrates five different approaches to drawing a design and asks the participant to complete the remaining five designs without repeating any, including the ones provided by the examiner. All errors are corrected during practice. The examiner then administers row 1, reminding the participant to try to draw a new design each time and to use four lines to connect the dots. The examiner records time to completion, and then administers the second row (*if required*). Errors are categorized as either rule violations or as perseverations. Instances of rule violations include using more or less than four lines, drawing lines that are not connected to dots, or connecting dots from adjacent arrays. Only the first instance of a rule violation and the first instance of a preservative error are corrected *in each row*. Errors are summed across both rows for a final score. Number of unique designs are calculated by subtracting the sum of these errors (perseveration and rule violations) from 20 (maximum total number of unique designs) for each row.

In addition to these scores, perseverative distance is also calculated. Perseverative distance is defined as the number of items between a perseveration and the most proximal last occurrence of that same design. The distances of all perseverations are summed across both rows, and divided by the total perseveration score. In summary, this test yields five scores including: total time, total rule violations, total perseverations, perseverative distance, and number of unique designs. Scores for individual rows can also be calculated, and were used to examine the reliability of a brief version test.

The GPG scores that differed significantly across the OAs and AD groups were examined in relation to other neuropsychological measures to determine the discriminant and convergent validity of the test. A brief description of these neuropsychological tests is given below:

Visual Scanning

Visual Scanning—This test (Weintraub & Mesulam, 1985) consisted of 60 targets among an array of distractor items spread across an 8.5" X 11" page displayed horizontally. Participants were asked to find and circle all of the targets as quickly as possible. The dependent variable was the total number of targets identified in 60 seconds. Similar to processing speed tests that require visual search such as the Symbol Search and Digit Symbol subtests of the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1997), this visual scanning test is also a timed measure.

Attention and Working Memory

Digit Span—This subtest from the Wechsler Memory Scales - Third edition (WMS-III; Wechsler, 1997) required participants to repeat a series of digits, beginning with only two

and increasing until the participant failed two consecutive items at a given series length. The second part of the test required participants to recite the numbers read aloud by the examiner in the reverse order. The dependent variables were the total raw scores on each of the forward and backward components of the task.

Spatial Span—This WMS-III (Wechsler, 1997) subtest required participants to remember a series of spatial locations on a board, beginning with only two and increasing until the participant failed two consecutive items at a given series length. The second part of the test required participants to recall the locations demonstrated by the examiner in the reverse order. The dependent variables were the raw scores on each of the forward and backward components of the task.

Learning and Memory

Philadelphia Repeatable Verbal Learning Test (PrVLT; Price et al., 2009)—The PrVLT is a list-learning task modeled after the 9-word California Verbal Learning Test (Libon et al., 1996) in which participants are required to learn 9 words (comprising three different categories: fruit, tools, and furniture) over the course of five trials. This is followed by an interference trail, and the short and delayed recalls, respectively. The primary dependent variables included total immediate recall across the 5 learning trials, and short and long delayed recalls and recognition (discriminability index) presented after 20-40 minutes.

Biber Figure Learning Test (Glosser et al., 1989)—This **visuospatial** list learning task consists of 9 black and white geometric designs presented over five trials. Designs were presented one at a time in a fixed order, for three seconds each. During the test phase, participants were asked to draw as many designs as they could remember. This was followed by an interference phase. After a 20 to 30 minute delay, participants were again asked to recall as many designs as possible, and subsequently to copy each of the stimuli to ensure that constructional abilities required for intact performance did not affect memory performance. Each drawing was scored according to strict guidelines on a scale of zero to three. Dependent variables included total immediate recall across the 5 learning trials, short and long delayed recalls and recognition (discriminability index) presented after 20-40 minutes.

Pre-morbid IQ

Wechsler Test of Adult Reading (WTAR; David Wechsler, 2001)—This brief test consists of 50 words that become increasingly complex and have several irregular pronunciations. The WTAR was normed on a large national sample and has shown to possess excellent psychometric properties (Spreen, 1998).

Verbal Fluency

Controlled Word Association Test (COWA; Spreen, 1998)—For this test, participants have to generate as many words as they can beginning with a specific letter in one minute; the letters were F, A and S. The average number of words generated was used as the dependent variable.

