8-4-2015

Construct Validity of the D-KEFS Sorting Test: Standard and Non-Standard Administration with a Non-Clinical Undergraduate Sample

Sunshine M. Collins
Indiana University of Pennsylvania

CONSTRUCT VALIDITY OF THE D-KEFS SORTING TEST: STANDARD AND NON-STANDARD ADMINISTRATION WITH A NON-CLINICAL UNDERGRADUATE SAMPLE

A Dissertation

Submitted to the School of Graduate Studies and Research

in Partial Fulfillment of the

Requirements for the Degree

Doctor of Psychology

Sunshine M. Collins

Indiana University of Pennsylvania

August 2015

Indiana University of Pennsylvania
School of Graduate Studies and Research
Department of Psychology

We hereby approve the dissertation of

Sunshine M. Collins

Candidate for the degree of Doctor of Psychology

____________________ ______________________________________
David J. LaPorte, Ph.D.
Professor of Psychology, Chair

____________________ ______________________________________
William Meil, Ph.D.
Professor of Psychology

____________________ ______________________________________
William J. Farrell, Ph.D.
Assistant Professor of Psychology

ACCEPTED

______________________________________ ____________________
Randy L. Martin, Ph.D.
Dean
School of Graduate Studies and Research

Title: Construct Validity of the D-KEFS Sorting Test: Standard and Non-Standard Administration With A Non-Clinical Undergraduate Sample

Author: Sunshine M. Collins

Dissertation Chair: Dr. David J. LaPorte

Dissertation Committee Members: Dr. William Meil
Dr. William J. Farrell

This study had three major objectives: (a) to investigate the construct validity of the Delis-Kaplan Executive Function System Sorting Test, (b) to investigate possible improvements to the construct validity by modifying administration instructions based on theoretical grounds, and (c) to explore utility of Sorting Test optional scores.

Students in introductory psychology classes at a rural state university enrolled in the subject pool were sampled for this study (n=171). Participants ranged from 18-23 years old, were primarily freshman, and predominantly identified as White, non-Hispanic. Using this sample, the construct validity of the Sorting Test was not fully supported. Results were inconclusive regarding convergent validity; neither Confirmed Correct Sorts nor Free Sorting Description Score on the Sorting Test were significantly correlated with Categories Completed or Perseverative Responses on the Wisconsin Card Sorting Test (WCST), although Sort Recognition Description Score was significantly correlated with Categories Completed, r=.51, and Perseverative Responses, r=-.49 (all ps<.005). Nonsignificant correlations between the Sorting Test and Color-Word Interference suggested adequate divergent validity; however neither the WCST nor Letter Fluency yielded similar nonsignificant correlations with Color-Word Interference.

Correlation of the Sorting Test with Letter Fluency and WCST was significantly decreased under two experimental manipulations of task instructions. A Minimal Instruction condition generated a larger decrease than did a Feedback condition.

Significant correlation of the Sorting Test, Letter Fluency, or WCST with Color-Word was not present in either experimental condition. This represented a change from the Control condition in which Color-Word was significantly correlated with Letter Fluency, $r=.36$, Categories Completed, $r=.37$, and Perseverative Responses, $r=-.38$ (all $ps < .008$). No significant correlations were found between Set-Loss Sorts on the Sorting Test and Failure to Maintain Set on the WCST under either experimental condition.

Results suggested that the Sorting Test was not sufficient to replace the WCST as a measure of executive function associated with the dorsolateral prefrontal cortex. Theoretically derived alternative administrations significantly diminished the correlation between the two tests, suggesting that current administration instructions are adequate.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

CHAPTER 1

LITERATURE REVIEW

Clinical neuropsychology is the applied science that allows for the description of brain-behavior relationships (Lezak, Howieson, Loring, Hannay, & Fischer, 2004). These relationships are present in a variety of domains. Assessment of these brain-behavior relationships is most familiar to the general public in relation to the domain of cognitive abilities. This is because investigation into cognitive functioning began before neuropsychology, with educational psychologists who were attempting to measure intelligence (Binet & Simon, 1948). Early tests of intelligence led to the development of the Army Alpha and Army Beta screening measures used during World War I and this practical application of assessment demonstrated the potential for companies founded to create such tests (Gottfredson & Saklofske, 2009). It was not until the late 20th century that psychologists began studying individuals with brain injuries to understand cognitive functions more completely (Jones-Gotman, 1990).

Neuropsychological testing is the comprehensive evaluation of an individual's functioning across many domains for the purposes of diagnosis, treatment, prognosis, and lesion localization, among others. Domains that are commonly overlapping as well as evaluated separately, in a neuropsychological assessment, include intelligence, attention, language, personality, orientation, perception, memory, construction, concept formation, motor performance, executive function, and judgment. By utilizing a multi-method approach to information gathering, clinical neuropsychology has developed in the present age to be capable of assessing individuals with a diverse set of behavioral presentations.

The ability of neuropsychologists to provide information appropriate for diagnosis, treatment, prognosis, planning of care, evaluation of treatment outcomes, research, and addressing neuro-forensic questions is predicated upon the enhancements and availability of multiple information sources that support complex interpretations of human behavior. Strides in neuroimaging, the neural underpinnings of behavior, and multifaceted assessment approaches have enhanced such abilities for each subsequent generation of neuropsychologists. Neuropsychological assessment has enhanced our knowledge of disorders in multiple spheres of practice, including neurological, medical, and psychiatric (Brooks, Strauss, Sherman, Iverson, & Slick, 2009). This multifaceted approach extends assessment beyond the initial cognitive domain into other domains, such as language, psychomotor, visuospatial, and executive functions. Of all of these, it is argued that executive functions were key to the development and evolution of modern thinking (Coolidge & Wynn, 2001).

Executive Functions

Although executive functions have been an important focus of neuropsychology in recent years, there is no consensus as to what the term executive functioning means (Salthouse, 2005; Wecker, Kramer, Hallam, & Delis, 2005) and there is evidence to support recognizing executive functions as both unitary and diverse (Miyake et al., 2000). Which specific cognitive abilities to include under the nomenclature *executive function* is a matter of great debate in the literature of neuropsychological testing. This lack of consensus can be seen in the variety of variables evaluated on different measures of executive functions, such as the Delis-Kaplan Executive Function System (D-KEFS) or the Neuropsychological

Assessment Battery by Stern and White, which have very few individual tests in common (Salthouse, 2005). Similar to the intelligence tests from which they developed, tests of executive function developed alongside developing theories of the construct.

Frontal Lobes

Beginning with the theory of a higher order central executive capacity that was responsible for coordination of lower order processes (Baddeley, 1992), executive functions were initially understood to be a construct of cognitive control and working memory (Harding, Harrison, Breakspear, Pantelis, & Yücel, 2014). These trends in the historical course of identifying and defining the executive functions highlighted the frontal lobes as the site of executive functions (Stuss & Alexander, 2000a). The frontal lobes are an area of the brain that serve to interconnect major sensory and motor systems, thus serving as an integration site for all of the feedback loops and information that result in behavior. The effect of this integration on behavior is pervasive because it includes pathways from both the posterior cortex and the limbic system, such that the anterior section of the frontal lobes is processing information from both external and internal stimuli to inform behavior (Middleton & Strick, 2000).

Since the frontal lobes were highlighted as the gateway to assessing executive functions, the conceptualization of this brain region has undergone extensive revision. Conceptualized initially as a single entity, the frontal lobes are now understood as a variety of different systems that impact functional organization (Miller, 2007). In the beginning of the 20th century, a more complex breakdown of the frontal lobes was identified. In his classic 1909 work, Brodmann developed a cytological architectonic approach to understanding the frontal lobes (2006). Neuroimaging has become a key method for studying the localization of executive functions in the frontal lobes by enabling clear description of frontal lobe neuroanatomy activated through executive function task demands.

Use of X-ray, computerized tomography, and magnetic resonance imaging enabled precise neuroanatomical description that can be enhanced by neuropsychological testing that explicates the behavioral, emotional, and functional impact of patient specific structural injuries to a high degree of specificity. Modern neuroanatomical study of the frontal lobes produced an extensive literature of cytoarchitechtonic mapping, combining architectonic maps and functional imaging data to analyze structural-functional correlations of the cerebral cortex (Amunts & Zilles, 2001). For this approach, neuroimaging analysis is made of classic Brodmann maps of these areas, a process more intricate than gross anatomical structure presentation. Whether by functional region, cortico-cortical networks, or organization of anatomy by vascular territories, all imaging studies are hampered by incomplete anatomic localization of activation areas (Damasio, 1991; Goldman-Rakic & Friedman, 1991). Functional localization of executive functions in the frontal lobes provides a limited base from which to interpret brain-behavior relationships because it cannot account for the interindividual variability identified through neuroimaging and cortical mapping (Amunts, Schleicher, & Zilles, 2007).

The functional localization approach has produced a problem in the literature and our current understanding of executive functions. Functional localization uses anatomical definitions of executive function. Psychological definitions of executive function focus on deficit and ability in functional capacities. The combination of these approaches has led to inconsistent definitions of executive function in the literature. The terms executive, supervisory, and frontal are used both interchangeably and used to reflect slightly different constructs. These changes in labeling and defining the construct have led to ambiguity in the literature and the field. (Stuss & Alexander, 2000). Researchers and clinicians now vary in their definition of executive function by anatomical structures included, functional capacities included, and even relevance of executive function to overall cognitive functions.

These changes in the definition of executive functions have impacted test development. Currently, a variety of tests of executive functions are available owing to an absence of a generally accepted theory and structure of the latent construct being measured (Gottfredson & Saklofske, 2009). In a 2005 article, Salthouse investigated the relationship pattern between executive functions and established cognitive abilities, finding little evidence for the unique contributions of the executive function measures beyond that contributed by the cognitive measures with the cognitive impact of age held constant (2005). Findings such as these illuminate the failure to clearly define and operationalize executive functions.

Defining the Construct

The construct of executive functioning is comprised of component performances that are clustered differently based on theory. Lezak et al. propose that executive functions are comprised of four primary elements: volition, planning, purposive action, and effective performance with an additional area of self-regulation often included (2004). They define volition as the complex behavior that allows for the formation of intention and the conceptualization and initiation to achieve this intention. Motivation and drive are conceptualized as necessary but insufficient preconditions to the successful implementation of volition. Planning is conceptualized as the ability to identify and organize multi-step plans to successfully achieve a goal or intention. Research has demonstrated that planning is also comprised of multiple preconditions, such as prospective reasoning and flexible adaptation (Pavawalla, Schmitter-Edgecombe, & Smith, 2012). The ability to take these plans and intentions and implement them is defined as purposive action. Purposive action, then, is the self-serving activity achieved through the proper application of initiation and the ability to appropriately maintain, switch, and stop behavior sequences as appropriate (Lezak et al., 2004). Self-regulation influences productivity, flexibility, perseveration, and a myriad other areas that require controlled performance (Lezak et al., 2004). Effective performance is related in that it is impacted by self-monitoring, correction, and regulation. Each of these four primary elements are comprised of activity-related behaviors and are necessary for successful navigation of our social world.

