A Re-examination of the Convergent and Divergent Validity of the Ruff Figural Fluency Test

An essay submitted in partial fulfillment of

the requirements for graduation from the

Honors College at the College of Charleston

with a Bachelor of Science in

Psychology

Shelby Stohlman

May 2016

Advisor: Thomas P. Ross, Ph.D

Abstract

The convergent and divergent validity of standard and qualitative indices of the Ruff Figural Fluency Test (RFFT; Ruff 1988) was examined using a sample of healthy undergraduates (N = 90). Participants (83.3% Caucasian; 80% female) were administered a comprehensive battery of neuropsychological tests including measures of executive functioning (EF), intelligence, working memory, motor speed and dexterity. The RFFT did not correlate with measures of working memory or motor dexterity; however, the RFFT correlated with measures of intelligence (e.g., vocabulary, block design) and with motor and visual scanning speed. RFFT performance correlated with fluency and planning domains of EF, but not with interference control and cognitive flexibility. Taken together, these findings provide mixed support for the construct validity of the RFFT as a measure of EF. Evidence for convergent validity was seen in only a subset of EF tasks, while correlations of a similar magnitude were observed on measures of intelligence (indicating modest divergent validity). Consistent with previous findings (e.g., Ross, 2014) the correlations among RFFT scoring indices did not support Lezak et al.'s (2012) model of score interpretation. Although strategic clustering was positively associated with novel design production, a negative relationship between perseverative responding and design output was not observed. Future research using more diverse samples is needed to better examine the relationship between figural fluency and intelligence, as well as the role of strategy use in effective RFFT performance.

Key Words: Ruff Figural Fluency Test, Validity, Fluency, Executive Functioning

A Re-examination of the Convergent and Divergent Validity of the Ruff Figural Fluency Test

Executive Functions (EFs) are among the most abstract and crucial behaviors that enable us to live and adapt cognitively, emotionally and socially. Descriptions of EFs include one's ability to plan and make decisions, self-monitor, establish goals, initiate and execute goal-oriented behavior, and to shift strategies when necessary (Lezak, Howieson, Bigler & Tranel, 2012; Strauss, Sherman & Spreen, 2006; Stuss & Benson, 1986). These functions are complex in nature and can be impaired by neurological and psychiatric disorders (Heilman & Valenstein, 2012; Soble, Donnell & Belanger, 2013). Although impairment can be constrained to specific facets of executive functions, central nervous system trauma or disease will often result in damage to more than one ability. Typically, EF deficiencies are indicative of frontal lobe damage; however, EFs can be affected by impairment to other parts of the brain (Lezak et al., 2012; Stuss & Benson, 1986). Although there is no model for EFs that is universally agreed upon, Lezak et al. (2012) contend that executive functions can be divided into five major components: volition, planning and decision making, purposive action, effective performance and self-regulation.

According to Lezak et al. (2012), *volition* is one's ability to act intentionally; it requires the ability to formulate a goal, motivation and self-awareness. If a person lacks volition, he or she may still be able to perform complex behaviors, but only if he or she is given explicit and detailed instructions. One is also unable to contemplate future events, assume responsibilities and consider abstract goals. There are only a few formal tests that examine one's capacity for volition. The examiner must take into account the patient's responses as well as reports given to him or her by those who are close the patient. The Iowa Scales of Personality Change (ISPC) consists of a set of questions that are answered by someone who is close to the patient and has known him or her before and after the brain injury, providing the examiner with a quantification of change in the patient's behavior. The ISPC examines the following volitional constructs: social withdrawal, apathy, lack of initiative, dependency, insensitivity and social inappropriateness (Lezak et al., 2012).

Planning and decision making abilities allow us to assess what needs to be done prior to initiating a behavior to achieve a goal. In order to plan, one must take into account circumstantial changes, environmental relationships and understand alternative choices while considering the pros and cons. There are also relatively few formal tests that assess planning and decision making; however, the manner in which the patient approaches various tests provides examiners with some insight into how the patient plans and makes decisions. Some of the tests and tools that are used to assess planning and decision making include the following: tower tasks, storytelling tasks, questioning the patient, the patient's use of space in drawings, the Iowa Gambling Test and maze tests (Lezak et al., 2012). Tower task paradigms require participants to place disks onto pegs to build a specific design by using the least amount of moves, while adhering to rules or constraints. Tower tasks measure one's competency for planning, decision making and adherence to constraints. Patients with frontal lobe pathology exhibit a lack of planning, poor judgment, impulsivity and inefficiency on this task (Strauss et al., 2006).

Purposive action involves the execution of a previously planned, self-serving activity. This requires a person to elicit complex sequences of behavior in a cohesive manner. Non-routine and novel activities are at the core of purposive action. These are far more susceptible to impairment than routine, automatic tasks. Purposive acts are intentional; they are controlled by the executor. The ISPC measures the following components of purposive action in everyday tasks: lack of persistence, perseveration and lack of stamina (Lezak et al., 2012).The Tinkertoy Test presents a patient with an opportunity to plan and independently execute a complex activity using a Tinkertoy set to make anything he or she wants.