Data Analysis

Basic demographic variables including age, education, and gender were compared between the OAs and AD groups using t-tests and chi-square analyses. Age was significantly different between the AD and OA groups for the CUMC site, but not for the University of Pennsylvania's Penn Memory Center site. After statistically controlling for the site, three aspects of reliability were calculated for the GPG including split-half reliability across rows 1 and 2, inter-rater reliability, and test-retest reliability after one year. In order to calculate both inter-rater reliability and test-retest reliability, Pearson's correlations were obtained. To evaluate inter-rater reliability, two trained raters from each site (the examiner and another rater) independently calculated the scores for three GPG variables; perseverations, perseveration distance and rule violations, for 88 participants (53 OAs and 35 ADs). Test-retest reliability was calculated for these GPG variables one year after the initial assessment in a subset of the sample. GPG scores were then compared between the two groups (OAs versus AD) using Analysis of Covariance (ANCOVA) controlling for demographic and site differences. Receiver-operating characteristics (ROC) analysis was used to examine the efficacy of GPG scores in distinguishing individuals with AD from OAs. Finally, we evaluated the extent to which performance on the GPG was correlated with other neuropsychological scores. In case of missing data for the GPG variables and the neuropsychological measures, the cases containing the missing data were deleted, and the analysis was conducted without these values.

Results

Demographic Variables

After controlling for site, age (OA; $\bar{X}$ = 75.70, SD = 8.38, Range = 57-97 years, and AD $\bar{X}$ = 77.57, SD = 7.67, Range = 57-100 years) and education (OA; $\bar{X}$ = 15.99, SD = 2.64 and AD; and AD; $\bar{X}$ = 15.36, SD = 2.70) were compared between OAs and AD groups using ANCOVA, and gender (OA; Males =56 (33.5%), Females = 111 (66.5%) and AD; Males = 47 (42.7%), Females = 63 (57.3%)) was compared across these two groups using a chi-square analysis. Both age (F $(1, 273)$ = 4.59, p = .03) and level of education (F $(1, 267)$ = 3.75, p = 0.05) were found to be significantly different between these groups. Chi square analysis was not significant (χ^2 (df) = 2.40 (1), p = 0.12).

Reliability

Reliability for a brief version of this test—We investigated the relationship between row 1 and row 2 scores on all five GPG variables for participants if they completed the original version of the task with both rows (OAs group, n = 20; AD group, n = 15). The results demonstrated that, for the OA group, the number of perseverations correlated significantly between rows (r = 0.75, p < 0.001), as did time to completion (r = 0.69, p = 0.001) and the number of unique designs (r = 0.77, p < 0.001). This was not the case for perseverative distance (r = −0.05, p > 0.05). For the AD group, the time to completion correlated significantly between the rows (r = 0.90, p < 0.001). This was not the case for perseverations (r = 0.49, p > 0.05), perseverative distance (r = −0.14, p > 0.05), rule violations (r = −0.02, p > 0.05) and the number of unique designs (r = 0.48, p > 0.05). Given

the utility of creating a brief assessment, the remainder of the analyses will focus on row 1 variables.

Inter-rater reliability—To assess inter-rater reliability, intraclass correlations (ICC; absolute agreement) were calculated for the baseline testing in a subset of individuals whose tests were double scored (n = 88). For all three variables examined, the ICC showed a high degree of agreement. The number of perseverations ($\alpha = 0.99$, $p < 0.001$), perseverative distance ($\alpha = 0.99$, $p < 0.001$) and rule violations ($a = 0.99$, $p < 0.001$) in row 1 were significantly correlated across raters.

Test-retest reliability—Pearson correlations were used to estimate the relationship between GPG variables at time 1 and time 2 in the set of individuals who underwent repeat annual testing (103 OAs and 34 ADs). Table 1 shows that the test-retest reliability for each group after controlling for site. For the OA group, four out of five variables were highly significant; the total number of perseverations ($r = 0.39$, $p < 0.001$), rule violations ($r = 0.31$, $p < 0.01$), time ($r = 0.47$, $p < 0.001$), and number of unique designs ($r = 0.56$, $p < 0.001$). For the AD group, the number of perseverations ($r = 0.34$, $p = 0.05$) and the number of unique designs ($r = 0.39$, $p = 0.05$) was significantly correlated at the two time-points.

Between Group Differences in OAs versus AD

After adjusting for age, education, and site, separate one-way ANOVA analyses demonstrated that of the five row 1 GPG variables examined, 3 were significantly different between the OAs and AD groups (Table 2). The three variables that differed significantly were number of perseverations [$F (1, 264) = 53.71$, $p < 0.001$], rule violations [$F (1, 264) = 28.60$, $p < 0.001$] and number of unique of designs [$F (1, 264) = 96.29$, $p < 0.001$].