Fuster provides a different definition of executive functions (1991). He suggests that the prefrontal cortex temporally organizes behavior by mediating cross-temporal contingencies (Fuster, 1991). From this perspective the prefrontal cortex is responsible for integrating behavior with sensory information that is not present currently, but either was in the past or will be in the future (Fuster, 1991). Fuster, therefore, places the prefrontal cortex at the pinnacle of integration within the perception-action cycle (1991). Deliberate sequential behavior results in a circular pattern of cybernetic influences such that a cycle is formed that portrays the flow between sensation and movement (Fuster, 1991). This cycle is relevant to Fuster's hypothesized role of the prefrontal cortex because the addition of time to the cycle, in the form of distance between movement and sensation, calls for additional cognitive functions to close the cycle These basic functions, short-term memory and preparatory motor set, are represented in the dorsolateral prefrontal cortex and are temporally symmetric in that one is retrospective and one prospective (Fuster, 1991). The principle of cross-temporal contingency, therefore, is universally applicable to all sequences of behavior which have substantial temporal distance between mutually contingent events (Fuster, 1991). This can be summarized as "If now this, then later that; if earlier that, then now this" (Fuster, 1991, p. 60). The prefrontal cortex as the temporal organizer of behavior presumes that executive areas in the lateral cortex of the frontal lobes are hierarchical, such that higher areas are associated with higher levels of complexity in the information being integrated to produce behavior (Koechlin, Ody, & Kounelher, 2003). Fuster's overarching position is that the frontal lobes are responsible for the development of temporal gestalts (Shallice & Burgess, 1991). As is the case with many of the definitions of executive function, Fuster's definition developed as more information became available regarding the structure of the frontal lobes, in this case the neural dynamics of the frontal hierarchy in behavioral action (Fuster, 2000, 2004). Fuster argues that the confusion about the prefrontal cortex today stems from the basic mistake of focusing on one prefrontal cortex function to the neglect of others and localizing any of these functions to a precise prefrontal cortex location (2001). This localizing approach remains active in research, however, with researchers employing neuroimaging techniques to image the activity of many inputs, enabling precise localization of neuronal computations of specific functions (Priebe & Ferster, 2010).

Drawing from Luria's conceptualization of the frontal lobes as a programming, regulating, and verifying system resulting in behavior, Norman and Shallice developed a third definition of executive functions. In Norman and Shallice's model, there is a lower and higher level such that the lower level is responsible for relatively routine activity and the upper level is not engaged unless a novel activity occurs which the lower level is insufficient to manage (Shallice & Burgess, 1991). In this definition, behavior is selected in a condition-action problem solving format which triggers an appropriate program similar to a schema (Shallice & Burgess, 1991). These schema/programs require resources and so their implementation must be monitored to prevent two schema that require the same resources from being activated (Shallice & Burgess, 1991). In Norman and Shallice's model, this monitoring is controlled by the Supervisory System and it is damage to this Supervisory System that results in the classic disorders observed following prefrontal lesions (Shallice & Burgess, 1991).

A developmental model for evaluating executive function is also appropriate. Research on the neural correlates of cognitive control in children indicates that increasing age is associated with the development of executive functions (Lamm, Zelazo, & Lewis, 2006). Executive functions tend to develop later than other neurocognitive functions, such as language and memory, and tend to decline in old age more rapidly as well. (Zelazo, Craik, & Booth, 2004). One framework for the development of executive function across the lifespan identifies bimodal gains in working memory, planning, and problem solving, first at 15-19 years and again at 20-29 years (Luca et al., 2003).

Unitary Versus Diverse Construct

Ultimately, the question of how to measure executive functions is impacted by how these functions are conceptualized, either as a unitary or diverse construct (McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010). The terms executive functioning and frontal lobe functioning are often used interchangeably, although executive functions are currently being perceived as the more fluid abilities associated with the frontal lobes (McCabe et al., 2010). When executive functions are conceived of as a set of various constructs that are related loosely by their connection to the frontal lobes, intuitive approach to measurement becomes the application of multiple tests to assess each individual component construct that is gathered under the umbrella term executive functions (McCabe et al., 2010). Those who take a more unitary conceptualization are also supported in the literature, with a recent study evidencing that many executive function tasks appear to share an underlying reliance on attentional control during goal oriented activity (McCabe et al., 2010). Recently, a new theoretical framework has been developed that integrates these disparate findings concluding that executive functions are both unitary and diverse, impacted by genetics, related to clinical and social phenomenon, and show developmental stability (Miyake & Friedman, 2012). The tools used to measure executive functions are tasked with describing a constellation of abilities generated by the frontal lobes, such that each measure could conceivably be measuring a cluster of abilities with little overlap to that measured by a different but similar tool. This possibility highlights that standardized measures of executive function would benefit from theoretically derived investigation of their construct validity. Tools with vague or unresearched executive function ability constellations could provide imprecise or incomplete data during neuropsychological testing.

Executive Dysfunction

Based on the above theories of executive function, dysfunction is understood to cause behavior change that can be used to directly and indirectly identify affected brain structures of individuals with executive dysfunction. Assessment tools for neuropsychological testing attempt to measure observable performance in these areas of behavior such that clarity of unobservable neurological functioning is achieved. Earlier, the theoretical definition of the construct of executive functioning was reviewed using a top-down approach. To use a bottom-up approach, executive dysfunction can be used to define the same construct. The wide range of behavioral correlates to frontal lobe damage can be organized, to some degree, by portioning the anterior section of the frontal lobes into three major subdivisions, each with their own connections to different thalamic nuclei, cortical and subcortical structures, and neural pathways (Owen, 1997). Damage in one of these three key areas can be observed through specific deficits in appropriate behavior. It is this brain-behavior relationship that is explicated through neuropsychological assessment. An accurate definition of the executive functioning construct will enable such testing to specifically identify the brain structure or pathway that is dysfunctional.

The first subdivision of the frontal lobes, the dorsolateral prefrontal cortex, is responsible for processing information used to engage in a range of cognitive functions.

Damage in this area results in deficits in the areas of control, regulation, integration of cognitive activities, judgment, and insight (Knight, 1991). These deficits can impact other processes, such as attention and memory, such that the observed process dysfunctions obscure the impaired executive functioning that catalyzed the deficit. In the case of attention, deficits in the ability to shift and maintain set and engage with novel information are responsible for diminished attentional capacities. Similarly, impairment in executive functions resulting in inertia and lack of spontaneous engagement can underlie observed memory problems associated with frontal lobe damage (Daffner et al., 2000).

The second subdivision, the medial regions (also called the cingulate or the limbic cortex) is responsible for drive and motivation because it connects the cortex with affective integration centers in the diencephalon (Lezak et al., 2004). Damage here results in limited capacity for emotion (Knight, 1991). This is a vital area for independent functioning, as individuals with low drive states often have reduced interest in initiating or maintaining social or vocational pursuits, despite intact life sustaining drives (Dimitrov, Grafman, & Hollnagel, 1996). Recall that the first definition of executive functions highlighted volition as a key component of intact executive functioning.

The last subdivision is the orbital frontal cortex, sometimes called the basal or ventral frontal cortex, and is responsible for impulse control and regulating and maintaining set and ongoing behavior (Malloy, Bihrle, Duffy, & Cimino, 1993). Damage to this area can result in disinhibition, impulsivity, aggressive outbursts, and sexual promiscuity (Eslinger, 1999). Orbitofrontal damage is characterized, then, by affective disturbances (Knight, 1991). In addition to the possible disinhibition resulting from damage to this area, this is also the site of understanding the future consequences of behavior; so poor decisions result from either or both of these when damage is received here (Bechara, 2004; Clark et al., 2008). This has been most notably described in the case of Phineas Gage, a railroad worker who evidenced such significant changes in personality following on-the-job damage to his orbital prefrontal cortex that his friends stated the he was a different person (Stuss, Gow, & Hetherington, 1992).

Importance of Executive Function

Although the emphasis on cognitive functioning as vital to a healthy lifestyle is not without merit, it is executive functions that provide the higher-level cognitive skills necessary to maintain an independent and meaningful life (Eslinger, 1996). Executive functions are necessary for many cognitive, emotional, and social skills as well as the ability to respond appropriately to the environment when novel situations occur. Thus, individuals with deficits in executive function are unable to successfully navigate the myriad experiences necessary for independent functioning in the modern world. If executive functions remain intact, an individual with significant cognitive loss can still operate independently in the world (Archibald, Mateer, & Kerns, 2001). Whereas cognitive deficits are typically seen in a discrete area, impairments in executive functions more often have global presentations with a variety of behavioral manifestations (Teuber, 2009). It is this global presentation that makes executive function impairments so incapacitating as far as self-care, employment, and social relationships (Gioia, Isquith, Kenworthy, & Barton, 2002). Impairment in executive functions often renders a person incapable of independent living, even in instances where retention of cognitive function is high, because impairments in executive functioning diminish a person's ability to plan, engage in multi-step activities, think abstractly and flexibly, and even to monitor and motivate their own performance (Goldstein & Naglieri, 2013). In a study of 95 community dwelling healthy older adults, executive functions were found to be directly related to performance based instrumental activities of daily living (Vaughan & Giovanello, 2010). Of particular import in terms of independent functioning, individuals with frontal lobe damage often exhibit anosognosia, or an inability to recognize their own deficits (Stuss, Picton, & Alexander, 2001). This is particularly important in terms of independent functioning because it renders the individual unlikely to engage in adaptive behaviors to moderate their difficulties, making independent functioning unlikely and treatment planning challenging. Even subtle impairments in executive functions have been shown to greatly hinder treatment effectiveness (Swick, Honzel, Larsen, Ashley, & Justus, 2012). Neuropsychological rehabilitation has often been overlooked for this population because of the poor prognosis following significant damage to the frontal lobes (Prigatano, 1991).

Failure in prospective memory, or remembering to remember, is another feature of executive function impairment that results in limited independence because individuals with damage to the frontal lobes may forget to go to work or other appointments, engage in self-care rituals such as bathing, and even to change clothes (Sohlberg & Mateer, 2001). Ettenhofer, Hambrick, and Abeles (2006) note that there are an increasing number of researchers examining the way that executive functions impact a wide range of neuropsychiatric conditions from schizophrenia to the normal aging process.

Neuropsychological assessment is also interested in identifying and extolling preserved functions, as they are informative for treatment planning and also provide information about strengths that can be utilized to develop adaptive behavior to accommodate other weaknesses (Vaughan & Giovanello, 2010). As Vaughan and Giovanello (2010) note, following their findings that a significant relationship exists between executive functions and performance based instrumental activities of daily living, experimental measures of executive functions can be used to predict daily function and provide appropriate treatment strategies by highlighting areas where cognitive training can be used to improve an individual's successful completion of activities of daily living.

Measurement of Executive Function

The service of integration performed by the anterior frontal cortex has not always been well understood and modern science continues conducting research that refines and redefines these functions (Stuss, 2011). In the early half of the 20th century, individuals with frontal lobe lesions and other types of damage were assessed using standard intelligence measures, resulting in test scores that indicated intact intellectual functioning (Lezak et al., 2004). These data were confusing because the frontal lobes were believed to be the seat of highest cognitive functions because of their recent evolution and increase in size (Lezak et al., 2004). Ultimately, it was discovered that, although general cognitive functions such as skill and other static information remained intact after frontal lobe damage, higher-level processes were disrupted (Tranel, Kemmerer, Adolphs, Damasio, & Damasio, 2003). These higher-level processes include the interaction between major systems, affecting the way that the sensory systems and the limbic system impact behavior

(Koechlin, Basso, Pietrini, Panzer, & Grafman, 1999). This is crucial as elements of arousal, affective states, and motivational states are all impacted by a disruption in the integration of these systems (Malloy et al., 1993). Individuals with executive function impairments are seen to fail to approach problems correctly, lack effective planning and integration of environmental feedback, and be stymied by a strong feature of a stimulus such that they lose the set of the task being worked on in the moment (Eslinger, Grattan, & Geder, 1995). As such, measurement tools should contain a stimulus feature that creates opportunities to assess patients on these features of approach style.