Effective performance encompasses one's ability to monitor, change and improve aspects of his or her performance. Many people who suffer frontal lobe impairment are unable to correct their mistakes and perform haphazardly without exhibiting any evidence of planning. Performance is greatly affected by self-regulation. Those who perform unsuccessfully are often unable to understand their mistakes (Lezak et al., 2012). The Wisconsin Card Sorting Task is sensitive to deficits in monitoring and changing performance, particularly in the case of frontal-lobe impairment. This task requires the participant to sort cards according to a rule that the experimenter defines. The subject is not told what this rule is, but is informed whether his or her responses are correct or incorrect; this requires the participant to learn through trial and error. Then, after a certain period of time, the experimenter changes the sorting rule without warning; requiring the participant to adapt and shift his or her strategy. This task requires mental flexibility, as the participant must switch from following one rule to following another. Patients with frontal lobe lesions tend to make perseverative errors and often keep applying the initial rule to the new situation, despite being informed that his or her responses are incorrect (Gazzaniga, Ivry & Mangun, 1998).

Self-regulation includes one's ability to be productive and to switch one's mode of thinking. Reduced productivity can be accounted for by a gap between motivation and action. Many patients who lack self-regulation may talk about "doing things," but never actually do them. In patients with brain injuries, there is often a large decrease in productivity due to slower responding. Mental flexibility refers to one's ability to change a thought pattern or strategy in accordance to changing circumstances (Lezak et al, 2012). Mental flexibility is often assessed using the Alternate Uses Test, which requires both convergent and divergent thinking. This task requires subjects to write down as many ways to use common objects as they can generate. When subjects provide an obvious use for an object, such as using a brick to build a house, they are demonstrating convergent thinking; when subjects provide an uncommon use for an object, such as using a brick as a chair, they are demonstrating divergent thinking. Those who are mentally flexible are able to partake in more divergent thinking; in contrast, those with frontal-lobe lesions tend to state the typical use of the object and have trouble utilizing divergent thinking. D-KEFS Color-Word Interference Test (Delis, Kaplan & Kramer, 2001) also measures mental flexibility and the ability to inhibit previously learned responses. In this task, participants are first asked to name colored blocks; then they are asked to read the names of colors written in black ink; then the participants are asked to name the color of the ink that a word has been printed in; lastly, the participant must switch between naming the ink colors and reading the names of colors (Strauss et al., 2006).

Self-regulation encompasses various types of fluency including verbal fluency, categorical fluency, action fluency, writing fluency and design fluency. For example, when asked to name the color of ink in which words are printed, the patient must self-regulate by inhibiting a habitual response in favor of a novel, atypical manner of responding. Certain fluency tests that assess mental flexibility, self-regulation and response times have been useful in assessing brain damage in patients, particularly in the frontal lobes (Ruff, Light & Evans, 1987). Fluency is a broad cognitive ability; the comprehensive assessment of this construct requires different tests that examine various types of fluency. Verbal fluency tends to be associated with left hemisphere functions, while figural fluency tends to be associated with the right hemisphere (Kraybrill & Suchy, 2008; Ruff, 1988). The COWAT assesses verbal fluency by requiring participants to say as many words as they can think of that begin with a certain letter during a 60-second trial (Strauss et al., 2006). Due to the hemispheric specialization of verbal and nonverbal abilities, different tests have been designed to assess spatial and figural fluency (Strauss et al., 2006).

The Ruff Figural Fluency Test (RFFT; Ruff, 1988) was originally proposed as a measure of nonverbal fluency, analogous to measures of verbal fluency. The RFFT's purpose is to examine one's nonverbal capacity for mental flexibility, divergent thinking and the executive coordination of sustained productivity (Ruff, 1988). The test contains five parts, each containing a different pattern of dots. In part 1, the dots are presented concentrically. Parts 2 and 3 also utilize the same concentric presentation as part 1, but each contains various stimuli that serve as distractors. Parts 4 and 5 use a different arrangement of dots without including any distractors. The client is presented with three squares which contain a total of five dots each. He or she must connect two or more dots by always using straight lines. The goal is to make as many figures as possible, but each response must be unique. Therefore, the respondent must address the goal to generate numerous designs while avoiding repetitions of a previous response. The RFFT imposes a 60 second time constraint for each of its five parts (Ruff, 1988).

Scoring the RFFT is based on production (novel design output) and quality. The *production score* is based on the total number of unique patterns produced. The number of *perseverative errors* (design repetitions) made by the client are counted, and subtracted from the total number of designs to yield the novel output. The examiner then calculates the *error ratio score* by dividing the number of unique designs by the number of perseverative errors. Ruff (1988) developed qualitative scores to identify rotational and or enumerative strategies a client may use in order to maximize novel design output on the RFFT. Lezak et al. (2012) contend that "generally, the greatest productivity with fewest perseverations is achieved by persons who quickly develop and then maintain a strategy so that each square no longer calls for a unique solution but rather the pattern for a long series of squares that has been predetermined by the strategy" (p. 700). According to Ruff, production strategy scores can be used to assess whether deficiencies in design production are due to initiation or planning inabilities. Problems with initiation would be shown by a low output of designs; whereas deficiencies in planning abilities would be seen by a substantial amount of output, along many perseverative errors and failure to implement strategies (Ruff, 1988; Lezak et al., 2012).

Various studies have assessed the reliability and validity of the RFFT. *Interrater reliability*, also known as interscorer reliability, refers to the degree of agreement or consistency between two or more scorers on a particular measure (Cohen, Swerdlik & Sturman, 2010). Ross, Foard, Hiott and Vincent (2003) reported good to excellent interscorer agreement (r_{icc} totals ranging from .79 to .95) for RFFT indices. Higher reliability was observed for enumerative strategies relative to rotational strategies (Ross et al., 2003). Other studies have reported good to excellent interscorer reliability for the number of unique designs and other qualitative indices that assess strategy utilization on the RFFT (see Strauss et al, 2006).