Data for both the rows was available for only the CUMC site. After adjusting for age, analyses showed similar results. The four variables that significantly differed between OA and AD groups were number of perseverations [$F (1, 32) = 22.56$, $p < 0.001$], perseverative distance [$F (1, 32) = 4.45$, $p < 0.05$], rule violations [$F (1, 32) = 11.19$, $p < 0.01$] and number of unique of designs [$F (1, 32) = 33.48$, $p < 0.001$].

ROC Analysis—As the number of perseverations and unique designs differed between groups and showed the most robust psychometric properties, receiver operating characteristic (ROC) analysis was conducted to determine the optimal cut-off score for differentiating between OA and AD groups. The ROC analysis for perseveration in row 1 produced an area under the curve (AUC) of 0.76 (Figure 2) and yielded an optimal cut-off score of 4 perseverations with a sensitivity of 76% and a misclassification rate of 37%. The ROC analysis for unique designs in row 1 produced an area under the curve (AUC) of 0.81 (Figure 3) and yielded an optimal cut-off score of 15 unique designs with a sensitivity of 81% and a misclassification rate of 36%.

Neuropsychological correlates—Partial correlations were conducted between GPG row 1 perseverations and other neuropsychological measures, and between unique designs from row 1 and other neuropsychological measures, adjusting for site and global cognition (MMSE) within each group. Neuropsychological data for PrVLT, Biber Figure Learning

Test, and WTAR standard scores was obtained from both sites and, data for Digit Span, Spatial Span, Visual Scanning and Verbal fluency was obtained from only one site (CUMC).

For the AD group, a significant positive correlation was obtained between unique designs from row 1 and visual scanning (Visual Scanning Total Targets). No other significant correlations were found for the perseverations and unique designs. For the OA group, number of perseverations on row 1 was negatively correlated with working memory (Digit Span Backward), and with immediate and delayed verbal and visuospatial memory (PrVLT total immediate, PrVLT short and long delay recall, Biber total immediate, Biber short and long delay recall, and recognition discriminability index). In the case of the number of unique designs from row 1, a positive correlation was found for pre-morbid intelligence (WTAR Standard score), and for immediate and delayed visuospatial memory (Biber immediate, short and long delay recalls and, recognition discriminability index). See Tables 3 and 4 for full correlational results examining perseverations and unique designs, respectively.

DISCUSSION

We have examined performance on the GPG test in non-demented older adults and individuals with AD, and examined its psychometric properties. One of the main goals of the current study was to determine if a brief version of the GPG Test, using only one row of designs rather than two, would be useful in characterizing performance on the test. Indeed, the analyses revealed that perseverations, time to completion and number of unique designs are highly correlated across rows in the OA and the AD groups. Previously it has been proposed that several trials of the designs, such as those on the RFFT, may be redundant (Lee et al., 1997). Our study supports this idea. However, we acknowledge that increasing the number of trials may serve to increase the test-retest reliability of the test.

We therefore focused the remaining analyses of test reliability and validity on row 1 variables. Perseverations, perseverative distance and rule violations were all found to have high inter-rater reliability. Examination of test-retest reliability, suggested that only the number of perseverations and the number of unique designs on row 1 was correlated over time for both OAs and AD groups. With regard to between group differences, number of perseverations, rule violations and number of unique designs differed between the OAs and the AD groups. A cut-off score of 4 or more perseverations showed a moderate degree of sensitivity (76%) and specificity (37%) in distinguishing individuals with AD and healthy older adults. Similarly, a cut-off score of 15 or less unique designs showed a moderate degree of sensitivity (81%) and specificity (36%).

Our results are consistent with those of the original Glosser and Goodglass (1990) study using the GPG in which perseverations were significantly different between the patient and control groups. Findings from the current study may at first appear inconsistent with those of a recent study by Possin and colleagues (2012) in which only patients with FTD, and not those with AD, were found to differ from healthy older adults with regard to number of perseverations on a design fluency task. However, this version of the task gives subjects 60 seconds to generate as many designs as possible; as such, patients with AD generated

significantly fewer designs ($M = 8.6$) as compared to healthy older adults ($M = 20.6$). When this difference in the total number of designs is considered, the AD group had proportionately more perseverations than the healthy older adults even though the raw number was comparable (approximately 1.5). Infact, our study also found a consistent pattern wherein the number of unique designs produced was significantly fewer in the AD group compared to healthy older adults. Thus, patients with AD achieved fewer total correct designs by virtue of the fact that they made more perseverations than healthy older adults.