Further complicating this process of defining the construct of executive functions, conflicting research has not been able to conclusively indicate the stability of the construct. Measures of executive functions have been found to vary in their heritability, with measures such as the Wisconsin Card Sorting Test 64-Card Version (WCST) evidencing no heritability (Vasilopoulos et al., 2012). Executive functions have been shown to have a very high stability and reliability in older adults as a latent construct when measured at two time points and with 5 common measures, including the WCST and several items on which individual subtests of the D-KEFS are based, such as the Stroop Color-Word Test and the Trail Making Test (Ettenhofer et al., 2006). Conflicting research supports a frontal aging hypothesis, however, that suggests that changes in the frontal lobes following aging result in declines in complex cognition. This hypothesis further supposes that these structural and functional changes in the frontal lobes due to age take place at a faster rate than they do in other brain areas, with the dorsolateral prefrontal cortex evidencing change more significantly than even the orbital frontal areas (McCabe et al., 2010). It is also difficult to parse out what is truly unique to executive functions as they have also been shown to correlate closely with the cognitive abilities of reasoning and perceptual speed (Salthouse, 2005). As such, measurement tools need to engage executive functioning discriminate of cognitive reasoning and perceptual speed to the extent possible.

Observation

As previously noted, damage to the prefrontal cortex often does not affect cognitive function directly through loss of skill or previously held information, and so individuals with frontal lobe damage can perform reasonably well on standard ability tests. The testing environment that involves an examiner controlling organization of the time, activities, and providing directives to the examinee accommodate the diminished executive functions that are prevalent after frontal lobe damage (Frederiksen, 1986). This does not mean, however, that it is impossible to observe or assess for executive dysfunction in a standard testing format. Use of informant information from family, coworkers, and friends is valuable in obtaining information about possible executive dysfunction (Lezak et al., 2004). These reports are likely to include concerns about apathy, carelessness, poor judgment, diminished adaptability to novel situations, as well as limited social sensibility (Lezak et al., 2004). Informant report may also be helpful in identifying limited practical and social judgment. Elements of carelessness can be assessed by behavioral observation in session. Additionally, careful observation for possible perseverations can provide valuable information about executive functions as stimulus boundedness is a common deficit after frontal lobe damage (Tate, 1999). Executive functions are also at work in the assessment setting during timed tasks such that an individual with executive dysfunction is unlikely to successfully engage in anticipatory sequencing or efficiency in responding through planning and organization (Barkley, 1997). This will be observable during timed tasks in which such individuals engage in unusual or inefficient means to produce the requested response. Prefrontal cortex damage may also be observable in an assessment setting through absence of self-monitoring to enable self-correction (Eslinger et al., 1995). These findings may be heightened by a failure to adequately interpret time and make judgments about recency, size, and number (Levin, Benton, & Grossman, 1982). Such poor orientation to time compounds slowed response to complex tasks, making timed tasks more difficult for those with prefrontal cortex damage (Levin et al., 1982). Assessment measures must, therefore, be intentional in their treatment of time with respect to tasks tapping executive functioning.

Behavioral Disturbance

Lezak et al. advises that one of the best ways to use observation in an assessment setting to identify executive dysfunction is to be attuned to the presence of five general areas of behavioral disturbance (2004). These areas of behavioral disturbance are manifested to varying degrees, are often overlapping, and are the natural correlates of the deficits that have previously been expounded upon. The first area is that of starting difficulties. This is when an individual exhibits a decrease in spontaneity and productivity often from the decrease in drive and motivation discussed earlier. Individuals who have problems starting frequently are able to verbalize plans but do not initiate any of the described behavioral responses to a situation themselves (Eslinger et al., 1995). This is obviously seen to a varying degree of severity but in its most severe form an individual will lack all initiative beyond that which is dictated by bodily functions (Sohlberg & Mateer, 2001). Therefore, measures of executive function that include prompts after 30 seconds without a patient response may be less effective in testing the limits for individuals with some deficits in executive functioning.

The next area of behavioral disturbance is that of making and maintaining mental or behavioral shifts (Walsh, 1978). During assessment this may be observed globally, as perseveration and rigidity due to frontal lobe deficits often appear across tasks as a result of difficulties with attentional and novelty issues (Bechara, 2004). Therefore, effective measures of perseveration and rigidity may incorporate novelty as orchestrated by the administrator.

Problems in stopping are another area of behavioral disruption that can be anticipated in assessing individuals with frontal lobe damage (Bechara, 2004). This includes the impulsivity and disinhibition explored previously, as well as overreactivity and difficulty in delaying gratification (Stuss, 1991). Observation of this behavioral disruption would best be generated by measures that provide an opportunity for disserving behavioral repetition. In these measures, the repetition can more clearly be tied to disinhibition, as the impact of the behavior is not preferable to the patient.

Individuals with frontal lobe deficits often demonstrate anosognosia, a disruption of self-awareness resulting in limited capacity to recognize functional limitations—both on a global scale and for acute errors being enacted in the present moment. (Lezak et al., 2004). This has particular relevance to social settings, as it can result in a lack of empathy for others due to a failure to understand how one is impacting those around them (Anderson, Bechara, Damasio, Tranel, & Damasio, 1999). Although this is a serious limitation experienced by those with frontal lobe deficits, a lack of self-criticism often results in individuals feeling euphoria and self-satisfaction as well as decreased anxiety compared to a non frontal lobe damaged population (Stuss & Alexander, 2000b). In an assessment setting, this behavior change may be most observable in an absent concern for social conventions such that the examinee does not engage with the examiner in expected ways and may fail to respond normally to external stimuli (Rolls, Hornak, Wade, & McGrath, 1994). A measurement tool with extensive introductory material that the administer must review for the patient may be less successful at generating the opportunity to observe this absence of social convention by artificially extending the time period during which response to external stimuli is anticipated. Such an extension could have the effect of obscuring observable behavior on this dimension.

Lastly, a concrete attitude is another behavioral change that may be observable during testing of individuals with frontal lobe damage (Goldstein & Naglieri, 2013). This can be assessed by evaluating for abstract thinking and reasoning throughout the course of the testing experience. An individual with such damage would be unlikely to engage in foresight, goal directed behavior, and planning—resulting in a literal response set (Lezak et al., 2004). This is not to say that abstract conceptual abilities are not possible for those with frontal lobe damage, but rather, that they are unlikely to process information in this way spontaneously and are often found to experience a loss of perspective (Lezak et al., 2004). A tool that imposes abstract sets onto the testing situation rather than allowing for the patient to generate such sets organically may best measure this concrete thinking.

Deficit Measurement

Currently, neuropsychological assessment is predicated upon deficit measurement such that identification of areas of deficiency is the primary focus of assessments (Lezak et al., 2004). This is achieved by comparing data obtained to normative data organized by age, education, or another meaningful category. Neuropsychological assessment is made possible through the long-standing areas of educational and intelligence testing which provide well-defined operationalization and standardized tests to establish deficit measurement. Additional evaluation of pervasive and subtle elements such as executive functions can be obtained by evaluating their effect on various cognitive activities as well as through the use of specifically designed assessment measures. Tests designed to assess latent abilities, like intelligence or executive functions, are different from those designed to test achievement in that achievement tests need only provide items that probe the ability across a continuum of a content domain whereas tests of latent abilities must instigate the deployment of the latent ability of interest (Gottfredson & Saklofske, 2009). The intention of tests of executive functions are to instigate the latent ability of interest in the controlled environment of the testing room, such that the unobservable can be revealed through its effect on behavior (Gottfredson & Saklofske, 2009). Thus, when behavior instigated by the test is consistent with the construct that the test purports to activate, construct validity is demonstrated (Gottfredson & Saklofske, 2009).

Utilizing deficit measurement to assess functioning presupposes an existent norm within the larger population or a prior score of the individual's with which to make a comparison. This method of evaluation for a deficit in comparison to another score is called a *comparison standard* (Lezak et al., 2004). Neuropsychological assessments are versatile in that they can be conducted utilizing a normative or an individual score. This means that when the subject of interest has previously existing test data, these data can be used to evaluate for deficits and when such data do not exist or are not available, normative data from a relevant sample can be utilized. The tests used to conduct deficit measurement may look at species-wide performance measures, customary standards, or population average as the comparison point when utilizing a normative standard for comparison.

Although normative and individual assessments are possible, a deficit measurement utilizing baseline data of the individual is often considered superior. One reason for this is that deficits of adult onset cannot be assessed meaningfully using normative data unless the skill being assessed is one that should be within the purview of all adults, thus rendering any difference from the normative data meaningful (Lezak et al., 2004). If this is not the case and the ability is normally distributed throughout the population, it is difficult to ascertain what the discrepancy between current and prior performance may be using only normative data.

An initial estimate of prior functioning begins a deficit measurement assessment. This will be relatively simple for species-wide norms and will require more specific information for normatively distributed skills. Given the necessity of individual comparison for those skills that are normatively distributed in the adult population, much of a neuropsychological assessment is conducted using intraindividual comparisons.

Direct measurement of a deficit is preferable but baseline data for the individual are often not available, necessitating indirect measurement (Rabbitt, 2004). In this indirect method, comparison is made of current performance with a best estimate of premorbid functioning (Rabbitt, 2004). This estimate can be obtained from a variety of sources and often will incorporate more than one source, such as self-report, informant report, and performance on discrete tasks that have been shown to be stable after most injuries. One such example is the use of specific tasks such as word reading tests, like the Reading test of the Wide Range Ability Test, which are acknowledged as effective in estimating premorbid mental ability (Chan, Shum, Toulopoulou, & Chen, 2008). Another method for establishing an estimate of premorbid function is the best performance method, in which the highest score in an assessment is taken to be indicative of premorbid global functioning and comparison of the other scores is made to this highest one. Once a methodology for establishing the comparison standard is achieved, the discrepancy between the expected and actual performance can be measured and a cognitive deficit can be identified where statistically significant differences are observed. This presupposes that the measures used to assess actual performance can reliably be administered so as to legitimate opportunities for standardized comparison. Measures that lack construct precision would be poor instruments with which to establish a comparison standard.

Standardized Assessment Measures

Examination of executive functions must evaluate the capacities that comprise the major aspects of these functions. Executive functions can be impaired by deficits that are manifest at different stages of behavior sequencing, so it is important to identify at what stage the observed executive dysfunction is occurring, by evaluating the component parts of executive functions (Grant & Adams, 2009).