Test-retest reliability refers to the consistency of scores generated from the same people on two different administrations of the same test (Cohen et al., 2010). Studies have reported acceptable score stability for the number of novel designs produced (e.g., $r = .8$); however, the stability of qualitative indices is modest (e.g., $r = .4$) and perseverations are poor ($r = .3$ to .4) (see Ross, 2014; Strauss et al., 2006). Psychometric studies have revealed significant *practice effects.* For example, Ross (2014) reported an average gain of 15 designs occurred when participants were retested using a short interval of six to seven weeks. Studies employing longer test-retest intervals have demonstrated significant practice effects after a one-year period (see Strauss et al., 2006).

Multiple studies have examined the criterion-related validity of the RFFT. *Criterion-related validity* is established by demonstrating that test scores can predict a participant's status or standing on a known criterion. (Cohen et al., 2010). Ruff, Light & Evans (1987) contend that the RFFT is sensitive to lesions of the right-anterior hemisphere. Patients with right frontal or right fronto-central lesions performed significantly worse than those with left-frontal lesions. Also, anterior lesions affected performance more than posterior-cerebral lesions. In a review of the psychometric studies, patients with traumatic brain injuries produced less designs and were more likely to perseverate (Strauss et al., 2006). These findings indicate that the RFFT is highly sensitive to the pathology of brain areas believed to mediate and enable executive functions.

Examinations of the RFFT's construct validity are few in number and have focused on models of score interpretation in addition to the relationship between the RFFT and other measures. Using a large sample of healthy college students ($N = 102$), Ross (2014) examined the psychometric properties of the RFFT, and in doing so, evaluated Lezak et al.'s (2012) model for score interpretation. This investigation also sought to better understand the contribution of strategy use on effective RFFT performance. Given the premises offered by Lezak and her colleagues, Ross (2014) hypothesized that a positive correlation between production strategy indices and unique design output would result, as well as a negative correlation between production strategy indices and perseverative errors. Ross reported a correlation of $r = .46$ between novel design output and strategy usage. Correlations between qualitative indices of strategy use and perseverative responses were non-significant, with the exception of percent of designs included in strategies, which was negatively correlated with perseverations ($r = .26$). The switching scores were positively associated with novel design output; however, these switching scores did not indicate a shift in cognitive flexibility because the participants were switching between non-clustered designs, as opposed to clustered designs. In contrast to Lezak et al.'s (2012) assertions, perseverations (i.e., repetitions) were positively correlated with novel design output. Therefore, the use of strategies was not necessary for a participant to perform effectively. Moreover, studies by Ross (2014) and Gardner et al. (2013) show that participants who employ strategies can still perform within the impaired range, while those who do not employ strategies can perform at normal to high average levels. Taken together, these findings suggest that effective performance on the RFFT is neither dependent on strategy use, nor on avoiding perseverative errors; however, both of these skills allow for a better score. At present, it is unclear as to whether strategic responding reflects a preferred cognitive style or some indication of other cognitive abilities, as assessed by the RFFT (Gardner, Vik & Dasher, 2013; Ross, 2014).

Convergent validity occurs when scores on a test undergoing construct validation correlate highly in the predicted direction with scores on previously validated measures of the same construct; *divergent validity* occurs when there is little relationship between test scores and other variables that were developed to assess different constructs (Cohen et al., 2010). Ross

(2014) examined the convergent and divergent validity of RFFT scores in a sample of healthy young adults (*N* = 102). Ideally, low coefficients with measures of intelligence, letter fluency, verbal learning and working memory would provide support for the divergent validity of the RFFT. In contrast, high correlations between the RFFT and other, putative measures of executive functioning would provide evidence for convergent validity. Ross (2014) found that the production of novel designs correlated with other known measures of executive functioning, providing some support for the convergent validity of the RFFT. However, smaller and less compelling correlations were observed between strategy scores and other EF measures.

Ross (2014) also reported mixed evidence for the RFFT's divergent validity. The RFFT moderately correlated with select subtest scores of the WAIS-III and working memory. WAIS-III Vocabulary subtest and North American Adult Reading Test (NAART) performance was linked to strategy utilization, but was not associated with novel design production. Working memory for verbal indices did not correlate with any facet of the RFFT. Nonverbal working memory measures correlated with strategy use but did not correlate with unique design output (Ross, 2014). Performance on the Block Design correlated strongly with design output and qualitative measures of switching and clustering. Taken together, these findings were consistent with Ruff's (1988) position that the RFFT is more sensitive to right hemisphere functions (Ruff, 1988).

Despite having excellent criterion-related validity, the findings of Ross (2014) and others (e.g., Gardner et al., 2013) indicate limited support for the construct validity of RFFT scores. Strauss et al. (2006) note that "executive functions" are not the sole predictors of effective performance on the complex EF tasks used by clinicians. Performance is impacted by age, education level, IQ and practice effects. They also report that correlations between the RFFT and other figural fluency tasks vary by condition. Performance on the RFFT has been consistently linked to semantic fluency (i.e., category fluency) performance rather than phonemic fluency performance (i.e., letter fluency), which is believed to require more executive control (Abwender, Swan, Bowerman & Connolly, 2001). RFFT scores also correlated with motor speed and cognitive measures (Strauss et al., 2006). Ultimately, the mixed findings for the validity of the RFFT warrant further study. The present study sought to investigate the convergent and divergent validity of the RFFT in a sample of healthy college students. More specifically, the contributions of intelligence, working memory, motor performance and executive functioning to effective performance were examined.

