Clustering, Switching & Verbal Fluency Assessment: A Critical Evaluation of Introspection and Interview-based Scoring Procedures

Antonio M. Bottos

Department of Psychology

The College at Brockport, State University of New York

June 11, 2019

A Senior Honors Thesis

Submitted in Partial Fulfillment of the Requirements
for Graduation in the Honors College

Thesis Advisor: David A. Abwender, PhD

Research Assistants: Peter Brand, Megan Fribance, & Danielle Colon

Presented at the 2019 National Conference for Undergraduate Research (NCUR)

Kennesaw, GA

April 11th, 2019

Presented at the 36th Annual Scholars Day

Brockport, NY

April 10th, 2019

Acknowledgements

First and foremost, I'd like to extend my gratitude to my family, friends and many mentors at The College at Brockport, State University of New York for supporting me in my journey as a growing scholar and professional. Additionally, I'd like to thank Peter Brand, Megan Fribance, and Danielle Colon for working diligently as research assistants helping with collecting data, scoring and encoding. Our hard work *together as a team* allowed this project to transform into fully reportable study with significant findings. I'd like to give special thanks to Dr. David Abwender for being my thesis director, the Department of Psychology for it's continued support in student research, and the Honors College for providing me with the opportunity to be the Principal Investigator of my very own honors thesis.

Last and certainly not least, I'd like to thank Barbara Thompson, Dr. Algernon Kelly, Professor Helen Cheves & CSTEP/McNair for dedicated mentorship as a Ronald E. McNair Scholar. The CSTEP/McNair program funded the printing of this study's poster and covered my travelling expenses to Kennesaw State University, GA to present at NCUR 2019.

As a first-generation college student, your support has been truly invaluable.

Warmest regards,

Antonio M. Bottos

Abstract

Semantic Verbal Fluency (SVF) is a neuropsychological assessment that requires respondents to rapidly generate words in specified semantic categories (e.g., different kinds of animals). It provides information on the integrity of examinees' executive function and semantic memory by looking at clustering of words generated and switching between clusters. According to recent literature, extant methods for scoring clustering and switching based on examiner-identified semantic associations among words should be abandoned in favor of scores derived from asking participants *post hoc* how they think they generated their words the way they did (Body & Muskett, 2013). Repeated measures ANOVA indicated that leading questions, a demand characteristic inherent in interviewing a participant after administering SVF, lead to more clusters, clustered words, cluster switches (CS) and decreased hard switches (HS) on interview-based scoring procedures across low demand and high demand queries – evidence supporting a threat to internal validity inherent in Body & Muskett's (2013) claim and methodology. Cueing three semantic categories (e.g., pets, sea creatures, farm animals) before administering SVF was also found to increase mean number of clusters, clustered words, CS and decrease HS compared to un-cued groups – evidence consistent with literature (Abwender, Preston, & Steffenella, 2003; Hurks, 2012) Additionally, HS on SVF traditional scoring measures correlated with total RFFT designs, while HS on SVF interview scoring measures correlated with total RFFT perseverations, suggesting the two scoring procedures are measuring different variables. No correlations were observed between CVLT-II clustering and SVF clustering using neither traditional scoring methods nor interview-based methods. We bring caution to the validity of interview-based methods employed by Body & Muskett (2013) as our preliminary study empirically supports a major threat to internal validity in their design.

Introduction

Verbal Fluency (VF) is a neuropsychological assessment that is used to assess an individual's ability to generate appropriate verbal output under specific guidelines. VF assessments are broken down into two main task variants: Letter (F-A-S) Fluency and Semantic Fluency (SVF). For a letter fluency task, participants are asked to overtly name all words that come to mind beginning with a particular letter (typically, F, A, or S). Similarly, for semantic fluency, participants are asked to name all words that come to mind from a specified semantic category. The most commonly used semantic category is "animals." Once the task commences, the participant has (usually) one minute to rapidly generate words under the stated guideline. The words are recorded (traditionally hand-written but sometimes audiotaped) by the examiner as they are spoken and analyzed later for clustering and switching of word output. Clustering and switching are representative of two major cognitive functions at work, and scoring such gives insight on the integrity of the respondent's semantic memory and executive function. The present study seeks to compare two approaches to identifying and scoring clusters and switches on SVF.

Clustering as a Measure of Semantic Memory

Clustering occurs when words are generated sequentially under a common subcategory, which demonstrates organization by association (Abwender, Swan, Bowerman & Connolly, 2001; Troyer, Moscovitch & Winocur, 1997). Extant scoring methods for animal fluency require the examiner to determine the presence of clusters based solely on objective indicators such as accepted zoologic relationships (e.g., felines), living environments (e.g. sea creatures), human use (e.g., beasts of burden), geographic relationships (e.g., Australian animals), etc. (Troyer et al.1997). Further examples of semantic clustering occur when a participant generates the words "fish, lobster, shark" during an animal naming task. These words share a similar living

environment, as they all belong to the semantic subcategory "Sea Creatures". Other examples of clustering include "cat, lion, cheetah" as sharing a zoological relatedness (i.e. feline) and "dog, hamster, guinea pig" as having a shared human use (i.e., common household pets; Abwender et al., 2001; Troyer et al.,1997). Additionally, traditional scoring methods proposed by Abwender et al. (2001) have found that words on SVF may also be clustered using principles of phonemic organization (i.e., clustering words on the basis of similar phonetic devices). Phonemic clustering can be demonstrated on the SVF when a participant sequentially states examples such as "bird, bear, baboon" (indicative of a common first letter), "frog, dog, hog" (indicative of rhyming phonemes) and "antelope, badger, crocodile" (representative of an alphabetical ordering). Past research has suggested that mean number of clusters and cluster sizes are positively correlated with the quality of semantic memory functions linked to the brain's temporal lobe (Troyer, Moscovitch, Winocur, Leach, & Freedman, 1998).

Switching as a Measure of Executive Function

Switching occurs when words generated sequentially transition from either a cluster to a single unrelated word (and vice versa) or when generated words shift from one cluster to another cluster (Abwender et al. 2001; Troyer et al.1997). These processes are referred to as hard switching and cluster switching, respectively (Abwender et al., 2001). Past literature has suggested switching is positively correlated with executive functions associated most closely with the brain's frontal lobe (Troyer et. at., 1997).