Given the robust psychometric properties of both the perseveration score and the number of unique designs produced, in addition to the presence of between group differences, the neuropsychological correlations and ROC analyses were conducted for both these variables. The pattern of correlations found between GPG variables and neuropsychological measures, across the OAs and the AD groups were somewhat unexpected. Specifically, we hypothesized that the GPG variables would relate most strongly with verbal fluency, another executive task, in both the OAs and AD groups. Interestingly, in the OAs group, GPG perseverations and unique designs were related to indices of visuospatial memory. Whereas the number of perseverations was correlated with both immediate and delayed visuospatial memory indices, and with working memory, the number of unique designs was correlated with the immediate and delayed visuospatial indices, and premorbid intelligence. Generally, perseverative behavior and the generation of unique designs are conceptualized as signs of executive dysfunction that in theory would be dissociable from memory abilities. However, work by Libon and colleagues (1996) as well as others (Lamar et al., 1997; Pekkala, Albert, Spiro Iii, & Erkinjuntti, 2008) has suggested that certain forms of perseverations have their basis in memory rather than executive deficits (Davis, Price, Ball, & Libon, 1999). Emerging evidence shows that different types of perseverations may be associated with different regions of the brain. For example, the presence of recurrent perseverations (i.e., inappropriate repetition of a previous response after a new intervening response) and continuous perseverations (i.e., prolonged repetition of the same response without interruption) might be associated with compromised left and right temporo-parietal regions respectively, whereas stuck-in-set perseverations (i.e. recurrence of an earlier perseverative response after a completely new task is introduced) may be associated with deterioration of the frontal systems.

In the AD group, number of unique designs and perseverations were not correlated with any of the administered neuropsychological tasks, suggesting that both of these elements of the GPG captures an element of cognition in this group that may not otherwise be assessed. Given the findings in healthy older adults, it is somewhat counterintuitive that memory would not relate to GPG in a group with severe memory impairment. However, it may be that floor effects on the memory testing may have affected the degree to which an association can be detected.

In summary, this study is unique in the sense that it is the first study to provide normative data for an untimed design fluency test in both healthy older adults and individuals with AD. Furthermore, we provide evidence regarding the robust psychometric properties of a brief version of the test (6-7 minutes on average), on which number of perseverations may assist in distinguishing between individuals with OAs and ADs. We have also shown that

performance on this task appears to reflect the integrity of visuospatial memory processes. Based on the current findings, clinicians and researchers might consider using the brief version of the GPG test as part of a dementia evaluation. The presence of at least 4 perseverations or less than 15 unique designs on row 1 would be more consistent with an AD profile than with normal cognition. The other variables such as perseverative distance and rule violations likely hold important information for the differential diagnosis of AD versus non-AD dementias including frontotemporal dementia, and ongoing work in our lab is examining patterns of performance across these groups. Future studies should examine the validity of this test with other design fluency tests along with other language and processing speed tests.

Limitations of the study include the fact that not all participants had complete data for all measures evaluated. For example, not all participants completed both rows of the GPG task or were seen for annual follow-up, reducing the sample size for examination of test-retest reliability. It is also important to acknowledge that while administration of an untimed design fluency task will facilitate conceptualization of executive abilities without contributions of motor and processing speed abilities, it is certainly the case that administration of comprehensive battery can also accomplish this goal. However, this teasing apart would require several additional steps, and would leave the examiner with only an inference (rather than evidence) that design fluency would have been intact in the context of intact speed and motor functioning.

In conclusion, this study includes a large sample of well characterized older adults and individuals with AD, offering a useful group in which to examine the psychometric and normative properties of the GPG test. Results support the feasibility and utility of a brief version of the GPG test, and indicate that number of perseverations and/or unique designs are the most useful variable in distinguishing individuals with AD from OAs.

REFERENCES

Baldo JV, Shimamura AP, Delis DC, Kramer J, Kaplan E. Verbal and design fluency in patients with frontal lobe lesions. Journal of the International Neuropsychological Society. 2001; 7(05):586–596. [PubMed: 11459110]

Bozeat S, Gregory CA, Ralph MAL, Hodges JR. Which neuropsychiatric and behavioural features distinguish frontal and temporal variants of frontotemporal dementia from Alzheimer's disease? Journal of Neurology, Neurosurgery & Psychiatry. 2000; 69(2):178–186.