Given previous findings on the construct validity of the RFFT, the following hypotheses were examined. While the RFFT is not an IQ test, it was anticipated that the RFFT will correlate with certain subtests of the Wechsler Scales. Specifically, it was hypothesized that the RFFT would correlate with the perceptual organization, processing speed and working memory subsections, but not with verbal abilities (e.g. vocabulary). Additionally, it was hypothesized that RFFT scores would correlate with scores for tests of motor speed and dexterity. Finally, the hypothesis that the RFFT would correlate with measures of EF was examined. Such correlations would indicate convergent validity for the RFFT as a measure of executive functioning. A diverse group of executive functioning measures were examined to determine which executive functions are tapped by the RFFT. Previous research has shown that verbal and nonverbal executive functioning are dissociable (see Shallice & Burgess, 1991); therefore, a low correlation between the RFFT and COWAT were anticipated. No specific predictions about the associations of the RFFT with other measures of executive functioning were offered.

Method

Participants

After obtaining the approval of the institutional review board, undergraduate students (*N* = 105) were recruited from introductory psychology courses at a medium-sized southeastern university. Participants received no compensation for taking part in the study; however, they did receive credit towards course requirements. After obtaining their informed consent, participants were first asked to fill out a questionnaire about their health history and demographic background. Participants were excluded from the study if they endorsed a history of neurological disorder(s) (e.g., TBI, epilepsy and cerebral palsy). Additionally, participants were excluded if they performed below the suggested cutoff on a measures of test-taking effort: specifically, a score of 14 or higher on the Dot-Counting Test (Boone, Lu & Herzberg, 2002). Finally, participants were excluded for failure to attend part two of the study. Of the initial sample of 105 participants, two were excluded for neurological disorders, six were eliminated for a DCT error score of 14 or higher and eight were excluded for missing part two of the study (note that some of the participants who were excluded may have met multiple exclusion criteria).

Of the remaining sample of 90 participants, 83.3% identified themselves as Caucasian, 6.7% as Hispanic, 3.3% as African American and the remaining 6% reported other ethnic identities. Eighty percent of the participants identified themselves as female and 20% identified themselves as male. When asked about handedness, 86.7% claimed right-handedness, 7.8%

claimed left-handedness and 5.6% reported ambidextrousness. The estimated mean Full Scale IQ of the sample obtained using the National American Adult Reading Test was 103.6.

Materials

The RFFT was used to examine nonverbal fluency and to explore the construct validity of this measure. More specifically, the relationship between RFFT performance and general intelligence, working memory, EF, motor skills, processing speed and perceptual organization were assessed to examine the convergent and divergent validity of the RFFT. Verbal abilities were assessed using the number of correct responses on the NAART (Blair & Spreen, 1989) and the total raw score on the Vocabulary subtest of the WAIS-IV (Wechsler, 2008). Perceptual-organizational abilities were examined using the total raw scores of the Block Design (Wechsler, 2008) and Matrix Reasoning (Wechsler, 2008) subtests of the WAIS-IV. Measures of working memory included the raw scores of the Symbol Span subtest of the WMS-IV (Wechsler, 2009) and Digit-Span Sequencing subtest of the WAIS-IV (Wechsler, 2008).

Measures of EF included the COWAT number of correct responses across 60-second trials and the number of clusters generated (Benton, Hamsher & Sivan, 1994), number of correct responses across 60-second trials using animals, vegetables and fruits as stimuli in the Category Fluency Test (Goodglass, Kaplan & Barresi, 2000), total time taken to complete the interference and interference-switching conditions of the Color-Word Interference test (Delis, Kaplan & Kramer, 2001), total time to complete the Trail Making Test Parts A and B (Reitan, 1986), the total achievement and planning scores on the Tower Test (Delis, Kaplan & Kramer, 2001) and auto detection variables of the 2&7 Test (Ruff & Allen, 1996).

Motor speed was assessed using the dominant hand completion time in seconds of the Grooved Pegboard Test (Matthews & Klove, 1964) and the average dominant hand taps across five trials in the Finger Tapping Test (Reitan & Wolfson, 1993). Finally, effort was measured using the Dot-Counting Test error score (Boone, Lu & Herzberg, 2002) and processing speed was assessed using the total raw score obtained on the Coding subtest of the WAIS-IV (Wechsler, 2008).

Procedures and Scoring

All participants were recruited from introductory level Psychology courses and received course credit for their participation. Participants self-registered and signed up for each testing appointment through the online experimental management system (SONA systems, Ltd, Version 2.72; Tallinn, Estonia). Once each student logged on, this system presented any active studies the student could participate in a randomized order. This procedure minimized the risk that many students would sign up from the same class and share information about the study. Upon arrival, each participant was given an informed consent form, which he or she was instructed to read and sign if he or she agreed to the terms of the study. After obtaining each participant's informed consent, he or she was given a health and demographic information questionnaire to complete. Then, the participants were given the various neuropsychological measures stated above.

This study required participants to complete two separate testing sessions. In session 1 of the study, the following measures were administered: RFFT, Dot-Counting Test, COWAT, 2 &7 test, Matrix Reasoning, Digit-Span Sequencing, Vocabulary and Coding subtests of the WAIS-IV, Action-Verb Fluency (Piatt, Fields, Paolo & Troister, 1990; Woods et al., 2005), Grooved Pegboard test and Color-Word Interference Test. In session 2 of the study, the following measures were administered: COWAT, Finger Tapping Test, Block Design, Symbol Span, Trail Making Test Parts A and B, Tower Tests, Category Fluency Test and the NAART.