Clinical Applications of SVF: Alzheimer's Disease, Schizophrenia & Parkinson's Disease

Persons with Alzheimer's Disease (AD) have temporal lobe dysfunction primarily resulting from the accumulation of neurofibrillary tangles (i.e., hyperphosphorylated Tau protein) and β-amyloid plaques in the brain (Danysz & Parsons, 2012). Patients with AD tend to

produce clusters that are noticeably simpler and smaller than those of non-demented persons (i.e., fewer category exemplars and types of categories; Binetti, Magni, Cappa, Padovani, Bianchetti & Trabucchi, 1995; Epker, Lacritz, & Munro Cullum, 1999; Testa, Fields, Gleason, Salmon & Beatty, 1998). Compared to either controls or non-demented Parkinson's patients, AD patients produce smaller and fewer clusters and also fewer switches (Epker et al., 1999; McDowd et, 2011; Tröster, Fields, Paul, Blanco, Hames & Salmon, 1998). Notably, severe-AD patients produce smaller clusters than do mild-AD patients, who in turn produce smaller clusters than do individuals with memory issues, who themselves produce smaller clusters than controls do (Binetti et al., 1995; Fagundo, López, Romero, Guarch, Marcos & Salmero, 2008). The sub-set of individuals with memory complaints who went on to develop dementia and be diagnosed with AD produced smaller clusters than those who did not develop dementia (Fagundo et al., 2008).

Similar to AD patients, schizophrenic individuals also exhibit temporal lobe dysfunction. Among schizophrenic persons, clustering does not strongly predict total word output, although among controls it does (Robert, Lafont, Medecin, Berthet, Thauby, Baudu & Darcourt, 1998). With few exceptions, research shows that individuals with schizophrenia produce smaller (Bozikas, Kosmidis, & Karavatos, 2005; Moelter, Hill, Ragland & Lundardelli, 2001; Robert et al., 1998; Zakzanis, Troyer, Rich, & Heinrichs, 2000) and/or fewer clusters than controls on SVF (Elvevåg, Fisher, Gurd, & Goldberg, 2002; Giovannetti, Goldstein, Schullery, Barr, & Bilder, 2003; Robert, Migneco, Marmod, Chaiz, Thauby, Benoit & Beau, 1997). Schizophrenic persons also tend to produce clusters in noticeably different patterns than controls (Moelter et al., 2001), typically featuring frequently-named animals to the exclusion of animals less frequently named (Sung, Gordon, Vannorsdall, Ledoux, Pickett & Pearlson 2012).

Likewise, Parkinson's disease (PD) is notable for deficits in frontal/executive function that accompany it (Bondi, Alfred, Kaszniak, Bayles, & Vance, 1993; Martignoni, & Calandrella, 2009); consequently, it is typical for PD patients to show normal clustering on semantic VF tasks (McDowd et al., 2011) but exhibit impairments on the letter VF task (thought to be the more executively driven version; Epker et al., 1999). Finally, temporal but not frontal lobe lesions have been associated with impairments in semantic clustering (Troyer et al., 1998). A dissociation in clustering performance between persons with left- versus right- temporal lobe (TL) lesions was found such that left-TL patients produced smaller clusters compared to right-TL patients (Troyer et al., 1998). Huntington's disease (HD) has little effect on the temporal lobes (Hedreen & Folstein, 1995), so it is makes sense that HD patients produce average cluster sizes on SVF (Testa et al., 1998).

Highlighting the vast literature depicting the numerous clinical applications of SVF, not only elucidates the assessment's significance, but also supports the validity of using traditional methods (Troyer et al., 1997; Abwender et al., 2001) to score clustering and switching as valid measures of cognitive function. Established scoring measures for SVF are widely used in both the literature and clinical practice of a neuropsychologist.

SVF: Introspection & Interview-Based Scoring Procedures

Recent literature has suggested that traditional methods for scoring clustering and switching on VF tasks (e.g., Abwender et al., 2001; Troyer et al., 1997) need to be revised to include qualitative measurement of self-report when scoring SVF tests (Body & Muskett, 2013). Body and Muskett suggest asking participants, after testing, to explain for themselves if and how their words make up clusters, which is inconsistent with the established guidelines predominantly found in the literature and in practice (i.e., well-established scoring methods set

forth by Troyer et al., 1997, and Abwender et al., 2001, involve using a validated scoring rubric, which details which words comprise clusters, and which comprise switches without the use of interviewing the respondent for an explanation of their thought process.) Additionally, the research underlying Body & Muskett's claim lacks sufficient methodological soundness to ensure validity. It is likely that there is an inherent demand characteristic in asking participants to "retro-introspect" about their reasons for ordering word output the way they did, by assuming that there even was a deliberate, conscious reason for clustering in the first place. By asking participants to retroactively explain their cognitive processes for generating words, it is possible that this demand in itself is implying that there was a conscious, reportable reason behind clustering, when in fact there may never have been a reason to begin with, prompting respondents to make up reasons on the spot. Work done by Heavey (2013) supports this claim, which found that an experimenter's mere assumption that there is an answer to their question can lead participants to produce answers even if there aren't any. For example, asking individuals how they are feeling typically elicits a feeling word as a response. However, the answer should usually be "nothing" because individuals have been found to actually experience pure feelings only a quarter of the time (for detailed explanation of investigation of pristine inner experience and ability to recall it, see Heavey & Hurlburt, 2008). In Body & Muskett's methodology, it was also highlighted that they asked participants to "Tell us how you came up with these words, and describe any links between them" (Body & Muskett, 2013). By asking participants to additionally explain *links* between words, this instruction may arguably have an even higher level of demand, as participants hearing this instruction may actively search for as many links between words as possible to satisfy the experimenter's instruction, running the risk of misremembering their own thought process or fabricating processes that hadn't actually