In addition the behavioral domains useful for observation of executive dysfunction in testing situations, direct assessment tools can be used to evaluate executive functioning. Although limited in quantity, there are measures that allow for the transfer of some of these executive functions to the examinee during the testing session, without sacrificing the structure and standardization of the procedure (Frederiksen, 1986). These instruments attempt to assess executive functions as a diverse construct, such that it is not uncommon to state that the measure or technique is assessing volition specifically, rather than executive functions generally (Lezak et al., 2004). Much like intelligence tests, which can be single factor focused, brief, or focus on particular populations (Gottfredson & Saklofske, 2009), tests of executive functions are varied in scope and focus. In a study using multiple discriminant function analysis, Stuss and Trites (1977) demonstrated that neuropsychological test scores are able to offer significant discrimination between groups including a control group, and two brain-damaged groups, one with a positive physical neurological exam and one with a negative exam. This provides support for the use of neuropsychological tests for the assessment of brain pathology. Often, a neurologist or other professional will request neuropsychological testing to confirm a diagnosis so the neuropsychological testing is often an adjunct to a physical neurological exam (Stuss & Trites, 1977). These findings offer preliminary data that neuropsychological tests provide better classification accuracy then do physical neurological exams due to their statistically higher sensitivity to cerebral pathology allowing for accurate discrimination of individuals with brain damage that had previously received a negative physical neurological exam (Stuss & Trites, 1977).

Evaluating the components of executive functioning in a traditional assessment procedure with performance based measures does present challenges. Studies show that frontal lobe damage can result in inconsistency in test performance (Stuss, 1991) and those with the most variability may be those with the most significant damage (Stuss, Pogue, Buckle, & Bondar, 1994). As previously noted, the structure of a traditional assessment makes it difficult to assess for executive dysfunction because the examinee is not being asked to self-direct or self-regulate their behavior (Stuss, 1991). The testing situation is such that the examiner controls pace, organization, structure, and even what materials to use at what time (Stuss, 1991). Additionally, a deficit in basic skills will artificially suppress scores on measures of executive functions making it difficult to assess executive functions even when preserved (Wecker et al., 2005). Additionally, studies (Heflin et al., 2011) have found conflicting results. Using 112 mildly cognitively impaired patients, Heflin et al. (2011) found that performance on a popular test of disinhibition and frontal lobe damage (Stroop, 1935) did not measure behavior disinhibition effectively even in a high risk population.

Current research evaluating measures of executive function for the precise construct of ability constellations being evaluated is needed to address the inconsistent findings in the literature.

Sorting Tests

The Wisconsin Card Sorting Test is a neuropsychological test that measures executive function. This test requires subjects to match a deck of cards to one of four key cards. Subjects receive feedback on their performance but no additional information on how to perform successfully. The Delis-Kaplan Executive Function System Sorting Test requires subjects to sort a set of cards into meaningful groups and to define the groups. Subjects are informed that they should make as many different sorts as they can. They make these sorts and provide these group descriptions over two rounds with different cards provided for each round. After this, two additional rounds occur in which the test administrator makes the sorts and the subject provides group descriptions of the administrator's sorts. Sorting tests like the WCST require participants to make a decision based on information that is both temporally separate and changing. In this case, successful performance can only be achieved through engaging in behavior that is contingent upon information from the recent past (Stuss & Trites, 1977). Unlike other measures of decision-making impairment which utilize real world contingencies, such as reward and punishment in the Iowa Gambling Task (Waters-Wood, Xiao, Denburg, Hernandez, & Bechara, 2012), sorting tests are absent of uncertain reward and loss schedules. In the WCST, a reward system is in place (Diamond, 1991), but the reward is in the form of being informed that they have made the correct response, which lacks the real-world value of the monetary reward represented in the Iowa Gambling Task.

The WCST is considered the classic test of dorsolateral prefrontal cortex function (Diamond, 1991) and is most often used to assess mental flexibility and set shifting, although it is also used to measure inhibition of previous sets, goal directed behavior after feedback, abstraction, and concept formation (McCabe et al., 2010). It is classified as a test of executive function because it requires participants to utilize strategic planning, organized searching, and modulate impulsive responding to successfully engage in the task

(Strauss, Sherman, & Spreen, 2006). The range of uses for the WCST can best be understood as a natural result of the varied history of sorting tests.

History of Sorting Tests

Mental set was tested with sorting tasks as early as 1900 when Ach developed a sorting task in which participants identified sorting principles using objects and cards (Nyhus & Barceló, 2009). Ach is notable for describing the paradigm that became the model for all subsequent sorting tests (Eling, Derckx, & Maes, 2008). This paradigm was to institute a task so that the experimenter could observe the process of concept formation (Eling et al., 2008). In 1920, sorting tasks were used by Goldstein to assess attitudes of brain-damaged patients, specifically their concrete and abstract thinking (Nyhus & Barceló, 2009). This is perhaps best depicted in the study of a patient dubbed Th. (Eling et al., 2008). This patient was given sorting tasks to complete and it was through this that Goldstein developed the concept of observing concrete and abstract thinking through sorting tasks, identifying healthy individuals as those capable of abstract thinking while brain damaged individuals were concrete in their thinking, often becoming rigid in how they interacted with the environment (Eling et al., 2008).

Attempting to trace the history of sorting tests and their use becomes difficult as it is unclear which historical figures were influenced by one another, working with or for one another, and even which ones were aware of work in similar areas being conducted at the same time as their own research in different locations (Eling et al., 2008). It appears that Gelb worked closely with Goldstein in furthering the application of sorting tests to brain-injured individuals, however, it is suggested that their work borrowed heavily from that of

Ach, although they were interested in using sorting tests to explore concept formation and abstraction and had close ties with Gestalt psychology (Eling et al., 2008). In 1924, Gelb and Goldstein developed the Gelb-Goldstein Color Sorting Test based on the Holmgren Test (Eling et al., 2008). This measure did not yield a total score, rather, Goldstein proposed that it was in observing the process of the subject completing the task that the administrator came to grasp the attitude of the subject (Eling et al., 2008).

Subsequent to these, the WCST was devised in 1948 by Grant and Berg (Nyhus & Barceló, 2009). The WCST was initially intended to assess abstract reasoning, concept formation, and integration of feedback to instigate behavior change (Nyhus & Barceló, 2009). Milner then modified the WCST with correction criteria in 1963 to enable its use as an assessment of prefrontal lobe dysfunction in individuals with brain lesions (Nyhus & Barceló, 2009). Since that time, other versions of the WCST have become popular, including a shortened version by Heaton (Nyhus & Barceló, 2009).

Other sorting tests include the Kasanin-Hanfmann Concept Formation Test first developed in 1953 and the Color Form Sorting Test first developed by Goldstein, Scheerer, and Weigl in 1941 (Lezak, Howieson, Bigler, & Tranel, 2012). The former was introduced to the United States by Kasanin and Hanfmann but originally developed by Vygostky, a Russian psychologist (Eling et al., 2008). The later measure is often called Weigl's Test and is the first sorting task to utilize the principles of sorting and shift to increase the screening possibilities of the test (Weigl, 1941). Another measure designed by Goldstein, Scheerer, and Weigl in 1941 was the Object Sorting Test (Goldstein, Scheerer, & Hanfmann, 1953). This task utilized everyday objects in place of more traditional token and block sorting tasks. Sorting tests continue to be developed today with multiple variants of the WCST, including a 2004 tactile version developed by Beauvais et al. to assess executive function in the visually impaired through palpating and then sorting stimuli (Beauvais, Woods, Delaney, & Fein, 2004). The same principles of sorting tests can be seen in a subtest of the Halstead-Reitan battery called Category Test which has been much copied in other tests (Eling et al., 2008), including the stand-alone measure the Booklet Category Test.

The WCST has been utilized as a test of prefrontal function for the last four decades, which accounts for the ample research literature available (Nyhus & Barceló, 2009). Research has shown that the WCST activates the dorsolateral prefrontal cortex as observed through blood oxygenation changes being measured by multichannel near-infrared spectroscopy in healthy adults (Sumitani et al., 2006). MRI and regional cerebral blood flow (rCBF) have also been used to evidence the activation of the dorsolateral prefrontal cortex in individuals completing the WCST (Berman, Torrey, Daniel, & Weinberger, 1992; Weinberger, Berman, Suddath, & Torrey, 1992). Another study confirmed the dorsolateral prefrontal cortex activation via positron emission tomography (Kirkby, van Horn, Ostrem, & Weinberger, 1996). The WCST is further supported as the gold standard for tests of executive function through two meta-analyses, which show that the WCST is sensitive to frontal damage, with the highest effect sizes being for dorsolateral damage, and that the WCST does not differentiate between left versus right lateralization of damage (Demakis, 2003). Meta-analyses such as these are important because there are individual studies indicating that the WCST does not discriminate between frontal and nonfrontal lesions (Nyhus & Barceló, 2009).

Delis-Kaplan Executive Function System (D-KEFS)

Of the executive function measures, there are also those that espouse to complete a wide range assessment of executive functions, of which the D-KEFS is one. The D-KEFS was authored by Dean C. Delis, Edith Kaplan, and Joel H. Kramer in 2001 and is published by Pearson (Pearson Education, Inc., 2012). It was designed to be the first nationally standardized instrument to assess executive function in both adults and children (Pearson Education, Inc., 2012). It is comprised of nine individual tasks that deliver discrete scores, which can be used individually to assess specific components of higher level executive functioning. The instrument does not produce an overall or composite score of executive functioning but does provide nine different tasks to assess frontal lobe performance. Pearson markets this instrument as an engaging test that utilizes game-like activities and removes examiner feedback that may elicit frustration in examinees (Pearson Education, Inc., 2012).

The design of the D-KEFS allows for multiple uses in clinical practice. The most obvious use is to identify deficits on tasks that are mediated by the frontal lobe (Pearson Education, Inc., 2012). Additional uses include identification of how deficits in executive function may impact activities of daily living and the ability of the examinee to participate meaningfully in various life tasks as well as indicating areas of rehabilitation and the ability to tailor recommendations to specific areas of dysfunction (Pearson Education, Inc., 2012). These additional uses of the D-KEFS are important for clinical practice in that they augment clinician feedback beyond identifying a problem with executive function to providing concrete and meaningful responses to diminish the impact of such deficits where it exists.

Part of how the D-KEFS supports the identification of these responses is that its composition of stand-alone measures allows for the isolation of the specific higher-level cognitive process that is not functioning optimally.

This benefit is contrasted with the reality that the individual tasks that comprise the D-KEFS are newer iterations of tasks that have been previously authored, normed, and produced. Thus, prior to the introduction of the D-KEFS, clinicians could have achieved similar results by using the individual tasks on which much of the D-KEFS tasks are based. This was likely made possible because the traditional versions of the Trail Making Test, Verbal Fluency, Twenty Questions, and Proverbs are in the public domain. Further, the D-KEFS was designed without theoretical rationale for the inclusion of its subtests, beyond that of incorporating verbal and nonverbal tasks (Pearson Education, Inc., 2012). The advantage of using the D-KEFS over the individual measures that some of the subtests are based on is described by the authors as the co-norming of all of these subtests on 1,750 individuals between the ages of 8 and 89. An alternative form for Verbal Fluency, Sorting, and Twenty Questions has also been provided by the authors and is normed on a sample of 295 participants. The D-KEFS also produces a variety of scores such that raw scores can be translated into standard scores and cumulative percentile ranks. Additional differences between the original forms of these subtests and their D-KEFS representations are the lengthening of the measures and the addition of items of low and high difficulty to minimize ceiling and floor effects. Many of the subtests on the D-KEFS are also broken down to provide scores for individual components that produce successful completion of the complex task, a new approach to clinical neuropsychology introduced by Edith Kaplan called the *process approach* (Delis, 2010). This process approach is helpful given that a deficit in basic skill areas underlying the more complex task of interest will artificially lower performance on executive function measures (Wecker et al., 2005). On its surface, this sounds like a great additional feature to some of the well-known versions of these subtests, however, some authors (Lezak et al., 2004) have argued that there is no established support for the additional time and energy necessary from both the examinee and examiner to obtain these additional scores. Lezak et al. (2004) give the example of the Trail Making Test, which has 12 primary measures and 12 optional scores, and the Sorting Test, which has 5 primary measures and 29 optional ones. An example of this is provided for the addition of semantic fluency to the Verbal Fluency task. The addition of this to the original format of the test did not provide further success at discriminating between individuals with focal frontal lobe lesions and those without (Baldo, Shimamura, Delis, Kramer, & Kaplan, 2001). In fact, the D-KEFS as a whole contains a total of 125 scores, 42 of which are labeled as primary and 83 of which are optional. In a test of significance relying on a $p < .05$, 5 of every 100 scores will be statistically significant based on chance alone. Given that the D-KEFS provides 125 scores, several scores from each administration can be expected to be elevated due to chance alone.