Budson A, Sullivan A, Mayer E, Daffner K, Black P, Schacter D. Suppression of false recognition in Alzheimer's disease and in patients with frontal lobe lesions. Brain. 2002; 125(12):2750–2765. [PubMed: 12429602]

Calderon J, Perry R, Erzinclioglu S, Berrios G, Dening T, Hodges J. Perception, attention, and working memory are disproportionately impaired in dementia with Lewy bodies compared with Alzheimer's disease. Journal of Neurology, Neurosurgery & Psychiatry. 2001; 70(2):157–164.

Davis KL, Price CC, Ball SF, Libon DJ. Error analysis of the CVLT-9. Age (in years). 1999; 77:6.05.

Delis, DC.; Kaplan, E.; Kramer, JH. Delis-Kaplan executive function system. Psychological Corporation; 2001.

Folstein, MF.; Folstein, SE.; McHugh, PR. Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician. Pergamon Press; 1975.

Glosser G, Goodglass H. Disorders in executive control functions among aphasic and other brain-damaged patients. Journal of Clinical and experimental Neuropsychology. 1990; 12(4):485–501. [PubMed: 1698809]

Glosser G, Goodglass H, Biber C. Assessing visual memory disorders. Psychological Assessment: A Journal of Consulting and Clinical Psychology. 1989; 1(2):82–87.

Goebel S, Fischer R, Ferstl R, Mehdorn HM. Normative data and psychometric properties for qualitative and quantitative scoring criteria of the Five-point Test. [Journal Peer Reviewed Journal]. The Clinical Neuropsychologist. 2009; 23(4):675–690. [PubMed: 18923969]

Jones-Gotman M, Milner B. Design fluency: The invention of nonsense drawings after focal cortical lesions. Neuropsychologia. 1977; 15(4-5):653–674. [PubMed: 896022]

Keil K, Kaszniak AW. Examining executive function in individuals with brain injury: A review. Aphasiology. 2002; 16(3):305–335.

Kluger A, Gianutsos JG, Golomb J, Ferris SH, George AE, Franssen E, Reisberg B. Patterns of Motor Impairment in Normal Aging, Mild Cognitive Decline, and Early Alzheimer'Disease. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences. 1997; 52(1):P28–P39.

Lamar M, Podell K, Carew TG, Cloud BS, Resh R, Kennedy C, Libon DJ. Perseverative behavior in Alzheimer's disease and subcortical ischemic vascular dementia. Neuropsychology. 1997; 11(4):523–534. [PubMed: 9345696]

Lee GP, Strauss E, Loring DW, McCloskey L, Haworth JM, Lehman RAW. Sensitivity of figural fluency on the five-point test to focal neurological dysfunction. The Clinical Neuropsychologist. 1997; 11(1):59–68.

Lewis SJG, Dove A, Robbins TW, Barker RA, Owen AM. Cognitive impairments in early Parkinson's disease are accompanied by reductions in activity in frontostriatal neural circuitry. The Journal of Neuroscience. 2003; 23(15):6351–6356. [PubMed: 12867520]

Libon DJ, Mattson RE, Glosser G, Kaplan E, Malamut BL, Sands LP, Cloud BS. A nine—word dementia version of the California verbal learning test. The Clinical Neuropsychologist. 1996; 10(3):237–244.

Mez J, Cosentino S, Brickman AM, Huey ED, Manly JJ, Mayeux R. Dysexecutive Versus Amnestic Alzheimer Disease Subgroups: Analysis of Demographic, Genetic, and Vascular Factors. Alzheimer Disease & Associated Disorders. 2012

Pekkala S, Albert M, Spiro Iii A, Erkinjuntti T. Perseveration in Alzheimer's disease. Dementia and geriatric cognitive disorders. 2008; 25(2):109–114. [PubMed: 18075249]

Possin K, Chester SK, Laluz V, Bostrom A, Rosen HJ, Miller BL, Kramer JH. The Frontal-Anatomic Specificity of Design Fluency Repetitions and Their Diagnostic Relevance for Behavioral Variant Frontotemporal Dementia. Journal of the International Neuropsychological Society. 2012; 18(5):834–844. [PubMed: 22835330]

Possin K, Feigenbaum D, Rankin KP, Smith GE, Boxer AL, Wood K, Kramer JH. Dissociable executive functions in behavioral variant frontotemporal and Alzheimer dementias. Neurology. 2013; 80(24):2180–2185. [PubMed: 23658382]

Price CC, Garrett KD, Jefferson AL, Cosentino S, Tanner JJ, Penney DL, Libon DJ. Leukoaraiosis severity and list-learning in dementia. The Clinical Neuropsychologist. 2009; 23(6):944–961. [PubMed: 19370451]

Regard M, Strauss E, Knapp P. Children's production on verbal and non-verbal fluency tasks. Perceptual and Motor Skills. 1982; 55(3):839–844. [PubMed: 7162920]

Ruff RM, Light RH, Evans RW. The Ruff Figural Fluency Test: a normative study with adults. Developmental Neuropsychology. 1987; 3(1):37–51.