occurred. Moreover, past research has indicated that a person's ability to introspect in general, has been shown to be fairly unreliable; a study by Johansson, Hall, Sikström, and Olsson (2005) had individuals choose one of two photos they preferred, later asking them to describe the intentions and mental processes behind their choice. At the point of being asked to describe their intentions behind their choice, the participants were shown the photo they had *not* chosen. In 70% of the cases, participants provided clear explanations for why they "chose" the photo that they hadn't chosen! Wilson and Schooler (1991) found similar findings by asking participants which of two pictures they preferred (participants made a whole series of such judgments) and after a brief delay asked them to explain why they preferred the ones they chose – while showing them some of the pictures they originally disliked. Once again, the majority of participants completely fabricated justification for "preferring" pictures that they previously identified as being disliked, showing how unreliable retro-introspections can be and how strongly experimenter demands influence them. When taking the findings from studies like these into consideration, it is likely that Body & Muskett's study possesses multiple threats to internal validity in the form of multiple levels of experimenter demand characteristics. In addition to threats to internal validity, a very small sample used in Body and Muskett's (2013) study (i.e., n=10) is indicative of a threat to external validity. It is nearly impossible to effectively generalize results from a behavioral study such as this one with a sample size as small as 10; not to mention that Body and Muskett (2013) failed to conduct any sort of statistical analysis. Their claims are purely based off of reporting select excerpts from their interviews, which are incomplete as they stand. Thus, with threats to both internal and external validity, the results from Body & Muskett (2013) cannot be uncritically applied to clinical and academic applications of VF assessment.

Research Question

The present study is aimed at investigating whether or not interview-based scoring measures are justifiable by comparing traditional, established methods of scoring clustering and switching against novel methods incorporating respondents' own accounts of why they produced words in the order they did (i.e., interview-based). We aimed to investigate how demand characteristics inherent in interviewing according to Body & Muskett's methodology may influence participant responses when self-reporting thought processes behind the VF test and if some connections identified by participants are better explained in terms of idiosyncratic reasoning. Specifically, we looked to compare interview query instructions that were high in level of demand characteristic (i.e., labelled as "high demand"; query instructions replicating Body & Muskett's 2013 methodology as written) against interview query instructions intended to generate lower demand characteristics (i.e., labelled "low demand"; query instructions that ask the participant to explain how they generated their words, excluding the latter half of Body & Muskett's methodology asking participants to "think of *links* between words"). We also aimed at exploring the effects of cuing and its relevance to SVF administration. Some literature highlights that when a participant receives a cue before the SVF test (e.g., task instructions that suggest to respondents that their words can be subcategorized as pets, African animals, or reptiles, which are semantic subcategories of the superordinate category "Animals"), the participant generates more clusters and larger clusters, exhibiting a higher level of organization overall (Abwender et al., 2003; Hurks, 2012). These findings are consistent with spreading activation within the semantic network model, which visually depicts how semantic memory is organized in clusters, and how the activation of a single node (semantic concept labelled with a word) increases the likelihood another closely related node or word also being activated (Quillian, 1962; Quillian

1968; Collins & Loftus, 1975). According to the semantic network model, words that are closer in association to the first node are likely to share a semantic commonality, thus being a part of a larger cluster or subcategory (Quillian, 1962; Quillian 1968). From this, our lab aimed at attempting to replicate the findings on cuing subcategories in SVF, and how the level of query present in the interview process (i.e., low demand vs high demand) may interact with cueing. Lastly, we aimed to see which clustering and switching scores SVF under set scoring measures (i.e., traditional vs interview) correlate with other tests established for semantic memory and executive function respectively.

Hypotheses

Hypothesis 1a (Main effect of scoring method): Asking participants, after administering VF, to explain why they said their words in the order they did creates a demand characteristic, so interview-based scoring will yield more semantic clusters, more cluster switches, and fewer hard switches than traditional scoring.

Hypothesis 1b (Main effect of demand level): Participants will report larger clusters, more cluster switches, and fewer hard switches on high demand query-based scoring ("explain how these words were generated and identify any links between each word") as opposed to low demand query-based scoring ("tell me how you came up with these words").

Hypothesis 2a (Main effect of cueing): Task instructions that cue participants to think of categories (e.g. pets, farm animals, sea creatures) results in spreading activation/priming, so cued VF will yield larger clusters, more cluster switches, and fewer hard switches than instructions that don't cue.

Hypothesis 2b (Demand x cue interaction): On interview-based SVF's, if both high demand query and cuing increase the number of clusters reported, (where cueing involves activating a semantic network, and giving a high demand query pressures a participant to make more links between words) we would hypothesize these two variables to have an interaction with each other. This interaction would be observed when the number of clusters on high demand/cued groups is synergistically much higher than expected compared to high demand/un-cued and low demand/cued groups (where clustering scores would be intermediate), and in low demand/un-cued groups (where clustering scores would be the lowest). We would not expect to see the same interaction under the traditional scoring method, because level of demand would not affect clustering in this case, since level of demand pertains strictly to the interview process.

Hypothesis 3 (Convergent validity): If interview-based scoring is different from traditional scoring with respect to its validity, we would expect clusters and switches with one method to convergently correlate better with established measures for executive function and semantic memory.

Methods

Participants

Undergraduate students at the College at Brockport, State University of New York (n = 104; mean age = 19.6 +/- 4.2; 77 females, 26 males, 1 non-reporting) were recruited to participate in this study. Inclusion criteria for participation consisted of (a) the participant is enrolled in Introduction to Psychology, (b) the participant is at least 18 years of age and (c) the participant has not participated in this study before. Participants were recruited via SONA, an online human participant pool management system operated by the Department of Psychology at The College at Brockport.

Measures

Semantic Verbal Fluency Test (Troyer et al., 1997; Abwender et al., 2001). The semantic verbal fluency test is a neuropsychological assessment that requires participants to rapidly generate as many words that come to mind under a given guideline (most typically the category is "animals"). By analyzing data depicting participant word clusters and switches, the experimenter (typically a clinical neuropsychologist) is able to ascertain the state of the respondent's semantic memory and executive function.

For each participant, interview-based scoring was done by the experimenter who ran the participant and traditional scoring was done by a different member of the research team who was blind to the participant's experimental conditions and to the results of the interview-based scoring. Likewise, interview-based scoring was done blind to the results of the traditional scoring.

California Verbal Learning Test II (CVLT II; Delis & Kramer, 2001). The CVLT II is a neuropsychological assessment that requires participants to learn a list of 16 words over a span of five trials. After each presentation of the list, participants are asked to overtly recollect the words from the list in any order and the responses are recorded. The list contains words from four semantic categories (but it is presented in a "scrambled" order). Participant recollection of the words in semantically organized clusters is associated with better semantic memory function. Participants, alternatively, sometimes recollect words based on serial position in the list (e.g., primacy and recency effects, wherein they recall, respectively, the first or last few words on the list in order), which is presumed to track integrity of so-called episodic memory. Computerized scoring calculates per trial serial and semantic clustering scores as a function of number of words recalled on each trial.