Correlation With the "Gold Standard"

A more detailed description of the Sorting Test, including psychometric properties, is provided later, however, a brief description is appropriate here. The Sorting Test has existed in a similar format as the California Card Sorting Test since 1992 (Delis, Squire, Bihrle, & Massman, 1992) and is also remarkably similar to the WCST. This subtest is comprised of two conditions: Free Sorting and Sort Recognition. In the Free Sorting condition, examinees are asked to use cardboard cards of varying qualities (e.g. color, size) to make two groups of three cards in each group. The intention is that the examinee will use either verbal or perceptual qualities to sort the cards. In the Sort Recognition condition, the examiner sorts the cards into two groups and the examinee is asked to describe how the cards in each of the groups are similar, thereby identifying the sorting principle used to sort them into the groups by the examiner. In a study utilizing two data sets of 328 and 7,000 participants respectively, Salthouse (2005) found that the number of correct classifications on the Sort Recognition portion of the subtest was weakly and nonsignificantly correlated with reasoning, spatial, and memory abilities and that the removal of the spatial construct yielded a statistically significant influence of reasoning ability.

Confirmed Correct Sorts on the Sorting Test has been demonstrated to have a modest correlation of .64 with Categories Achieved on the WCST although Perseverative Responses between the two tests seem unrelated with r= .15 (Beatty & Monson, 1996). Use of the Sorting Test to assess frontal lobe lesions is supported by data that indicate that patients with frontal lobe lesions provide fewer attempted sorts, correct sorts, and correct sort descriptions compared to control subjects (Beatty & Monson, 1996). In contrast, Parkinson’s patients did not display discernable differences from controls beyond the area of providing more perseverative sorts (Dimitrov, Grafman, Soares, & Clark, 1999).

Both measures have been shown to be effective in assessing executive ability in individuals with multiple sclerosis (MS) as well as to correlate with MRI indices of brain atrophy (Parmenter et al., 2007). Again, both measures discriminated between employed and disabled individuals and the D-KEFS did so after controlling for depression (Parmenter et al., 2007). If these similarities in performance hold true, ease of administration becomes a key factor in determining which measure to administer. One consensus panel indicated that the D-KEFS is an attractive alternative to the WCST because there are alternative forms of the Sorting Test, something not available with the WCST (Parmenter et al., 2007). This is a considerable advantage as it has been noted that the solution to the WCST is such that it is easily remembered for subsequent testing experiences, making it useful as a measure of executive function only once (Lezak et al., 2004).

Although this is true, the administration of each measure varies significantly in administrator burden, with the WCST requiring the administrator to do little more than provide a brief and uninformative introduction to the task and then to confirm or deny correct sorts for 64 trials. During the administration, the administrator makes the barest of notations to allow for correct scoring at the completion of the testing session. This differs significantly from the burden of the Sorting Test. In the Sorting Test, the administrator provides substantially more instruction throughout, and needs to record responses verbatim to enable appropriate scoring at the end. The test consists of a trial in which the examinee provides sorts as well as descriptions of the sorts and a trial in which the administrator provides sorts and the examinee is responsible only for providing descriptions. In both conditions, the Sorting Test examiner is actively recording responses as well as being mindful of timing constraints. Further, this test involves three sets of stimuli, as opposed to the one set provided for administration of the 64-card manual WCST.

Additional tasks that fall to the examiner in the administration of the Sorting Test is shuffling the stimuli in between each free sort, even as the examinee is still manipulating the stimuli because this is a continuously timed task (apart from the time during which the examinee is providing descriptions of their verbal or perceptual sorts).

Reflecting on means by which tests of executive function can instigate behavioral correlates of executive dysfunction, brief task comparisons appear to demonstrate disparate utility between the two measures. Of the two measures, only the WCST incorporates a strong stimulus feature that creates in vivo opportunity for set loss, a component reviewed earlier for its identification of patient task approach style. The WCST removes time burden from the administrator, allowing for testing of limits with patients demonstrating avolition and also enabling discrimination of performance from processing speed confounds. The Sorting Test incorporates both time constraints as well as prompts to initiate behavior after 30 seconds of nonresponse. The Sorting Test includes extensive preamble such that social interaction dysfunction may be obscured and task novelty mitigated. In contrast, the WCST administration merely informs that the clinician is not allowed to tell very much about how to do the test. The WCST is structured so as to allow for behavioral repetition evident of stopping deficits, whereas the Sorting Test removes such opportunities by asking the patient to lead each round of sorting. These contrasts suggest that further evaluation may be necessary. Ascribing a theoretical conceptualization model from which the measures approach executive function evaluation may enhance clinical utility of these measures. To fully classify the construct validity of the Sorting Test as the modern iteration of the WCST more information is needed.

Higher-order cognitive abilities specific to complex or novel task demand have been established in the literature as key elements of modern sophisticated thought development. Conceptualized as both a unitary and diverse construct, this constellation of fluid abilities is mantled under the executive functions nomenclature. Executive functions are fundamental to independent and successful functioning in a variety of settings and contexts.

Measurement of executive function component ability clusters is conducted through a range of methods, including direct assessment with tests specifically designed for this purpose. The permutations of these ability clusters are extensive, and vary dependent on theoretical construct conceptualization that differs by theorist. The tools used to measure these vast permutations must be understood by clinicians with respect to their theoretical construct conceptualization. The measures used to evaluate executive functions are the product of extensive history and research, but may require refining to best serve as a tool that clinicians can utilize to capture the unique and variable behavioral dysfunctions produced by patients.

Given this, it is important to assess the possible benefits as well as the shortcomings of these newer iterations compared to the original tests upon which they are based. A comparison of this kind is essential to evaluating the ultimate utility of the D-KEFS and identifying what, if anything sets it apart from its predecessors. Lezak et al. (2004) have argued that although the principle scores from the D-KEFS have acceptable reliability, the optional scores often have variability amongst age groups and low reliability. They support these claims with reliability data reported in the D-KEFS manual, which on the switching condition of Design Fluency, for example, varies from .13 to .58 depending on age. Further, the D-KEFS manual reports internal consistency reliability coefficients that are lower than would be preferred. In a 2005 study supporting the use of the D-KEFS subtest Design Fluency with patients with frontal lobe epilepsy, researchers (McDonald, Delis, Norman, Tecoma, & Iragui, 2005) found that the subtest did provide adequate discrimination to identify frontal lobe epilepsy distinct from temporal lobe epilepsy. A 2012 study (Possin et al., 2012) of this subtest found that repetition errors effectively discriminated frontotemporal dementia from other participants although total correct design scores from the same subtest did not differentiate the dementia patients from one another specifically. These findings provide support for the D-KEFS in that they indicate that the measure does tap frontal lobe functions specifically.

To meaningfully evaluate the relative improvement of the D-KEFS measures over their predecessors, comparisons between the two are warranted. Correlation coefficients reported in the testing materials range from .30 to .60 in a small sample comparison of the D-KEFS tests and the WCST for Categories Achieved. Correlation coefficients of .20 to .71 were achieved when considering Perseverative Responses on the WCST. The correlation between the Proverb Test Total Achievement score and the WCST Perseverative Response score was the strongest.

Examination of the factor structure of the D-KEFS shows that a three-factor model best fits the data provided in the D-KEFS technical manual (Latzman & Markon, 2010; Miyake et al., 2000). The factors: Conceptual Flexibility, Monitoring, and Inhibition were reflected in scores from different subtests. Conceptual Flexibility was reflected in three scores from the Sorting Test: Confirmed Correct Sorts, Free Sorting Description Score, and

Sort Recognition Description Score (Latzman & Markon, 2010). Monitoring was reflected in scores of category switching on the Verbal Fluency subtest (Latzman & Markon, 2010). Inhibition was reflected in scores from two subtests: the Trail Making Test and Color-Word Interference (Latzman & Markon, 2010). A different three-factor model identifying Inhibition, Updating, and Shifting factors (Miyake et al., 2000) is likely similar to the factors Inhibition, Monitoring, and Cognitive Flexibility, respectively. How well the D-KEFS captures dysfunction along these factors may be a useful gauge in evaluating its utility as a measure of executive function.

Evaluating D-KEFS Subtests as Modern Analogues of Established Measures

Given that the Sorting Test is the most studied of the D-KEFS subtests, a focus on this subtest seems appropriate in establishing the value and utility of the D-KEFS method of adapting existing tests to create an omnibus measure of executive function. Establishing evidence that the Sorting Test is linked to the same neural structures as a well-established measure of executive function, like the WCST, would provide evidence for convergent validity. Given the wealth of research on the WCST, this seems an appropriate measure to use to address this question.

Controlled Oral Word Association Test

Although the WCST is considered the gold standard for tests of executive functioning, there are other measures that are also well supported in the literature. One of these, the controlled Oral Word Association Test (COWAT) requires participants to spontaneously produce words under restricted conditions and limited time (Strauss et al., 2006). The COWAT is used often as a stand-alone measure and in many formats (Strauss et al., 2006) although it is a subtest from a larger battery of tests, the Multilingual Aphasia Examination (Benton, Hamsher, & Sivan, 1978). The COWAT has little administrator burden and can be conducted with just a stopwatch (Strauss et al., 2006). Its many benefits include multiple forms for reducing practice effects. Productivity, as measured on the COWAT, has been shown to be significantly correlated with measures of planning, self-monitoring, and self-regulation (Ownsworth & Shum, 2008) making it a relevant measure to this study. A study of patients with focal frontal lobe lesions found that the COWAT accurately classified the impaired participants (Baldo et al., 2001). Additional studies support the use of the COWAT to investigate executive functions in a variety of populations (Nutter-Upham et al., 2008; Sachs, 2011; Tucha, Smely, & Lange, 1999).

The D-KEFS subtest, Verbal Fluency, is based on the COWAT although it involves many more steps and scores owing to the process approach used in constructing the D-KEFS subtests. The first condition, Letter Fluency Total Correct (Letter Fluency), is precisely the same as the COWAT, utilizing the same time constraints, rule constraints, and even the same letters to prompt the participants' verbal production. The D-KEFS Technical Manual (Delis, Kaplan, & Kramer, 2001) reports that Letter Fluency, a subcomponent of the Verbal Fluency subtest, yields internal consistency coefficients from moderate to high and that test-retest reliability is from good to high. Due to the replicatory nature of the Letter Fluency Task, and acceptable internal consistency as well as test-retest reliability, the D-KEFS subtest Letter Fluency is an appropriate representation of COWAT or general verbal fluency performance.