Spreen, O. A compendium of neuropsychological tests: Administration, norms, and commentary. Oxford University Press; 1998.

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

Wechsler, D. WAIS-III/WMS-III technical manual. The Psychological Corporation; San Antonio, TX: 1997.

Weintraub S, Mesulam MM. Mental state assessment of young and elderly adults in behavioral neurology. Principles of behavioral neurology. 1985; 26:71–123.
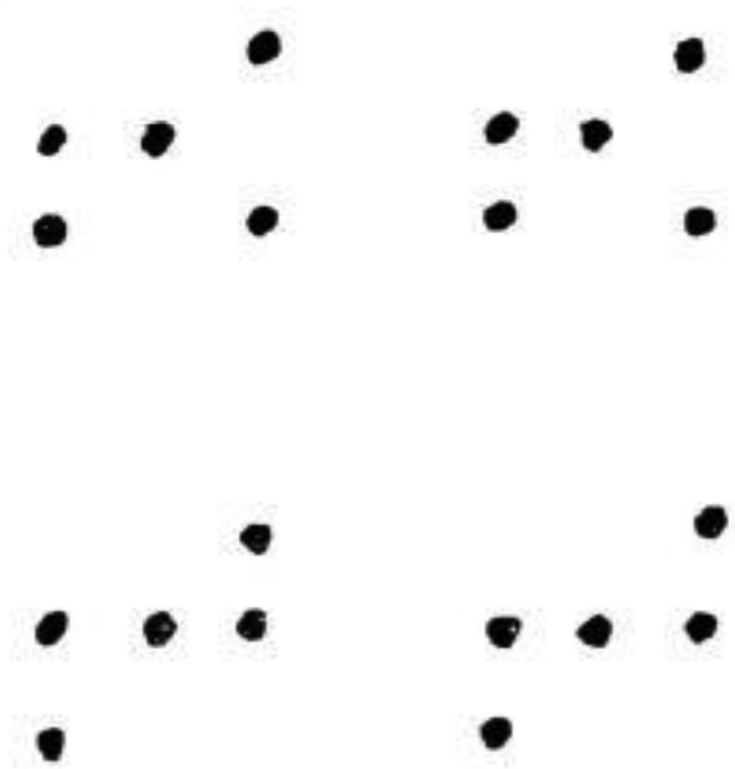
Figure 1.
Sample of the standarized test material for row 1 and row 2 for the GPG test (scaled down detail).
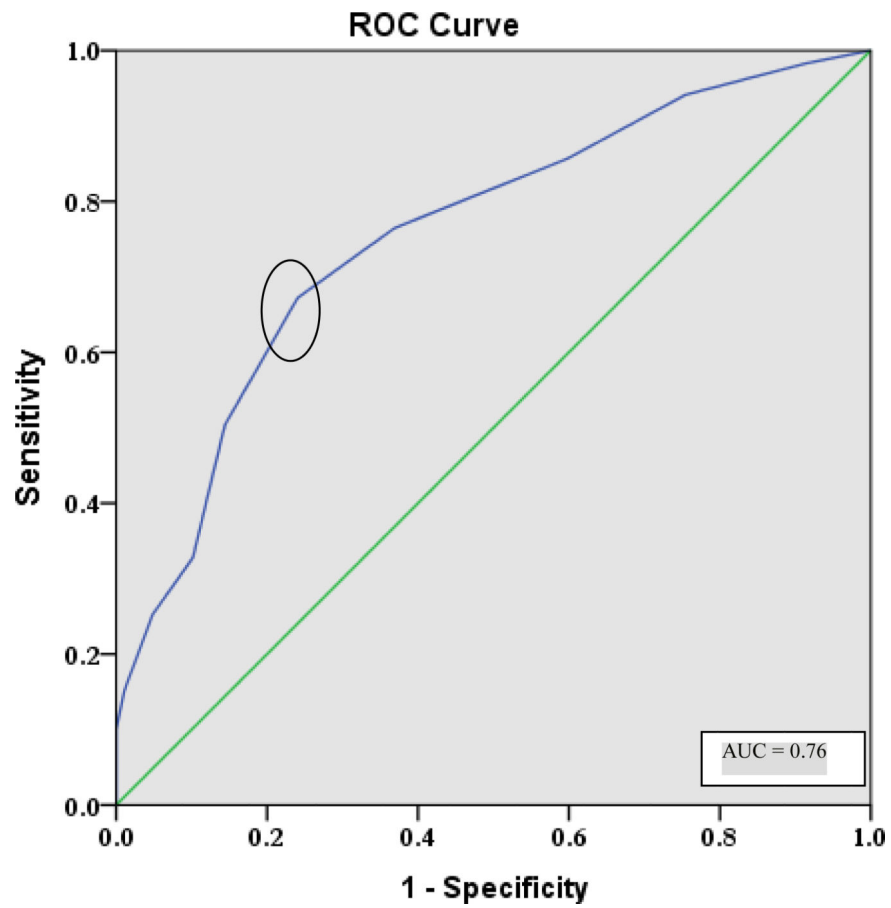
Figure 2.