Ruff Figural Fluency Task (RFFT; Ruff, 1996). The RFFT is a timed, neuropsychological assessment that requires participants to connect various patterns of dots provided on a sheet of paper in as many unique ways as possible, without making repeated designs. The number of unique designs, and repeated designs (perseverations) are measures of executive function (i.e., switching mental sets and self-monitoring deficit). As a visual-spatial analogue of verbal fluency, demanding production of a large number of different responses, it also can be analyzed for patterns (clusters, switches) in output.

Participants completed five 60" trials with sample trials prior to each trial. Each trial consists of 35 1.5" squares in a 5x7 array on standard letter size paper with five dots in each square. In the first three trials the dots are in the formation of a regular pentagon and in trials 4 and 5 the dots are in a different seemingly random pattern. In the first trial there are only dots in each rectangle but trials 2 and 3 use diamonds and lines, respectively, to mislead or attempt to

slow down participants. Each participant completes each of the five trials in the same order. In each trial, the participant must create as many designs as possible in 60" by drawing line segments (minimum = 1) that connect at least two of the five dots. Designs can be as simple or elaborate as participants wish. The basic score for the RFFT is the number of unique designs per trial (maximum = 35; although healthy examinees occasionally approach 35 designs in 60", it is exceedingly rare for someone to finish all 35 boxes before time for the trial expires). Designs that replicate ones previously created on that trial do not count toward the unique designs total but are tallied as perseverative errors (even if they are not, strictly speaking, perseverations).

Procedure

In a mixed-subjects true experimental design, the participants were randomly assigned to one of four SVF groups, each consisting of two crossed between-subject variables (i.e. [1] un-cued/low demand, [2] un-cued/high demand, [3] cued/low demand, or [4] cued/high demand). In each folder (one folder per participant), three clinical assessments (i.e., SVF, CVLT-II, RFFT) were presented in a randomized order to minimize test ordering effects. Under traditional scoring procedures set forth by Troyer et al. (1997) and Abwender et al. (2001), plus modified (viz., introspective) measures suggested by Body and Muskett (2013), we analyzed SVF scoring differences across traditional vs interview scoring, low demand query vs high demand query, un-cued SVF vs cued SVF and correlations between [1] clustering scores on SVF with semantic clustering on the CVLT-II and [2] switches on SVF with perseveration scores and total number designs created on the RFFT). These assessments are appropriate for Verbal Fluency because CVLT-II semantic clustering scores have concurrent validity with semantic memory, and the RFFT has concurrent validity with switching between rules, which is an executive function (Troyer et al., 1997; Abwender et al., 2001). Convergent validity is demonstrated if there is a

significant correlation between clustering and switching scores on SVF with scores on the corresponding assessments.

After completion of all assessments by the participant, the SVF's were scored two ways: (1) using traditional scoring procedures (Troyer et al. 1997; Abwender et al., 2001) and (2) using interview-based scoring procedures (Body & Muskett, 2013) representing our study's within-subjects variable. For the first between subject variable, un-cued SVF is reflected in the verbal fluency administration instruction "For this task, I want you to tell me as many animals as you can" compared to the cued instruction "For this task, I want you to tell me as many animals as you can. They can be farm animals, pets, sea creatures – any animals at all." The second between subject variable (low demand query vs high demand query) relates specifically to the level of demand used to query the participants thought processes behind producing their words during the fluency task, and is a component of the interview process. The low demand query involved researchers stating "I'd like for you to tell me how you came up with these words the way you did (followed by an abridged query following each word; e.g., "How did you think of cat? How did you think of dog?")" and the high end demand query involved researchers stating "I'd like for you to tell me how you came up with these words the way you did, and would like for you to tell me any links between them" (followed by an abridged query following each word; e.g., "How did you go from cat to dog? How did you go from fish to zebra?"). The high demand query is designed to replicate Body and Muskett's methodology as closely as possible, while the low demand query is designed to represent the same question, except with a lower level of experimenter demand, that doesn't ask participants to report "links."

Results

Hypothesis 1a: Repeated measures ANOVA compared SVF scores derived under the traditional scoring method with SVF scores derived under the interview-based method, and found that the interview-based SVF scoring measure on average yielded significantly more semantic clusters and cluster switches (see Table 1 & Figure 1). No main significant differences were observed with the number of semantically clustered words and number of hard switches across scoring measure type (i.e. traditional vs interview-based). Thus, this hypothesis receives partial support.
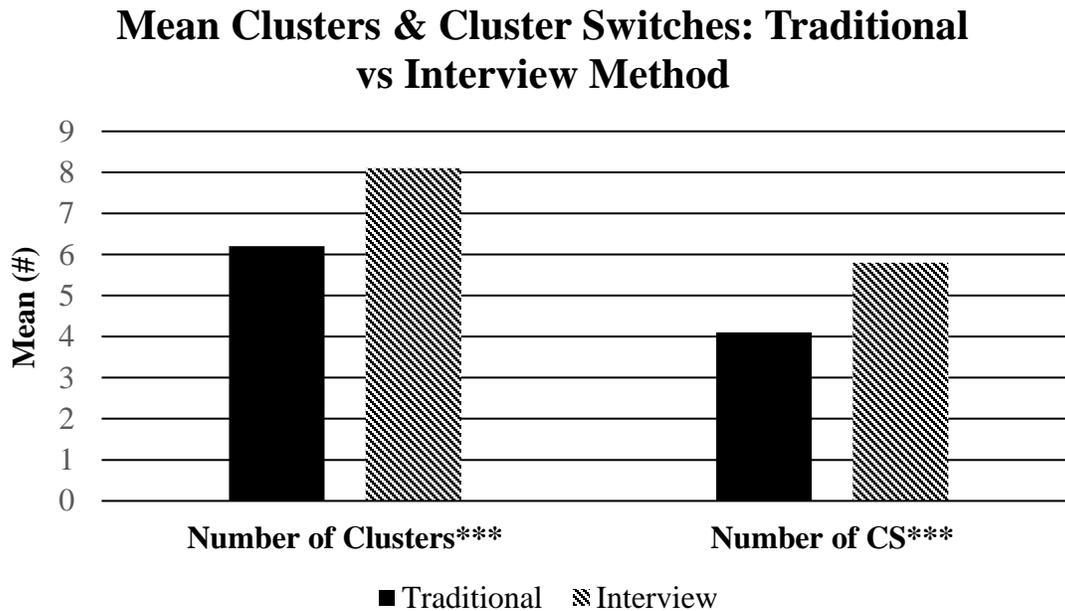
Table 1.
Statistical Comparison of SVF Means: Traditional vs Interview Scoring Methods