Stroop Test

The Stroop Test assesses for what is now called the Stroop Effect. It measures reaction time under three conditions: neutral, congruent, and incongruent. The Stroop Effect predicts that reaction time in the incongruent condition is slowest. This was first described by John Stroop after he conducted a series of three experiments and then published his findings regarding reaction time (Stroop, 1935). There have been a range of variants developed of this test over the years, but the Golden version is used frequently (Strauss et al., 2006). In this version, words are used to administer the different conditions with the first condition consisting of words of colors in black ink, the second condition words of colors in the congruent color ink, and the third condition with words of colors in incongruent ink (Golden, Freshwater, & Golden, 2003). There is some disagreement about the principles that result in a slower reaction time in the third condition, with some suggesting that the time to respond to an item on this condition is a combination of the time to read the word and the time to say the color; however, the accepted understanding in the field of neuropsychology is that this task depends on inhibition and suppression such that reaction time is contingent upon the successful employment of these during the task (Adams & Jarrold, 2009). Test-retest reliability is high for the Golden version of the Stroop Test, although practice effects are present (Franzen, Tishelman, Sharp, & Friedman, 1987).

The D-KEFS subtest, Color-Word Interference Test Condition 3 Inhibition (Color-Word) is based on the Stroop Test, more specifically the Golden version that is often categorized as a Stroop Color and Word Test. The D-KEFS version, again, contains additional components as the task has been broken down into its various elements to align with the process approach. Also, the D-KEFS version produces a variety of scores that are not necessarily meaningful corollaries of the traditional Stroop scores. Despite these disparities, the Color-Word obtains a score for a trial in which words are presented in an incongruent ink color and the participant is asked to name the color of the ink and not read the word. These instructions and the method of tracking and scoring the responses mirror the same trial from the Golden version of the Stroop Test. The D-KEFS Technical Manual reports that the test-retest reliability of the Color-Word is in the moderate to high range although this study, too, found evidence of practice effects (Delis et al., 2001). Due to the replicatory nature of the Color-Word task, and acceptable test-retest reliability, the third condition of the D-KEFS subtest Color-Word Condition is an appropriate representation of the Stroop Test or general inhibition performance.

Support for Revised Administration

In keeping with the Standards for Educational and Psychological Testing (American Psychological Association, 1999), any revisions to the Sorting Test need to demonstrate appropriate development, evaluation, and administration. Such study begins the process of exploring the possibilities by developing and testing new administrations that are founded in theory. The addition of theory-driven modifications to this empirically based test may improve the precision with which individual subtests tap the latent construct of interest, executive function, by honing in on those specific components of executive function that each subtest purports to measure. The D-KEFS is predicated upon a conceptualization of executive functions as a diverse set of abilities that are best assessed individually. The test is designed to include assessment of the spectrum of domains sensitive to executive function. The Sorting Test is intended to measure verbal and spatial concept formation, problem-solving initiation, and overall problem-solving (Ramsden, 2003). Additional components of executive functions purportedly measured by the Sorting Test include the ability to explain sorting concepts abstractly, to transfer sorting concepts into action, to inhibit previous sorting responses, flexibility of behavior, and inhibition of previous descriptions responses (Swanson, 2005). In its current form, the Sorting Test provides clear and precise instructions on how to sort the stimuli. It is directly stated in the standardized instructions that the participant is to sort the cards into two groups of three cards in each group and that the three cards in each group must be the same in some way (Delis et al., 2001). Further, the Free Sorting Condition begins with a Practice Card Set and two example items led and narrated by the examiner (Delis et al., 2001). This amount of guidance and the controlled environment of the task are in conflict with the nature of the functions being assessed. As noted earlier, unlike ability tasks, tests of latent constructs must instigate the deployment of the latent construct to complete the task (Gottfredson & Saklofske, 2009). A high amount of examiner control in the testing environment is necessary for standardized assessment but is counterproductive to instigating the deployment of executive functions. It is difficult to assess executive function because standardized testing protocol limit the examiner's ability to initiate self-directed and self-regulated behavior from clients(Stuss, 1991). The paradigm upon which Ach founded sorting tests is that concept formation can be observed through the participant's approach to and completion of the task (Eling et al., 2008). These elements are removed when the

Sorting Test is delivered under the current administration procedures, decreasing the task's ability to tap executive functions as well as it potentially could.

Although there are a variety of definitions of executive function, the failure of the current administration procedure in tapping executive functions can be demonstrated through examining any one of them. Using Lezak et al.'s (2004) conceptualization of executive functions, this level of guidance for the task decreases the necessity of volition in that the conceptualization of the behavior necessary to achieve the tasks is provided. The necessity of planning is greatly diminished with the current instructions because the participant has been informed through both verbal instructions and two in-the-moment examples how to organize the multiple steps of the task to provide the correct answer. Purposive action and self-regulation are adequately assessed with the current instructions because the directions do not limit the necessity for implementing, switching, or stopping behavior sequences nor decrease the likelihood of perseveration or the need for flexibility or productivity. Given the limited ability to observe planning and volition with the current directions, a revision of the administration including less structure and guidance may improve the Sorting Test's ability to tap executive functions as defined by Lezak et al. (2012, 2004).

The current instructions for the Sorting Test also decrease the ability of the task to tap executive functions as conceptualized by Fuster (1991, 2001, 2004). Like Lezak et al. (2012, 2004), Fuster, too, identifies executive functions in a diverse rather than unitary way, consisting of elements that include planning and decision-making (Fuster, 2008). These elements would be impacted by the structured nature of the current administration instructions for the same reasons that were mentioned above. Further, Fuster's understanding of the prefrontal cortex as the temporal organizer of behavior (Fuster, 2004) and, thus, executive functions as mediations of cross-temporal contingencies (Fuster, 1991) is virtually unaddressed by the Sorting Test. In its current format, the Sorting Test does not have a time component in the cycle between movement and sensation such that additional cognitive functions are required to close the cycle. The limited need for the retrospective and prospective skills, short-term memory and preparatory motor set, make the Sorting Test a very poor assessment of executive functions as conceptualized by Fuster. Certainly the Sorting Test is not completely devoid of these elements, as some retrospective short-term memory is required to avoid perseveration on specific sorts and to avoid repetition, but the ability of the task to tap this conceptualization of executive functions may be improved by the addition of feedback throughout the task. The test authors have indicated that the absence of feedback on the Sorting Test is an intentional effort to minimize participant discouragement, particularly in children (Swanson, 2005). Although this is a reasonable goal, the removal of feedback from a task assessing executive function minimizes the task's ability to assess the participant's problem-solving, use of environmental feedback, and appropriate flexibility of behavior. These losses in exchange for providing a more comfortable examination do not seem warranted. Additionally, without the application of feedback, Fuster's conceptualization of executive functions cannot be addressed by this task. If the administration included feedback for the participant, then the task could be said to involve the use of temporally separate information with behavior and sensation separated by time such that executive functions are necessary to complete the task.

Norman and Shallice's model of executive functions (Shallice & Burgess, 1991) is also not addressed by the current Sorting Test administration. The directions for the task, coupled with the Practice Set containing two demonstration items, likely instigate the beginning development of a program or schema for how to approach the task. Norman and Shallice's model states that behavior is activated in a condition-action problem solving format such that where a program or schema exists for the appropriate response to the situation, and adequate resources are available to run that program or schema, then only the lower level is engaged, that which is responsible for routine activity. Granted, two demonstration items does not constitute the development of a routine, but with the additional consideration of the length of the task, the absence of feedback that may indicate a need for changed behavior, and the repetition of the task across two stimuli sets in the first condition, it is less likely that the Sorting Test is accessing the higher level responsible for novel activity, executive functions.

To address these shortcomings and bolster the Sorting Test's ability to tap executive functioning, specifically planning, temporal organization of behavior, and problem solving for novel activity, a revised administration procedure is warranted.

CHAPTER 2

CURRENT INVESTIGATION

This study examined the construct validity of specific measures used to assess executive functioning: sorting tests. For the purposes of this study, three executive function construct conceptualizations were evaluated for relevance to research questions; however, a more comprehensive table depicting additional conceptualizations of this construct is included below for reference (Table 1). The construct tapped by two executive function measures was examined. The overlap in the construct measured by the two tests was examined. The viability of one measure as a substitute for the other was explored.

Table 1

Components of Executive Function According to Multiple Theories

	Theorist	Components of Executive Function
Examined in this study	Lezak	Volition, planning, purposive action, effective performance
	Fuster	Perception-action cycle, temporal gestalts, hierarchal temporal organization of behavior, cross-temporal contingency
	Norman and Shallice	Supervisory attentional system
	Delis et al.	Flexibility of thinking, inhibition, problem-solving, planning, impulse control, concept formation, abstract thinking, creativity
Other notable theories	Baddeley and Hitch	Central executive, phonological loop, visualspatial sketchpad
	Lafleche and Albert	Concurrent manipulation of information: cognitive flexibility, concept formation, cue-directed behavior
	Borkowsky and Burke	Task analysis, strategy control, strategy monitoring
	Anderson et al.	Attentional control, cognitive flexibility, goal setting
	Hobson and Leeds	Planning initiation, perseveration and alteration of goal directed behavior
	Piquet et al.	Concept formation, reasoning, cognitive flexibility
	Elliot	Solving novel problems, modifying behavior in light of new information, generating strategies, sequencing complex actions
	Banich	Purposeful and coordinated organization of behavior, reflection and analysis of the success of the strategies employed

Note. Adapted from "The Elusive Nature of Executive Functions: A Review of our Current Understanding," by M.B. Jurado and M. Rosselli. 2007, *Neuropsychology Review*, *17*(3), p. 213–233. doi:10.1007/s11065-007-9040-z

Letter Fluency Total Correct (Letter Fluency) and Color-Word Interference Test Condition 3 Inhibition (Color-Word) served as additional measures of comparison with which to assess the Delis-Kaplan Executive Function System (D-KEFS) Sorting Test

(Sorting Test), in conjunction with the other well regarded measure, the Wisconsin Card Sorting Test 64-Card Version (WCST). As Color-Word is focused on inhibition and the Sorting Test focuses more heavily on different aspects of executive function—including planning, problem-solving, and set switching—evidence that these measures were not strongly related would support the discriminant validity of the Sorting Test.

It was proposed in this study that the Sorting Test would evidence convergent validity when compared to the WCST and D-KEFS Letter Fluency and divergent validity when compared to the D-KEFS Color-Word. Additionally theoretically-derived alternative administration of the Sorting Test would improve the correlation between the Sorting Test and the WCST and Letter Fluency while not significantly improving the correlation between the Sorting Test and Color-Word. The relative merits of optional scores from the Sorting Test were considered as well, specifically those that may have been represented by scores on the WCST.

Experimental Administration Revisions

For this study, two modified administrations of the Sorting Test were developed. Derived from the aforementioned theoretical considerations of executive function, a Minimal Instruction condition and a Feedback condition were administered. Reduction of the frequency, quantity, and specificity of the Sorting Test instructions was the experimental manipulation in the Minimal Instruction condition. It was believed that a Minimal Instruction condition would activate executive functions more than the original version by requiring self-directed and self-regulated behavior, including planning of how to manipulate the stimuli and problem solving for novel activity.