Receiver operating characteristics (ROC) curve depicting sensitivity and 1-specificity values for separating individuals with AD from healthy elders. The optimal cutoff of greater than or equal to four perseverations on row 1 is circled.
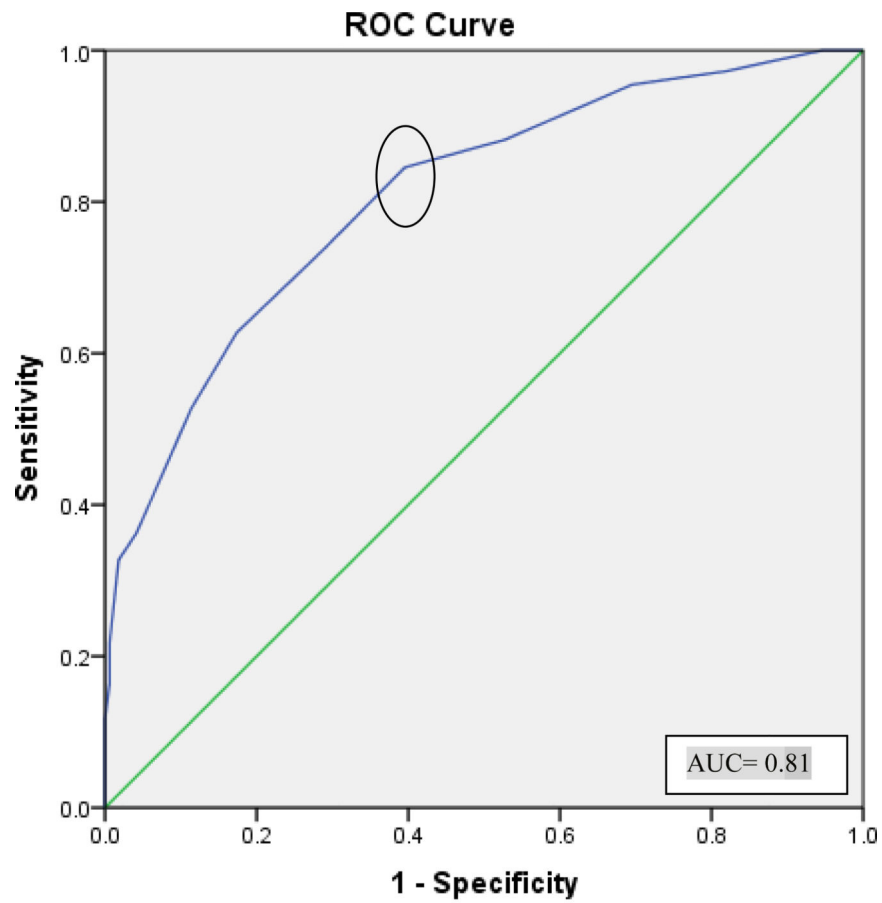
Figure 3.
Receiver operating characteristics (ROC) curve depicting sensitivity and 1-specificity values for separating individuals with AD from healthy elders. The optimal cutoff of greater than or equal to 15 or fewer unique designs on row 1 is circled.

Table 1

Pearson Moment Correlations showing the test-retest reliability for the OAs and ADs for row 1 after one year after controlling for site.