SVF Scoring Method	Mean Number of Clusters***	Mean Number of Words Clustered	Mean Number of Cluster Switches ***	Mean Number of Hard Switches
Traditional	6.2 (± 1.8)	12.6 (± 4.3)	4.1 (± 2.0)	4.8 (± 3.1)
Interview-based	8.1 (± 3.3)	12.6 (± 5.6)	5.8 (± 3.5)	4.8 (± 3.3)

*Note: *** = p<.001; N=104 (Both Traditional & Interview)*

Figure 1.

Statistical Comparison of Mean Clusters & Cluster Switches: Traditional vs Interview

### Mean Clusters & Cluster Switches: Traditional vs Interview Method

*Note: *** = p < .001; The figure above depicts findings highlighted in Table 1, under columns labeled "Mean Number of Clusters" and "Mean Number of Cluster Switches; For Figure 1, CS= Cluster Switches.*

Hypothesis 1b: Compared to interview-based SVF's in low demand query groups ($n = 46$), high demand query groups ($n = 58$) on average yielded significantly more clusters, more clustered words, more CS, and fewer HS (see Table 2 & Figure 2). This fully supports this hypothesis.

Table 2.
Statistical Comparison of Interview-based SVF Means: Low Demand vs High Demand Query

Interview-based SVF: Query Demand Level	Mean Number of Clusters**	Mean Number of Words Clustered**	Mean Number of Cluster Switches**	Mean Number of Hard Switches*
Low Demand	7.2 (±2.7)	11.7 (±5.5)	5.5 (±2.9)	6.0 (±3.3)
High Demand	8.8 (±3.5)	13.3 (±5.6)	7.3 (±4.1)	3.9 (±2.9)

*Note: * = p < .05; ** = p < .01; n=46 (Low Demand); n=58 (High Demand)*

Figure 2.
Statistical Comparison of Means using SVF Interview Method: Low Demand vs High Demand

SVF Interview Method: Low Demand vs High Demand Query
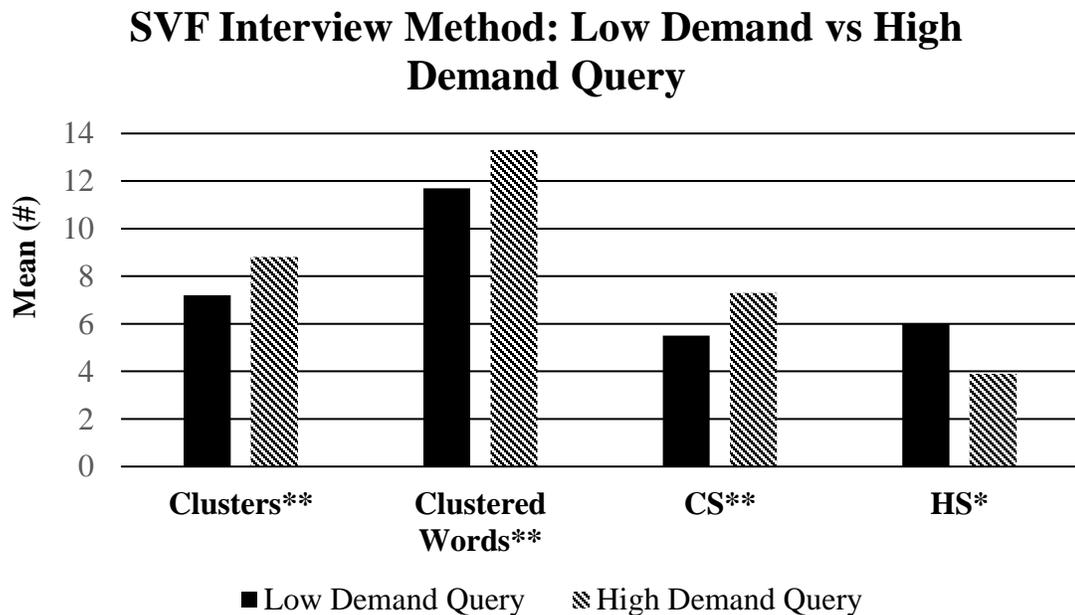
*Note: ** = p < .01, * = p < .05; The figure above depicts all findings highlighted in Table 2;*

For figure 2, CS=Cluster Switch, HS=Hard Switch.

Hypothesis 2a: Compared to un-cued participants ($n = 50$), cued SVF participants ($n = 54$) on average produced significantly more clusters, more clustered words, more cluster switches, and fewer hard switches under traditional scoring measures (see Table 3). This supports the hypothesis.

Table 3.
Statistical Comparison of Traditional-Based SVF Means: Un-cued vs Cued Groups

Traditional SVF Scoring: Presence of Cue	Mean Number of Clusters*	Mean Number of Words Clustered*	Mean Number of Cluster Switches*	Mean Number of Hard Switches*
Un-cued	5.7 (±1.8)	11.1 (±4.0)	3.6 (±1.9)	5.5 (±2.7)
Cued	6.5 (±1.7)	13.9 (±4.1)	4.1 (±2.0)	4.1 (±3.3)

*Note: * = p<.05; N=50 (Un-cued); N=54 (Cued); parallel findings from interview-based scoring are not shown here. Figure 2b shows parallel findings for mean clusters.*

Figure 3.