Provision of feedback to participants regarding the accuracy of their responses on the Sorting Test was the experimental manipulation in the Feedback condition. It was believed that a Feedback condition would instigate a temporal organization of behavior such that feedback from the environment would be utilized to prospectively plan behavior. As a byproduct of this design, it was expected that the Feedback condition administration would enable a more accurate assessment of flexibility and set switching abilities of test takers. This was assessed by determining if correlation between the Sorting Test with the WCST and Letter Fluency improved, because these two measures have already been noted for their use of these component abilities of executive functions.

As Color-Word has been identified as focusing primarily on a different component of executive function, inhibition for the condition of interest, the changed administration was not intended to affect the correlation between Sorting Test and Color-Word.

Hypotheses

It was hypothesized that (a) when minimal instructions or feedback were given for the Sorting Test there would be higher correlations between Confirmed Correct Sorts, Free Sorting Description Score (Description), and Sort Recognition Description Score (Recognition) of the Sorting Test with Categories Completed and Perseverative Responses of the WCST and Total Correct of Letter Fluency.

It was also hypothesized that (b) the administration of the Minimal Instruction condition or Feedback condition would not significantly improve the correlations between the Sorting Test and the Inhibition condition of the Color-Word.

The null hypothesis being tested, then, was that there would be no difference in the correlation between the Sorting Test experimental conditions with WCST, Letter Fluency, or Color-Word when compared to the same correlations with the Control condition.

Lastly, there were hypotheses regarding select optional scores on the Sorting Test and how they may correlate to the WCST. It was hypothesized that (c) the Repeated Sorts score from the Sorting Test might correlate strongly with the Perseverative Responses score of the WCST and the Set -Loss Sorts score from the Sorting Test might correlate strongly with the Failure to Maintain Set score from the WCST. These hypotheses were conjectured because these score pairings purport to examine similar phenomena: perseveration and failure to maintain set, respectively. High correlation between these scores would support examination of other optional scores from the Sorting Test that also lack research.

Summary

Sorting tasks of various sorts have been utilized for over 100 years to assess frontal lobe functioning. The latest iteration of these types of tasks is found on the D-KEFS, however, as the review above indicates, there are some conceptual, administrative, and psychometric problems with this new measure. Specifically, the Sorting Test fails to adequately assess planning, temporal organization of behavior, and problem solving for novel activity, key elements of executive functioning identified by researchers of the frontal lobes and executive functions. It also provides a surplus of scores that have not been satisfactorily supported in the literature as reliable or valid for assessing executive functions and evidences variable correlation with well-established measures of executive function. Thus, a more thorough evaluation of administration utility, psychometric characteristics, and relationship to well-established measures of frontal lobe functions appears warranted. This study examined control data from a non-clinical undergraduate sample to assess the relationship between the Sorting Test and two well-established gold standards of frontal lobe functioning: the Wisconsin Card Sorting Test 64 Card Version (WCST) and Letter Fluency Total Correct (Letter Fluency). Further, the relationship between the Sorting Test and the Color-Word Interference Test (Color-Word), a measure purported to assess a different area of executive function, was also examined. In addition, given problems noted above regarding the administration of the Sorting Test, alternative administrations were conducted with a normal sample and those results were compared to those obtained from controls, who received the standard administration instructions. Also, the relationship of the alternative administrations of the Sorting Test with WCST, Letter Fluency, and Color-Word was evaluated. Finally, as noted above, the Sorting Test provides numerous scores, most of which have not demonstrated adequate reliability or validity in the literature to date. Select optional scores were examined to determine if they correlated with well-established measures of frontal lobe functions.

CHAPTER 3

METHOD

Participant Characteristics

A total of 171 participants were identified for study inclusion, 58 for the control group and 113 for the experimental group. Age of participants ranged from 18-23 years old. Approximately 53 percent of control subjects were 18 years old, 28 percent were 19 years old, and 14 percent were 20 years old. The control group was predominantly White, non-Hispanic (71%) and female (79%). Control subjects were freshman or sophomore undergraduates, with a predominant 72% identifying as freshman. Comprehensive control group demographic characteristics are displayed in (Table 2) below.

Table 2

Participant Demographics in the Control Condition

		Control group *n* =58	% of control
Age	18	31	53.40
	19	16	27.60
	20	8	13.80
	21-23	3	5.20
	Total	58	100
Culture	Asian or Pacific Islander	1	1.70
	Black	11	19.00
	White, non-Hispanic	41	70.70
	Hispanic or Latino	3	5.20
	Bi-Racial	2	3.40
	Total	58	100
Gender	Females	46	79.30
	Males	12	20.70
	Total	58	100
Class	Freshman	42	72.40
	Sophomore	16	27.60
	Total	58	100
Traumatic brain injury	No	48	82.80
	Yes	10	17.20
	Total	58	100.00

The 113 experimental group participants were assigned to two conditions, described further below. The 56 participants in Minimal Instruction condition were predominantly 19 years old (45%), White, non-Hispanic (73%), and freshman undergraduates (77%). Minimal Instruction condition participants were approximately even across gender (52% female, 48% male). Comprehensive Minimal Instruction condition demographic characteristics are displayed in (Table 3) below.

Table 3

Participant Demographics in the Minimal Instruction Condition

			Minimal Instruction condition				
			Group1 *n*=29	% of group 1	Group2 *n*=27	% of group 2	% of Cond. 1 *n*=56
Age	18		9	31	5	18.50	25
	19		14	48.30	11	40.70	44.60
	20		1	3.40	5	18.50	10.70
	21-23		5	17.20	6	22.20	19.60
		Total	29	100	27	100	100
Culture	Asian/ Pac. Isl.		2	6.90	1	3.70	5.40
	Black		6	20.70	5	18.50	19.60
	White, non-Hisp		20	69	21	77.80	73.20
	Hisp/ Latino		1	3.40	0	0	1.80
		Total	29	100	27	100	100
Gender	F		17	58.60	12	44.40	51.80
	M		12	41.40	15	55.60	48.20
		Total	29	100	27	100	100
Class	Fresh		26	89.70	17	63	76.80
	Soph		1	3.40	6	22.20	12.50
	Jun		1	3.40	1	3.70	3.60
	Sen		1	3.40	3	11.10	7.10
		Total	29	100	27	100	100
Hx test	No		27	93.10	25	92.60	92.90
	Yes		2	6.90	2	7.40	7.10
		Total	29	100	27	100	100
TBI	No		23	79.30	24	88.90	83.90
	Yes		6	20.70	3	11.10	16.10
		Total	29	100	27	100	100
ESL	No		27	93.10	26	96.30	94.60
	Yes		2	6.90	1	3.70	5.40
		Total	29	100	27	100	100
Misc	No		28	96.60	26	96.30	96.40
	Yes		1	3.40	1	3.70	3.60
		Total	29	100	27	100	100

Note. Hisp. = Hispanic; Pac. Isl. = Pacific Islander; Hx test = previous exposure to a sorting test; TBI = traumatic brain injury; ESL = English as second language; Misc.= other condition of possible relevance but of low frequency in the study sample (e.g., low reading level).

The 57 participants in the Feedback condition were predominantly 19 years old (47%), White, non-Hispanic (70%), and freshman undergraduates (75%). The Feedback condition was 46% female and 54% male. Comprehensive Feedback condition demographic characteristics are displayed in (Table 4) below.

Table 4

Participant Demographics in the Feedback Condition

		Feedback condition				
		Group3 *n*=29	% of group 3	Group4 *n*=28	% of group 4	% of Cond. 2 *n*=57
Age	18	6	20.70	5	17.90	19.30
	19	11	37.90	16	57.10	47.40
	20	6	20.70	4	14.30	17.50
	21-23	5	17.20	3	10.70	14
	24-26	1	3.40	0	0	1.80
	Total	29	100	28	100	100
Culture	Asian/ Pacific Isl.	3	10.30	1	3.60	7
	Black	4	13.80	6	21.40	17.50
	White, non-Hisp	21	72.40	19	67.90	70.20
	Hispanic/ Latino	1	3.40	2	7.10	5.30
	Total	29	100	28	100	100
Gender	F	13	44.80	13	46.40	45.60
	M	16	55.20	15	53.60	54.40
	Total	29	100	28	100	100
Class	Fresh	20	69	23	82.10	75.40
	Soph	7	24.10	2	7.10	15.80
	Jun	1	3.40	2	7.10	5.30
	Sen	1	3.40	1	3.60	3.50
	Total	29	100	28	100	100
Hx Tested	No	27	93.10	26	92.90	93
	Yes	2	6.90	2	7.10	7
	Total	29	100	28	100	100
TBI	No	20	69	24	85.70	77.20
	Yes	9	31	4	14.30	22.80
	Total	29	100	28	100	100
ESL	No	26	89.70	27	96.40	93
	Yes	3	10.30	1	3.60	7
	Total	29	100	28	100	100
Misc	No	27	93.10	28	100	93
	Yes	2	6.90	0	0	7
	Total	29	100	28	100	100

Note. Hisp. = Hispanic; Pac. Isl. = Pacific Islander; Hx test = previous exposure to a sorting test; TBI = traumatic brain injury; ESL = English as second language; Misc.= other condition of possible relevance but of low frequency in the study sample (e.g., low reading level).

Conditions were evaluated for significant differences in demographic characteristics through chi-square tests and independent t-tests. As the demographic characteristics of culture, class, gender, and age were collected as binary, nominal, or ordinal categorical variables, conditions were evaluated for independence along these characteristics with chi-square tests. IQ, as a continuous variable, was evaluated for significant difference between conditions with an independent t-test. When compared to each other, all conditions had similar representations of culture groups and current college classes. Specifically, Minimal Instruction condition was found to not significantly differ from the Control condition in terms of culture group, $x^2(6) = 2.96$, or current college class, $x^2(3) = 9.50$, at the Bonferroni-adjusted significance level of (.05/4=.013). No significant difference in culture group, $x^2(3) = 1.85$, p>.05, or current college class, $x^2(3) = 6.96$, p>.05, was found between Feedback condition as compared to the Control condition. Similarly, Minimal Instruction condition and Feedback condition were not significantly different in terms of culture group, $x^2(3) = 1.19$, p>.05, or current college class, $x^2(3) = 1.11$, p>.05. Minimal Instruction condition and Feedback condition were also not significantly different in terms of IQ, t(104) = .23; age, $x^2(4) = 2.90$; or gender, $x^2(1) = .43$ (all ps>.05).

Some differences between conditions for demographic characteristics were significant. Participant age was significantly different between experimental conditions and the Control condition at p<.05. After adjusting for family wise error, Minimal Instruction condition, $x^2(3) = 13.22$, and Feedback condition, $x^2(4) = 15.83$, were both significantly different at the Bonferroni-adjusted significance level (all ps<.013). Specifically, age represents a medium association in Minimal Instruction condition by control, with a Cramer's statistic of .34, p<.01. Between Feedback condition and the Control condition, the standardized residual for age group 18 years old was beyond 1.96 with the odds of a participant being 18 years old 4.80 times higher in the Control condition than in Feedback condition, based on the odds ratio.

Both experimental conditions were also found to significantly differ from the Control condition in gender representation. Using a Bonferroni-adjusted significance level, both Minimal Instruction condition, $x^2(1) = 9.59$, and Feedback condition, $x^2(1) = 13.94$, were significantly different from the Control condition in terms of gender at p<.013.