All tests were administered and scored according to the procedures in the published manuals. When scoring COWAT responses for qualitative indices (e.g. clusters and switches), the present study used procedures developed by Troyer and colleagues (1997). The order of tests administered was counterbalanced to control for possible order effects. The testing sessions lasted approximately 55 to 65 minutes. Before beginning test administration, each of the participants was told that he or she would be completing a variety of tasks, some of which would be more difficult than others. He or she was instructed to do his or her best and work as quickly and accurately as possible during all of the timed tasks.

All tests were administered by advanced psychology majors who received training by a Ph.D level psychologist. Two raters scored the RFFT protocols for traditional and qualitative indices. Each rater received the scoring procedures stated below and scored several practice protocols until near perfect agreement was achieved. All protocols were then scored by both raters. The raters compared scored protocols for all cases, and if discrepant, carefully inspected the manual to determine the correct score for data analysis.

The present study examined standard and qualitative indices of the RFFT. The standard indices consist of the number of novel designs, the number of perseverations and the error ratio (Ruff, 1988). Qualitative indices included *enumerative* and *rotational production strategies*, mean cluster size and the percentage of designs in clusters. Ruff (1988) states that if a strategy is *rotational* the following rules must apply: the same number of dots or lines must be used; the rotation must be systematic and one or more dots must be present throughout each configuration.

An *enumerative* strategy is scored according to the following rules: the same figure must be present in each subsequent design and there must be a systematic adding or subtracting of one line from the former figure to the latter figure (Ruff, 1988).

Mean cluster size and percent of designs in clusters were calculated using the procedure by Ross and colleagues (2003). In order to calculate mean cluster size, one would first calculate the sum of each cluster. Once these have been added together, the sum is divided by the total number of *enumerative* and *rotational* clusters. For example, if a participant produced two clusters having sizes of 1 and 3, the mean cluster size would be 2 or (1 + 3)/2. The *percentage of designs in strategies* score was calculated by taking the total number of designs within clusters and dividing that number by the total number of designs produced. For example, if a participant produced 12 designs and 6 of these designs were included in clusters, the percentage of designs in clusters would be (6/12) or 50%.

Results

Data were first inspected for outliers and other violations of univariate normality. Estimates of skewness and kurtosis were within acceptable parameters. The means and standard deviations for RFFT indices are displayed in Table 1. The present study reported a mean total novel design output of 51.51. The data obtained are roughly commensurate with other studies that used parts 1, 4 and 5 of the Ruff Figural Fluency Test on a healthy sample of college students (see Ross, 2003; Ross, 2014).

The correlations among various Ruff Figural Fluency Test indices are shown in Table 2. The total number of perseverative responses correlated with the total novel design output (r = .321, $p < .002$). The use of strategic clusters was associated with total novel design output, meaning that the total novel design output correlated with both the total number of enumerative clusters (r = .32, p < .002) and the total number of rotational clusters (r = .48, p < .001). Total switching was strongly associated with total novel design output (r = .79, p < .001).

The means and standard deviations for other neuropsychological measures are shown in Tables 3 and 4. As can be seen in Table 3, the estimated full scale IQ of the current sample (N = 90) using the NAART was 103.62.

The correlations between indices of the RFFT and other neuropsychological measures are shown in Tables 5 and 6. RFFT mean total novel design output correlated with the Vocabulary subtest of the WAIS-IV (r = .24, p < .026), the COWAT total number correct (r = .30, p < .01), the Block Design Raw Score subtest of the WAIS-IV (r = .23, p < .032), the Finger Tapping Test (r = .28, p < .009), the 2&7 Auto Detection Accuracy and the Trail Making Test Part A (r = .44, p < .001). The RFFT total novel design output did not correlate with the Grooved Pegboard Test or the Matrix Reasoning total raw score. For the qualitative indices, there were very few significant correlations overall, except for the COWAT, where slightly larger r values were observed for the strategy scores, relative to total design output.

Discussion

The purpose of this study was to reexamine the convergent and divergent validity of the Ruff Figural Fluency Test and evaluate the findings according to Lezak's model of score interpretation (Lezak et al., 2012).

The data obtained in this study are roughly commensurate with other studies that used parts 1, 4 and 5 of the Ruff Figural Fluency Test on a healthy sample of college students (Ross, 2003; Ross, 2014). For example, Ross (2014) reported a mean total novel design output of 55.7.

This full scale IQ estimate is roughly commensurate with previous studies. For example, Ross et al. (2003) used the NAART and estimated a full scale IQ of 108.1 for their sample of healthy college students (N = 90).

There was mixed evidence for Lezak's model of score interpretation. According to Lezak, the greatest novel design output with the fewest number of perseverations is achieved by those who develop and maintain a strategy. The findings of this study indicate that clustering is positively correlated with one's number of novels responses; however, clusters were not negatively correlated with perseverative responses as implied by Lezak et al. (2012). This suggests that people can do well on the RFFT without having to employ strategies.

There was mixed support for the hypothesis that the RFFT would correlate with measures of perceptual organization. While the RFFT total novel design output and use of rotational clusters did correlate with the Block Design subtest of the WAIS-IV, it did not correlate with the Matrix Reasoning subtest of the WAIS-IV. This suggests that the RFFT may be more strongly associated with the Block Design subtest as it is a test requiring speed of information processing. The Coding subtest of the WAIS-IV was associated with both the number of perseverative responses on the RFFT as well as the number of total switches on the RFFT. This finding is consistent with the position that the RFFT is related to the Block Design subtest because of the speed requirement that is common to both tasks, rather than perceptual organization per se. This hypothesis was examined post-hoc with a partial correlation analysis. After controlling for the variance shared with the WAIS-IV Coding, the correlation between the RFFT and Block Design remained significant, suggesting that the relationship between the RFFT and Block Design is complex and needs to be examined further.