Statistical Comparison of Clustering Means with SVF Interview Method: Query & Cue
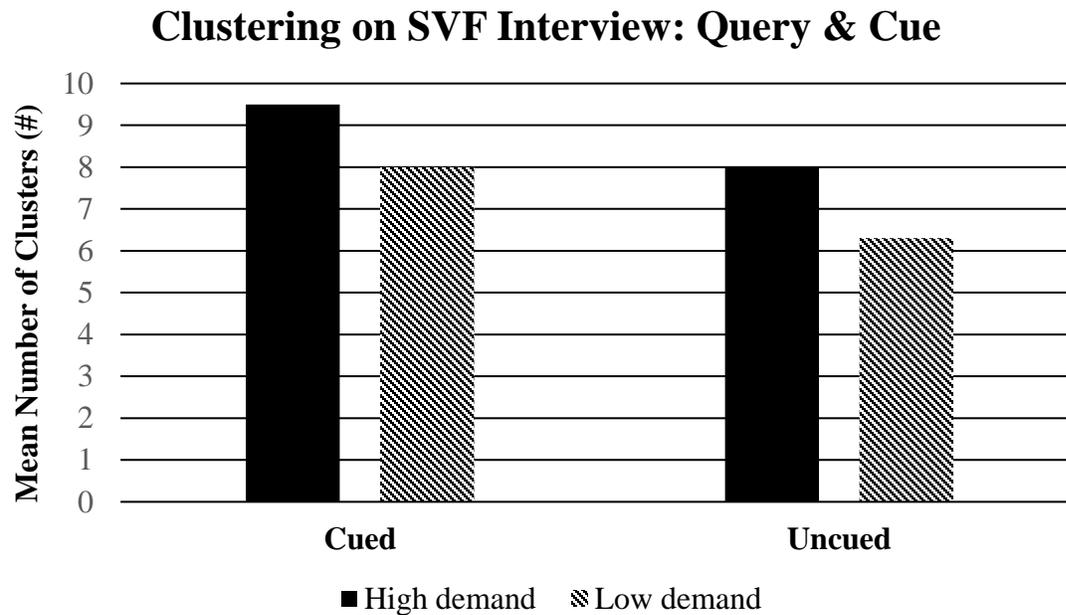
Note: Parallel findings for interview-based scoring from Table 2a are shown, but only for mean number of clusters; N=50 (Un-cued); N=54 (Cued)

Hypothesis 2b: Under interview-based scoring, no significant interactions were noted with regards to average number of clusters (i.e., Query level & Cue). With mean number of clusters as the dependent variable, clustering was lowest in the un-cued/low demand group,

approximately equal in un-cued/high demand and cued/low demand groups and highest in the cued/high demand (see Figure 3).

  Hypothesis 3: With traditional SVF scoring, correlation analysis found that Hard Switch (HS) count correlated with total unique designs created on the RFFT ($r = .20$; $p < .05$). However, with interview-Based SVF scoring, we found that HS count instead correlated with total repeated designs on the RFFT ($r = .26$; $p < .01$). Neither the interview method nor the traditional method yielded clustering scores that significantly correlated with semantic clustering on the CVLT-II.

Discussion

Based off our mixed-subjects true experimental design, we summarize the following findings: [1] a main effect of scoring method was observed, where interview-based SVF yielded more word clusters and more cluster switches than traditional scoring ; [2] significant differences were observed across low demand and high demand query groups on the interview based SVF, where high demand groups yielded more clusters, more clustered words, more CS and fewer HS; [3] across un-cued and cued traditionally-scored SVF, cued SVF lead to an increased number of clusters, more clustered words, more CS and fewer HS; [4] no significant query (i.e., low demand vs high demand) x cue (un-cued vs cued) interactions were observed on interview based SVF, where cue and query level increased the number of clusters as separate variables; and [5] based off our correlational analysis with SVF switching and RFFT performance, traditional-based HS correlated with total number of unique designs created on the RFFT, but *interview-based* HS correlated with total number of perseverative errors (i.e., repeated designs) on the RFFT. No significant correlations were observed between CVLT-II performance and SVF clustering or switching scores.

Comparison of SVF scores generated by interview versus traditional methods reveals the effect of experimenter demand. Moreover, varying the interview demands reveals that asking participants on SVF to "Tell us how you came up with these words, and state any links between them" leads to even more clusters, more CS and fewer HS. This finding suggests that the demand inherent in Body & Muskett's interview methodology does significantly alter responses. This finding is consistent with the literature depicting the effect of demand characteristics on participant responses (Heavey & Hulbert, 2008; Johansson et al., 2005; Wilson & Schooler, 1991). Heavey and Hulbert (2005) found people only experience pure feelings a quarter of the

time, but put under the demand of asking "How are you feeling?", the majority of people will make up feelings instead of saying "nothing" which might be a more accurate depiction. Additionally, Wilson and Schooler (1991) and Johansson et al. (2005) found that when people are presented a picture of an item (e.g., peanut butter) that was previously identified as disliked by the same people, the majority of people will fabricate reasons for why they liked this item, when given the demand "Why do you like this picture?" It can be inferred that the demand present in Body and Muskett's (2013) interview method pressures participants to assume that there were reportable links between words and therefore to fabricate links between words, resulting in fewer HS, but more CS and more clustered words.

A comparison of un-cued vs. cued groups indicates that cueing a semantic category improves performance on clustering and cluster switching. This is consistent with the findings presented in the literature (Abwender et al., 2003; Hurks, 2012), where giving a cue in the task instructions before delivering a VF test yields more clustering and less switching. It fits within the semantic network model of memory (Quillian, 1962; Quillian 1968; Collins & Loftus, 1975). Under the semantic network model, cueing a semantic subcategory such as "pets" during the animal naming SVF likely "primes" this specific network cluster, and perhaps animal clusters in general, increasing the likelihood these categories will be exhausted more thoroughly. This might explain why we see more clusters, clustered words, cluster switches (often representing overlap between networks of clusters) and fewer HS (representing the failure to cluster or switch to another cluster) on cued SVF.

In terms of the hypothesized interaction between query x cue, no significant interactions were observed; the main effects appears to be simply additive, partially inconsistent with Hypothesis 2b. When cued was paired with high demand instructions, the largest mean of

clusters was observed (but not significantly larger than expected, which would have indicated an interaction) and when un-cued was paired with low demand query, the lowest mean of clusters was observed. When cued was paired with low demand, or un-cued was paired high demand query the resulting means were practically equivalent, and were in between the two previously stated highest and lowest means (see figure 2b for more details). From this, we conclude that both query level and cue increased number of clusters reported on interview-based SVF's to a similar magnitude, but without one variable moderating the effect of the other. We note that this finding is only seen with interview-based SVF's, which makes sense considering that level of demand query (low vs high) has nothing to do with the traditional method of scoring SVF (i.e., this method does not take a person's self-reported introspections into account).