	AD				OA			
	N	Time 1 (Baseline)	Time 2 (One Year)	Pearson's *r*	N	Time 1 (Baseline)	Time 2 (One Year)	Pearson's *r*
Perseverations	34	4.47 (2.23)	5.06 (3.10)	0.34 *	103	3.03 (2.03)	3.04 (2.33)	0.39 ***
Perseveration Distance	34	6.72 (2.51)	6.17 (3.27)	0.10	103	7.18 (3.74)	6.99 (4.05)	0.00
Rule violations	34	2.74 (3.78)	2.68 (2.72)	0.33	103	0.68 (1.28)	0.77 (1.57)	0.31 **
Time (Seconds)	34	409.91 (203.84)	443.62 (240.97)	0.33	103	422.17 (226.09)	392.60 (229.99)	0.47 ***
Unique Designs	34	12.79 (4.03)	12.26 (3.32)	0.39 *	103	16.29 (2.48)	16.19 (2.80)	0.56 ***

Note:

***
p 0.001

**
p 0.01

*
p 0.05

Table 2

Between group differences on the four GPG variables for row 1 and for both the rows

	Row I						Both Rows					
	N	OA	N	AD	F	p	N	OA	N	AD	F	p
Perseverations	162	3.19 (2.09)	107	5.64 (3.03)	53.71	0.000	20	4.90 (4.90)	15	10.67 (3.74)	22.56	0.000
Perseveration distance	160	6.80 (3.57)	107	6.02 (2.67)	2.65	0.105	20	14.49 (4.80)	15	12.04 (2.75)	4.45	0.043
Rule violations	162	0.85 (1.60)	107	2.39 (3.22)	28.60	0.000	20	0.85 (0.93)	15	3.20 (2.93)	11.19	0.002
Time (Seconds)	142	381.64 (225.35)	102	418.94 (243.36)	2.14	0.145	20	936.65 (525.01)	14	809.36 (550.38)	0.39	0.537
Unique Designs	162	15.96 (2.55)	107	11.96 (3.90)	96.29	0.000	20	34.25 (3.34)	15	26.13 (4.64)	33.48	0.000

Table 3

Correlation between GPG Perseverations from row 1 and various neuropsychological tests, after controlling for the MMSE scores and site.

		OA			AD	
	N	**Pearson Correlation**	**Sig. (2-tailed)**	**N**	**Pearson Correlation**	**Sig. (2-tailed)**
Average Verbal Fluency (FAS)	43	−.249	0.100	26	−.313	0.105
PrVLT Total Immediate	134	−.217	0.011*	67	−.063	0.605
Biber Total Immediate	134	−.292	0.001**	67	−.014	0.910
PrVLT Short Delay	134	−.166	0.053*	67	.138	0.259
PrVLT Long Delay	134	−.137	0.111	67	.101	0.410
PrVLT Discriminability index	134	−.237	0.005*	67	.051	0.676
Biber Short Delay	134	−.188	0.028*	67	.101	0.407
Biber Long Delay	134	−.260	0.002**	67	.146	0.231
Biber Discriminability Index	134	−.306	0.000**	67	.119	0.330
WTAR Standard Score	134	−.165	0.056	67	−.137	0.260
Digit Span Forward	43	−.130	0.393	26	−.054	0.786
Digit Span Backward	43	−.396	0.007*	26	.085	0.668
Visual Scanning - Total Targets	43	−.011	0.943	26	−.296	0.126
Spatial Span Forward	43	−.084	0.584	26	−.007	0.972
Spatial Span Back	43	−.066	0.665	26	.058	0.771

Table 4

Correlation between GPG Unique Designs from row 1 and various neuropsychological tests, after controlling for the MMSE scores and site.

		OA			AD	
	N	Pearson Correlation	Sig. (2-tailed)	N	Pearson Correlation	Sig. (2-tailed)
Average Verbal Fluency (FAS)	43	.280	0.062	26	.335	0.082
PrVLT Total Immediate	134	.063	0.465	67	.033	0.787
Biber Total Immediate	134	.210	0.014*	67	−.031	0.799
PrVLT Short Delay	134	.092	0.289	67	.080	0.515
PrVLT Long Delay	134	.086	0.318	67	−.132	0.280
PrVLT Discriminability index	134	.159	0.065	67	.122	0.319
Biber Short Delay	134	.234	0.006*	67	.080	0.512
Biber Long Delay	134	.267	0.002**	67	−.041	0.738
Biber Discriminability Index	134	.252	0.003**	67	.089	0.468
WTAR Standard Score	134	.205	0.017*	67	.200	0.099
Digit Span Forward	43	−.017	0.910	26	.195	0.319
Digit Span Backward	43	.185	0.224	26	−.002	0.994
Visual Scanning - Total Targets	43	−.010	0.947	26	.369	0.054
Spatial Span Forward	43	−.107	0.485	26	.011	0.955
Spatial Span Back	43	.039	0.798	26	−.216	0.269