The hypothesis that the RFFT would correlate with measures of working memory was not supported. No indices of the RFFT correlated with the Digit Span Sequencing subtest of the WAIS-IV, or the Symbol Span subtest of the WMS-IV. Additionally, the hypothesis that the RFFT would not correlate with verbal abilities was also not supported. The RFFT correlated with the Vocabulary subtest of the WAIS-IV. This suggests that general intelligence may influence RFFT scores and that verbal and nonverbal domains of fluency may not be dissociable in a healthy sample.

There was mixed support for the hypothesis that the RFFT would correlate with scores for tests of motor speed and dexterity. While the RFFT correlated with the Finger Tapping Test, it did not correlate with the Grooved Pegboard test. These findings suggest that RFFT performance requires motor speed more than dexterity. This may be because the RFFT only requires subjects to perform basic motor skills (i.e., to connect dots with straight lines) and does not require much fine-motor control.

The RFFT correlated with certain EF tests, but not others. The Ruff 2 & 7 Auto Detection Accuracy score correlated with both the RFFT novel design output and RFFT total switching scores. This finding supports the position that the RFFT imposes visual scanning speed as a task requirement. This may be due to the need to continuously scan previous responses to avoid perseverative responding; a hypothesis that should be explored in future studies.

The COWAT total number of correct responses correlated with RFFT novel design output, RFFT cluster scores and RFFT percentage of designs in strategies. These results indicate the RFFT best measures the fluency and production facets of EF. It is interesting to note that qualitative indices of strategic responding were more highly correlated with the COWAT than total design output, suggesting that qualitative indices require further examination. Additionally, these results provide further support to suggest that verbal and nonverbal fluency may not be dissociable in a normal healthy sample. The Tower Test total first move time correlated with the RFFT total number of switches, indicating that the RFFT may also tap into the planning component of EF. The RFFT did not correlate with D-KEFS Color Word Interference Test, or with the Trail Making Test. These findings indicate that the RFFT may not be a direct measure of cognitive flexibility, inhibition and control.

Overall, evidence for convergent validity was seen in only a subset of EF tasks, while correlations of a similar magnitude were observed on measures of intelligence, indicating modest divergent validity. Correlations with estimates of verbal and nonverbal intelligence suggest that intellectual functioning influences RFFT scores. This finding suggests that future normative studies should present results by education and IQ functioning. As an EF measure, the RFFT correlated with measures of fluency, planning and self-monitoring (visual scanning), but not with measures of cognitive flexibility or control of inhibition. These results suggest that EF is a multifaceted construct and that the RFFT relates to some, but not all of these facets. Taken together, these findings provide mixed support for the construct validity of the RFFT as a measure of EF.

There were several limitations to this study. First, the sample used in this study was not representative of the larger population with regard to age, gender, ethnicity and education. While the RFFT was examined as a measure of EF, several validated EF tests (e.g., Wisconsin Card Sorting Task, Multiple Errands Test) were not included in the present neuropsychological battery. Also, the current study would have been strengthened by using the entire WAIS-IV to better examine the relationship between the RFFT and intelligence. Therefore, future studies should use more diverse samples, include additional measures of EF and include the entire WAIS-IV battery to further explore the construct validity of the RFFT.

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.

Benton, A. L., Hamsher, K., & Sivan, A. B. (1994) *Multilingual aphasia examination.* (3rd ed.) Iowa City, IA: AJA Associates.

Blair, J. R., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test. *The Clinical Neuropsychologist, 3*, 129-136.

Boone, K. B., Lu, P., & Herzberg, D. S. (2002). *The dot counting test.* Los Angeles, CA: Western Psychological Services.

Cohen, R. J., Swerdlik, M. E., & Sturman, E. D. (2010). Reliability. In D. B. Hash (Ed.), *Psychological testing and assessment: An introduction to tests and measurement* (pp. 145-180). New York, NY: McGraw Hill.

Cohen, R. J., Swerdlik, M. E., & Sturman, E. D. (2010). Validity. In D. B. Hash (Ed.), *Psychological testing and assessment: An introduction to tests and measurement* (pp. 181-210). New York, NY: McGraw Hill.

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

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (1998). Executive functions and frontal lobes. In H. Boyer, S. Mosberg, R. Mixter & M. T. Kelly (Eds.), *Cognitive neuroscience: The biology of the mind* (pp. 423-464). New York, NY: W.W. Norton & Company, Inc.

Gardner, E., Vik, P., & Dasher, N. (2013). Strategy use on the Ruff Figural Fluency Test. *The Clinical Neuropsychologist, 27,* 470 - 484.

Goodglass, H., Kaplan, E., & Barresi, B. (2000). *The Boston Diagnostic Aphasia Examination (BDAE-3)* (3rd ed.). San Antonio, TX: Pearson.

Heilman, K. M., & Valenstein, E. (2012). *Clinical neuropsychology.* New York: Oxford University Press.

Kraybill, M. L., & Suchy, Y. (2008). Evaluating the role of motor regulation in figural fluency: Partialing variance in the Ruff Figural Fluency Test. *Journal of Clinical and Experimental Neuropsychology, 30,* 903-912.