The fact that HS on the traditional SVF correlated with total RFFT designs is consistent with the literature, however we were not able to show that CS correlate better with total RFFT designs, which is inconsistent with some literature (Abwender et al., 2001). Additionally, HS scored with interview methods seem to correlate better with repetition of designs on the RFFT (indicative of inadequate effort at self-monitoring), suggesting that HS reflect different neurocognitive processes depending on scoring method – a result never before reported in the literature. This finding may require further investigation, as many responses denoted as HS on interview-based SVF resulted from the participant saying "I don't know; It just came to mind; It was random" or failing to provide any connection between the words they said. For our sample of college students, some of these interview-based HS could represent true failures to cluster (consistent with traditional methods), however, as suggested by the data, interview-based HS also might represent lack of effort or willingness to invest in the task. In the case of the traditional SVF, our results still support the validity of using SVF hard switches as a measure of

executive function but further research must be done to verify the validity of using potential correlations between RFFT repeated designs and interview-based HS, as there is still a strong demand characteristic (i.e., empirically support threat to internal validity of interview-based SVF) to consider based off our previous findings. The quality of interview-based SVF data appears, then, to be a function of cognitive investment in the task, which is not so much the case under traditional scoring approaches.

As for clustering scores correlating with CVLT-II performance, neither clustering scores under the traditional method or interview-based method correlated with CVLT-II clustering, which is inconsistent with the VF literature (Abwender et al., 2001) since previous experiments *found correlations between traditionally scored* VF clustering and CVLT clustering. The only explanation we may be able to provide on this inconsistency, is that perhaps with our sample, self-monitoring deficits played a larger role in the CVLT-II, than previously conceived; hinting at possible presentation effects that may have been present in spite of the techniques employed in our methodology to randomize test order. Since this is the first study attempting to compare interview vs. traditional SVF scoring methods, how this effect interacts with newly suggested interview methods has not been previously reported. What we can mainly derive from our correlational analysis is that interview-based SVF and traditional-based SVF may be measuring different variables, based from our finding regarding interview-scored HS vs traditionally scored HS.

Limitations & Future Directions

Measures for gathering participants to take part in this study involved recruiting 104 undergraduate college students enrolled in Introduction to Psychology via an online human subject pool hosted by the Department of Psychology at College at Brockport (i.e., SONA

systems). Effectively this is a convenience sample that lacks techniques to promote the randomization of sampling. As a result, our female to male ratio is ~ 3:1 and the narrow distribution of age centers at 19.4 years old. With these sampling demographics, it is difficult to generalize the results of our study to males in general, and to females with ages far above and below 19.4 years old, and females that are not enrolled in college. Our study was also limited in the variety of tests used for correlational analysis, depending solely on the CVLT-II, RFFT, and SVF. If a wider variety of tests for semantic memory and executive function was used, then we may have been able to find better correlations for semantic clustering, and perhaps elucidate more about the validity of using interview-based SVF vs traditionally scored SVF.

As a future direction, our lab is looking to establish inter-rater reliability with verbal fluency scoring as there were four VF scorers in this study. Inter-rater reliability will be established when each traditional SVF is scored by two researchers, and each interview SVF is scored by two different researchers. As it currently stands, for each participant, the SVF has been scored once via the traditional method and once via interview method. Evidently, each SVF test would need to be scored once more using the traditional method and once more using the interview method for inter-rater reliability to be analyzed. High inter-rater reliability will be demonstrated if correlational analysis shows SVF scores are similar with high convergence across raters. Inter-rater reliability will give us the opportunity to see if one scoring method (i.e. Traditional vs Interview) is more reliable than the other, which is by definition a separate, but important component to consider next to each test's validity.

Conclusion

Overall, we do not have enough evidence to fully compare the validity and usefulness of Traditional SVF scoring methods vs. Interview-based methods, but our preliminary study suggests that these two methods may be measuring different variables, and that interview methods are susceptible to demand characteristics under Body & Muskett's Methodology. To some extent, our study presents preliminary evidence suggesting that the validity of interview-based scoring guidelines set forth by Body & Muskett (2013) may be limited, solely on the basis of how profoundly demand characteristics empirically influence participant responses; pressuring the participant to create more associations between words, which may be less representative of true cognitive function, and less free from experimenter manipulation.

References

Abwender, D. A., Swan, J. G., Bowerman, J. T., & Connolly, S. W. (2001). Qualitative analysis of verbal fluency output: Review and comparison of several scoring methods. *Assessment, 8,* 323-336

Abwender, D. A., Preston, J. N., & Steffenella, T. M., (2003). *The effect of instructional set on verbal fluency clustering and switching*. Poster session presented at the 111th Annual Convention of the American Psychological Association, Toronto.

Body, R., & Muskett, T. (2013). Pandas and penguins, monkeys and caterpillars: Problems of cluster analysis in semantic verbal fluency. *Qualitative Research in Psychology, 10,* 28 41.

Bondi, M. W., Kaszniak, A. W., Bayles, K. A., & Vance, K. T. (1993). Contributions of frontal system dysfunction to memory and perceptual abilities in Parkinson's disease. *Neuropsychology*,*7*, 89-102. doi:10.1037/0894-4105.7.1.89.

Bozikas, V. P., Kosmidis, M. H., & Karavatos, A. (2005). Disproportionate impairment in semantic verbal fluency in schizophrenia: Differential deficit in clustering. *Schizophrenia Research*, *74*, 51-59. doi:10.1016/j.schres.2004.05.001.

Binetti, G., Magni, E., Cappa, S. F., Padovani, A., Bianchetti, A., & Trabucchi, M. (1995). Semantic memory in Alzheimer's disease: An analysis of category fluency. *Journal of Clinical and Experimental Neuropsychology*, *17*, 82-89. doi:10.1080/13803399508406584.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82,* 407-428. doi:10.1037//0033-295x.82.6.407

Danysz, W., & Parsons, C. G. (2012). Alzheimers disease, β-amyloid, glutamate, NMDA receptors and memantine - searching for the connections. *British Journal of Pharmacology, 167,* 324-352. doi:10.1111/j.1476-5381.2012.02057.x

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System. Technical manual.* Psychological Corporation, San Antonio, TX.