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). New York, NY: Oxford University Press.

Matthews, C. G., & Klove, K. (1964). *Instruction manual for the Adult Neuropsychology Test Battery.* Madison, Wisc.: University of Wisconsin Medical School.

Piatt, A. L., Fields, J. A., Paolo, A. M., & Troster, A. I. (1999). Action (verb naming) fluency as an executive function measure: Convergent and divergent evidence of validity. *Neuropsychologia, 37,* 1499-1503.

Reitan, R. M. (1986). *Trail Making Test manual for scoring and administration.* Tucson, AZ: Reitan Neuropsychological Laboratory.

Reitan, R. M., & Wolfson, D. (1993). *The Halstead-Reitan Neuropsychological Test Battery: Theory and interpretation.* Tucson, AZ: Neuropsychology Press.

Ross, T. P., Foard, E. L., Hiott, F. B., & Vincent, A. (2003). The reliability of production strategy scores for the Ruff Figural Fluency Test. *Archives of Clinical Neuropsychology, 18,* 879-891.

Ross, T. P. (2014). The reliability and convergent and divergent validity of the Ruff Figural Fluency Test in healthy young adults. *Archives of Clinical Neuropsychology, 29,* 806-817.

Ruff, R. M. (1988). *Ruff Figural Fluency Test professional manual.* Odessa, FL: Psychological Assessment Resources Inc.

Ruff, R. M., Allen, C. C. (1996). *Ruff 2 & 7 Selective Attention Test professional manual.* Odessa, FL: Psychological Assessment Resources Inc.

Ruff, R. M., Light, R. H., & Evans, R. (1987). The Ruff Figural Fluency Test: A normative study with adults. *Developmental Neuropsychology, 3,* 37-51.

Shallice, T., & Burgess P. (1991). Higher-order cognitive impairments and frontal-lobe lesions in man. In H. S. Levin, H. M. Eisenberg, & A. L. Benton (Eds.), *Frontal lobe function and dysfunction* (pp. 125 - 128) New York: Oxford University Press.

Soble, J. R., Donnell, A. J., & Belanger, H. G. (2013). TBI and nonverbal executive functioning: Examination of a modified design fluency test's psychometric properties and sensitivity to focal frontal injury. *Applied Neuropsychology: Adult, 20,* 257-262.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd. ed.). New York: Oxford University Press.

Stuss, D. T., & Benson, D. F. (1986). *Executive system* (pp. 204-216). New York: Raven Press.

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.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition*. San Antonio, TX: Pearson.

Wechsler, D. (2009). *Wechsler Memory Scale–Fourth Edition*. San Antonio, TX: Pearson.

Woods, S. P., Scott, J. C., Sires, D. A., Grant, I., Heaton, R. K., & Troster, A. I. (2005). Action (verb) fluency: Test-retest reliability, normative standards, and construct validity. *Journal of the International Neuropsychological Society, 11*, 408-415.

Table 1

RFFT Means and Standard Deviations for Test Samples

Index	Mean	*SD*
RND	51.51	14.31
RPR	3.41	4.07
RER	0.06	0.07
RTC	3.47	2.76
REC	1.41	2.10
RRC	2.06	2.02
RMC	1.42	0.88
RPS	21.95	17.77
RTS	42.08	13.39

Note. RND = RFFT Novel Designs; RPR = RFFT Perseverative Responses; RER = RFFT Error Ratio; RTC = RFFT Total Strategic Clusters; REC = RFFT Total Enumerative Clusters; RRC = RFFT Total Rotational Clusters; RMC = RFFT Mean Cluster Size; RPS = RFFT Percentage of Designs in Strategies; RTS = RFFT Total Switches.

Table 2

Correlations among RFFT Scoring Indices

Index	RND	RPR	RER	RTC	REC	RRC	RPS	RTS
RND	1	0.32**	0.14	0.59**	0.03**	0.48**	0.42**	0.79**
RPR	0.32**	1	0.96**	0.11	0.10	0.05	0.01	0.57**
RER	0.14	0.96**	1	0.01	0.04	-0.04	-0.07	0.44**
RTC	0.59**	0.11	0.01	1	0.68**	0.65**	0.94**	0.06
REC	0.32**	0.10	0.04	0.68**	1	-0.11	0.67**	-0.07
RRC	0.48**	0.05	-0.04	0.65**	-0.11	1	0.59**	0.16
RPS	0.42**	0.01	-0.07	0.94**	0.67**	0.59**	1	-0.14
RTS	0.79**	0.57**	0.44**	0.06	-0.07	0.16	-0.14	1

Note. RND = RFFT Novel Designs; RPR = RFFT Perseverative Responses; RER = RFFT Error Ratio; RTC = RFFT Total Strategic Clusters; REC = RFFT Total Enumerative Clusters; RRC = RFFT Total Rotational Clusters; RPS = RFFT Percentage of Designs in Strategies; RTS = RFFT Total Switches.

* $p < .05$

** $p < .01$

Table 3

Selected IQ and Working Memory Scores

Index	Mean	*SD*
NAART	103.62	6.76
VOC	26.17	3.70
BD	47.31	9.23
MR	18.06	3.83
COD	80.98	11.00
DSS	9.73	1.83
SS	26.04	4.65

Note. NAART = National American Adult Reading Test; VOC = WAIS-IV Vocabulary subtest; BD = WAIS-IV Block Design subtest; MR = WAIS-IV Matrix Reasoning subtest; COD = WAIS-IV Coding subtest; DSS = WAIS-IV Digit-Span Sequencing subtest; SS = WMS-IV Symbol Span Total Raw Score.