Elvevåg, B. B., Fisher, J. E., Gurd, J. M., & Goldberg, T. E. (2002). Semantic clustering in verbal fluency: Schizophrenic patients versus control participants. *Psychological Medicine*, *32*, 909-917. doi:10.1017/S0033291702005597.

Epker, M. O., Lacritz, L. H., & Munro Cullum, C. C. (1999). Comparative analysis of qualitative verbal fluency performance in normal elderly and demented populations. *Journal of Clinical and Experimental Neuropsychology*, *21*, 425-434. doi:10.1076/jcen.21.4.425.890.

Fagundo, A. B., López, S., Romero, M., Guarch, J., Marcos, T., & Salamero, M. (2008). Clustering and switching in semantic fluency: Predictors of the development of Alzheimer's disease. *International Journal of Geriatric Psychiatry*, *23*, 1007-1013. doi:10.1002/gps.2025.

Giovannetti, T., Goldstein, R. Z., Schullery, M., Barr, W. B., & Bilder, R. M. (2003). Category fluency in first-episode schizophrenia. *Journal of the International Neuropsychological Society*, *9*, 384-393. doi:10.1017/S1355617703930049.

Heavey, C. L. (2013). Confronting the challenges of observing inner experience: The descriptive experience sampling method. In J. W. Clegg (Ed.), *Self-observation in the social sciences* (pp. 103-119). Piscataway, NJ, US: Transaction Publishers.

Heavey, C. & Hurlburt, R. (2008). The phenomena of inner experience. *Consciousness and Cognition, 17*. 798-810. doi: 10.1016/j.concog.2007.12.006.

Hedreen, J., & Folstein, S. (1995). Early loss of neostriatal striosome neurons in Huntington's disease. *Journal of Neuropathology And Experimental Neurology*, *54*, 105-120.

Hurks, P. P. (2012). Does instruction in semantic clustering and switching enhance verbal fluency in children? *The Clinical Neuropsychologist, 26,* 1019-1037. doi:10.1080/13854046.2012.708361

Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science, 310*, 116-119. doi:10.1126/science.1111709.

Martignoni, E. E., & Calandrella, D. D. (2009). Mental manifestations in Parkinson's disease. *Minerva Psichiatrica*, *50*, 27-44.

McDowd, J., Hoffman, L., Rozek, E., Lyons, K. E., Pahwa, R., Burns, J., & Kemper, S. (2011). Understanding verbal fluency in healthy aging, Alzheimer's disease, and Parkinson's disease. *Neuropsychology*, *25*, 210-225. doi:10.1037/a0021531.

Moelter, S. T., Hill, S., Ragland, J., Lunardelli, A., Gur, R. C., Gur, R. E., & Moberg, P. J. (2001). Controlled and automatic processing during animal word list generation in schizophrenia. *Neuropsychology*, *15*, 502-509. doi:10.1037/0894-4105.15.4.502

Robert, P. H., Lafont, V., Medecin, I., Berthet, L., Thauby, S., Baudu, C., & Darcourt, G. (1998). Clustering and switching strategies in verbal fluency tasks: Comparison between schizophrenics and healthy adults. *Journal of the International Neuropsychological Society*, *4*, 539-546. doi:10.1017/S1355617798466025.

Robert, P. H., Migneco, V. V., Marmod, D. D., Chaix, I. I., Thauby, S. S., Benoit, M. M., Beau, C. H., & Darcourt, G. G. (1997). Verbal fluency in schizophrenia: The role of semantic clustering in category instance generation. *European Psychiatry*, *12*, 124-129. doi:10.1016/S0924-9338(97)80200-3.

Ruff, R. M. (1996). *Ruff Figural Fluency Test*. Odessa, FL: Psychological Assessment Resources.

Quillian, M. R. (1962). A revised design for an understanding machine. *Mechanical Translation, 7,* 17-29.

Quillian, M. R. *Semantic memory*. Unpublished doctoral dissertation, Carnegie Institute of Technology, 1966. (Reprinted in part in M. Minsky [Ed.], Semantic information processing. Cambridge, Mass.: M.I.T. Press, 1968.)

Sung, K., Gordon, B., Vannorsdall, T. D., Ledoux, K., Pickett, E. J., Pearlson, G. D., & Schretlen, D. J. (2012). Semantic clustering of category fluency in schizophrenia examined with singular value decomposition. *Journal of the International Neuropsychological Society*, *18*, 565-575. doi:10.1017/S1355617712000136.

Testa, J. A., Tröster, A. I., Fields, J. A., Gleason, A. C., Salmon, D. P., & Beatty, W. W. (1998). Semantic fluency performance of patients with cortical and subcortical neurodegenerative diseases. *Aging, Neuropsychology, and Cognition*, *5*, 203-214. doi:10.1076/anec.5.3.203.612.

Tröster, A. I., Fields, J. A., Testa, J. A., Paul, R. H., Blanco, C. R., Hames, K. A., Salmon, D. P. & Beatty, W. W. (1998). Cortical and subcortical influences on clustering and switching in the performance of verbal fluency tasks. *Neuropsychologia*, *36*, 295-304. doi:10.1016/S0028-3932(97)00153-X.

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two

    components of verbal fluency: Evidence from younger and older healthy

    adults. *Neuropsychology, 11,* 138-146. doi:10.1037//0894-4105.11.1.138

Troyer, A., Moscovitch, M., Winocur, G., Leach, L., & Freedman, M. (1998). Clustering and

    switching on verbal fluency tests in Alzheimer's and Parkinson's disease. *Journal of the*

    International Neuropsychological Society, 4, 137-143

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the

    quality of preferences and decisions. *Journal of Personality and Social Psychology, 60,*

    81-192.doi:10.1037//0022-3514.60.2.181

Zakzanis, K. K., Troyer, A. K., Rich, J. B., Heinrichs, W. (2000). Component analysis of verbal

    fluency in patients with schizophrenia. *Neuropsychiatry, Neuropsychology, & Behavioral*

    Neurology, 13. 239-245.