Table 4

Means and Standard Deviations for Tests of Executive Function and Motor Speed

Index	Mean	*SD*
GPT	68.09	8.97
FTT	43.41	7.45
AVF	19.97	4.05
SI	46.53	10.40
SIS	54.42	10.40
TMA	22.14	6.63
TMB	52.20	14.79
SF	46.60	9.31
TTA	17.91	2.97
TTP	3.09	1.41
CTC	37.31	8.43
CNC	9.54	2.80
RAD	49.28	7.34
RCS	42.00	5.26

Note. GPT = Grooved Pegboard Test dominant hand time in seconds; FTT = Finger Tapping Test; AVF = Action-Verb Fluency; SI = Stroop Inhibition Time to Complete; SIS = Stroop Inhibition Switching Time to Complete; TMA = Trail-making test, Part A Time in seconds; TMB = Trail-making test, Part B Time in seconds; SF = Semantic Fluency Total Correct Words; TTA = Tower Test Total Achievement Score; TTP: Tower Test Planning Score; CTC = COWAT Total Correct; CNC = COWAT Number of Clusters; RAD = Ruff 2 & 7 Auto Detection Accuracy Score; RCS = Ruff 2 & 7 Controlled Search Accuracy Score.

Table 5

RFFT correlations with IQ and Motor Speed tests

Index	RND	RPR	RER	RTC	REC	RRC	RPS	RTS
NAART	0.21*	0.03	-0.04	0.27**	0.10	0.27*	0.24*	0.07
VOC	0.24*	0.12	0.05	0.14	-0.01	0.20	0.08	0.20
BD	0.23*	-0.01	-0.04	0.16	-0.03	0.24*	0.14	0.15
MR	0.10	0.03	-0.01	0.05	0.02	0.04	0.01	0.07
COD	0.18	0.22*	0.16	-0.05	0.04	-0.10	-0.15	0.27*
DSS	0.16	-0.18	-0.20	0.01	-0.08	0.10	-0.06	0.12
FTT	0.28**	0.09	0.05	0.31**	0.11	0.30**	0.27*	0.11
GPT	-0.12	-0.02	-0.01	-0.01	0.12	-0.14	0.01	-0.11
SS	0.06	0.03	0.03	0.01	-0.13	0.14	-0.04	0.09

Note. RND = RFFT Novel Designs; RPR = RFFT Perseverative Responses; RER = RFFT Error Ratio; RTC = RFFT Total Strategic Clusters; REC = RFFT Total Enumerative Clusters; RRC = RFFT Total Rotational Clusters; RPS = RFFT Percentage of Designs in Strategies; RTS = RFFT Total Switches; NAART = National American Adult Reading Test; VOC = WAIS-IV Vocabulary subtest; BD = WAIS-IV Block Design subtest; MR = WAIS-IV Matrix Reasoning subtest; COD = WAIS-IV Coding subtest; DSS = WAIS-IV Digit-Span Sequencing subtest; FTT = Finger-Tapping test; GPT = Grooved Pegboard test; AVF = Action-Verb Fluency; SS = WMS-IV Symbol Span Total Raw Score.

* $p < .05$

** $p < .01$

Table 6

RFFT Correlations with Executive Functioning Measures

Index	RND	RPR	RER	RTC	REC	RRC	RPS	RTS
RAD	0.22*	0.13	0.11	-0.03	-0.11	0.08	-0.12	0.26*
RCS	0.08	0.20	0.19	-0.05	-0.02	-0.05	-0.12	0.15
SI	-0.10	-0.01	0.03	-0.03	-0.18	0.16	0.06	-0.10
SIS	-0.14	-0.15	-0.12	-0.02	-0.17	0.15	0.01	-0.18
CTC	0.30**	0.16	0.11	0.33**	0.39**	0.05	0.27*	0.14
CNC	0.20	0.01	-0.05	0.16	0.16	0.05	0.12	0.10
TMA	-0.19	-0.11	-0.05	-0.10	-0.08	-0.05	0.01	-0.19
TMB	-0.17	-0.11	-0.09	-0.14	-0.03	-0.16	-0.07	-0.13
TTA	-0.09	-0.14	-0.11	-0.10	-0.14	0.01	-0.09	-0.08
TTP	-0.19	-0.12	-0.13	0.13	0.12	0.05	0.14	-0.21*
SF	0.22*	0.16	0.14	0.04	0.03	0.01	-0.01	0.24*

Note. ND = RFFT Novel Designs; RPR = RFFT Perseverative Responses; RER = RFFT Error Ratio; RTC = RFFT Total Strategic Clusters; REC = RFFT Total Enumerative Clusters; RRC = RFFT Total Rotational Clusters; RPS = RFFT Percentage of Designs in Strategies; RTS = RFFT Total Switches; RAD = Ruff 2 & 7 Auto Detection Accuracy Score; RCS = Ruff 2 & 7 Controlled Search Accuracy Score; SI = Stroop Inhibition Time to Complete; SIS = Stroop Inhibition Switching Time to Complete; CTC = COWAT Total Correct; CNC = COWAT Number of Clusters; TMA = Trail Making Test, Part A; TMB = Trail Making Test, Part B; TTA = Tower Test Total Achievement Score; TTP: Tower Test Mean First Move Time; SF = Semantic Fluency Total Correct Words.

* $p < .05$

** $p < .01$