Specifically, based on the odds ratio, the odds of a participant being a female were 4.57 times higher in the Control condition than in Feedback condition and 3.57 times higher in Minimal Instruction condition than in the Control condition.

Sampling Procedures

All participants were undergraduates in introductory psychology classes who were enrolled in the subject pool. With the approval of the departmental and Institutional Internal Review Board for research with human subjects, students enrolled in the subject pool were recruited for one of two studies. Participants recruited for this study were placed in the experimental conditions. Participants recruited for a study evaluating the relationship between executive function and tobacco dependence and response to treatment in undergraduates (Principle Investigator: William Meil, Ph.D.) were placed in the control condition. Students were not permitted to participate in both studies.

For recruiting the participants for the other study, research assistants received a list of possible participants from the subject pool along with their contact information. Recruitment was conducted via email and telephone, with alignment of availability and failure to respond to three attempted contacts as the only exclusion criteria. Of the resultant pool of participants, 58 subjects were randomly selected for inclusion in this study as the Control condition participants.

For recruiting the participants for the experimental conditions, students were able to schedule a research session for this study through Sona software. No exclusionary criteria were set at this stage of recruitment. This self-selection process yielded 113 experimental participants for part two of this study.

Sample Size, Power, and Precision

A power analysis was conducted using G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007), which established that for an effect size of .15 and a power of .80, approximately 50 participants were necessary for the Control condition and each experimental condition. In anticipation of administration, scoring, or participant-specific issues that could spoil individual datum, greater than 50 participants were recruited for each condition. Where data were not available or were "spoiled" on some but not all variables for a given participant, cases were excluded pairwise to retain power. Both the control and experimental condition samples were comprised of largely homogenous participants, with restricted range of age, culture, and educational attainment.

Measures and Covariates

North American Adult Reading Test (NAART)

The NAART is a performance-based proxy of IQ that was modified for applicability with North American (Blair & Spreen, 1989). The original measure, the National Adult Reading Test, was developed in 1978 and modified in 1991 (Nelson & O'Connell, 1978; Nelson & Willison, 1991). The NAART is comprised of 61 increasingly complex non-phonetic North American English words that participants are tasked to read aloud. Beginning with the words "debt" and "debris," the difficulty of the pronunciation and its deviation from apparent phonetic properties increases, ending with the words "talipes" and "synecdoche." Correct pronunciation of each word yields one-point and total points earned are transformed to an estimated standard score IQ. This information was desirable to assess for the confounding influence of IQ on performance of participants on executive function measures. Use of the NAART to approximate IQ is supported by high correlations between NAART scores and Wechsler Adult Intelligence Scale full scale IQ (FSIQ) scores, $r = .75$, $p<.001$ (Blair & Spreen, 1989), $r = .60$, $p<.001$ (McGurn et al., 2004). NAART scores have become a field standard for estimating IQ for both normal and clinical populations (Crawford, Allan, Cochrane, & Parker, 1990). NAART scores have been found to highly correlate with premorbid FSIQ for individuals with schizophrenia (Crawford et al., 1992), traumatic brain injury, dementia, and other neurological impairments (Johnstone, Callahan, Kapila, & Bouman, 1996) in addition to healthy controls (Wiens, Bryan, & Crossen, 1993)—making it an ideal measure to use with the undergraduate population sampled for this study.

Delis-Kaplan Executive Function System (D-KEFS), Sorting Test

The Sorting Test (Delis et al., 2001) is comprised of a Screening Pretest, a Practice Card Set, a Free Sorting condition, and a Sort Recognition condition, as summarized in Table 6 below and fully detailed in Appendix C. The Screening Pretest requires participants to read a list of words aloud and indicate if there are any unknown words on the list. The Practice Card Set contains instructions for how the task is to be completed, along with two demonstration items in which the administrator both shows and describes how to accurately complete the task. The Free Sorting condition requires participants to sort stimuli cards into two groups and then describe how the cards were sorted into each group. This is discontinued after four minutes of elapsed sort time, ten attempted sorts, or an indication from the participant that they are unable to construct any other sorts. Then a new set of stimuli cards are used to complete the task again following the same rules. Next, the Sort Description Condition requires the participant to describe how the examiner has sorted the stimuli cards into two groups. This is completed for both sets of cards. Several scores are produced with this subtest, including 5 primary and 29 optional scores. The main primary scores are Confirmed Correct Sorts, Description, and Recognition. Confirmed Correct Sorts are those sorts that match a target sort and are described at least partially accurately, Description is the score for how well the participant articulated how they sorted each group, and the Recognition is for how well the participant articulated how the administrator sorted each group.

Internal consistency for the Sorting Test is difficult to determine as sorts within a card set are interdependent (Delis et al., 2001). Internal consistency as estimated by comparison between the two card sets, which are independent and function as two equivalent half tests, ranges from .72 to .86 in the Free Sorting Confirmed Sorts Condition, .73 to .83 in the Free Sorting Description condition, and .70 to .81 in the Sort Recognition Total Condition for ages 16-89 (Delis et al., 2001). Data reported in the technical manual for lower age ranges were irrelevant to the present study, which used undergraduate participants. Test-retest reliability for the Description Confirmed Sorts Condition is .51 for all ages 8-89, .50 for Free Sorting Descriptions, and .60 for Sort Total (Delis et al., 2001). The correlations between Sorting Test measures are generally positive and part-whole correlation yielded robust results (Delis et al., 2001). Association between card sets is in the moderate range and free sorting and sort recognition correlate moderately with one another (Delis et al., 2001). The Sorting Test is the only subtest of the D-KEFS for which validity data are reported in the Technical Manual, however, all of the data are from studies of the California Card Sorting Test, the measure on which the Sorting Test is based. Presuming the equivalency of these measures, validity for the use of the Sorting Test has been demonstrated through studies with focal frontal lesions (Dimitrov et al., 1999), Parkinson's (Beatty & Monson, 1990), chronic alcoholism (Beatty, Katzung, Nixon, & Moreland, 1993), schizophrenia (Beatty, Jocic, Monson, & Katzung, 1994), multiple sclerosis (Beatty & Monson, 1996), lobectomy (Crouch, Greve, & Brooks, 1996), and neurologically intact young adults (Greve, Farrell, Besson, & Crouch, 1995) among others.

Delis-Kaplan Executive Function System (D-KEFS), Letter Fluency

Letter Fluency (Delis et al., 2001) is a spontaneous production task based on the COWAT (Benton et al., 1978), in which the participant is asked to produce as many words as possible under a set of restrictions in a limited amount of time. This particular verbal fluency task involves three trials, one each for the letters F, A, and S, completed in 60 seconds each. The restrictions on the task are that the participant is not to provide names of people, places, or numbers or the same word with different endings (e.g., -ing). The limited administrator burden (e.g., no specific materials required beyond a stopwatch, administration time under five minutes, and simple scoring), in addition to its psychometric properties, makes Letter Fluency Total Correct (Letter Fluency) a popular measure of executive functions.

Using total number of words for each letter to assess for internal consistency, coefficient alpha between F, A, and S was found to be .83 (Ruff, Light, Parker, & Levin, 1996; Tombaugh, Kozak, & Rees, 1999). Test-retest reliability in healthy adults is high, across short and long durations (e.g. one week to five years; Levine, Miller, Becker, Selnes, &

Cohen, 2004; Tombaugh et al., 1999). High inter-rater reliability has also been demonstrated (Ross, 2003). High correlations between .85 to .94 are found between phonemic fluency tasks, such as the FAS and CFL forms of the COWAT (Cohen & Stanczak, 2000).

Delis-Kaplan Executive Function System (D-KEFS), Color-Word

Color-Word Interference Test Condition 3, Inhibition (Color-Ward) (Delis et al., 2001) is a new iteration of the Stroop Test, first designed to assess for what John Stroop identified in a series of experiments (Stroop, 1935). The D-KEFS subtest most resembles the popular Golden version of the Stroop Color and Word Test (Golden et al., 2003) and like its predecessor assesses reaction time as it is effected by inhibition. This task is administered in four conditions: Color Naming, Word Reading, Inhibition, and Inhibition/Switching. The first two conditions are just as they sound, naming patches of color and reading words of colors written in black ink. It is the third condition that is of importance to this study. The Inhibition condition measures the length of time it takes to name the color of ink that words are printed in when the words spell out an incongruent color. This task requires that the reading of the word be inhibited while only the ink color is spoken aloud by the participant. This task is highly focused on the executive function of inhibition, making it an excellent tool for supporting the divergent validity of the Sorting Test, which focuses on slightly on inhibition. The validity of the Stroop Test has been assessed for a variety of populations including children with autism (Adams & Jarrold, 2009), groups (Hakoda & Sasaki, 1990), ADHD (Savitz & Jansen, 2003), and even when administered in Spanish to children (Armengol, 2002). Test-retest reliability is moderate to high for the D-KEFS version of the Stroop Test and reliability is good form most age groups (Delis et al., 2001).

Wisconsin Card Sorting Test (WCST), 64-Card Version

The Wisconsin Card Sorting Test (WCST) 64-card version (Kongs, Thompson, Iverson, & Heaton, 1981) requires the participant to match 64 stimulus cards to one of 4 keys cards. Each successful match to a target sorting rule is signaled by the examiner with the response "Correct" or "Right" and each unsuccessful match to a target sort is signaled by the examiner with the response "Incorrect" or "Wrong." After successfully matching the target sort consecutively ten times, the target sort rule is changed without notice. There are three target sort categories: color, form, and number. The test is complete after the 64 trials have been completed. The cards are numbered from 1-64 on the back to enable the examiner to administer the cards in the same order each time. The test takes approximately 20-30 minutes to complete and delivers 16 scores, although the main scores utilized for the assessment of executive functions are Categories Completed and Perseverative Responses. Categories Completed is the number of target sort categories achieved; with achievement of a category occurring after 10 consecutive correct sorts for the current category. Perseverative Responses are those that match a previously correct category (Parmenter et al., 2007).

Neuroimaging supports the use of the WCST to assess dorsolateral prefrontal cortex functions (Sumitani et al., 2006) and this is supported in meta-analyses (Demakis, 2003), as well. The validity of the WCST for assessing multiple populations has been demonstrated in the research, including multiple sclerosis (Parmenter et al., 2007), older adults

(Ettenhofer et al., 2006), autism (Ozonoff, 1995), Alzheimer's (Terada et al., 2011), and Parkinson's (Alevriadou, Katsarou, Bostantijopoulou, Kiosseoglou, & Mentenopoulos, 1999) to name a few. Test-retest reliability of the WCST is not meaningful to compute, as the test is not a reliable measure to use with subjects who have been previously exposed to the task. The nature of the problem-solving task is changed once an individual is aware of the mechanics of the task, the need to switch sets, and sort on new principles, so the measure is limited by its one-time use expiration.

Unlike the Sorting Test, the WCST is available in multiple formats including 64 card and 128 card manual versions as well as a computer-administered version. Many clinicians use the results from any of these versions as equal to another, but the manual and the computer versions of the WCST have been shown to have incomplete psychometric equivalence (Steinmetz, Brunner, Loarer, & Houssemand, 2010) so for this investigation, the commonly administered 64 card manual version will be used exclusively.

Research Design

Participants were recruited through Sona and presented to the scheduled research room in Uhler Hall on IUP's campus at the appointed time. At the start of the session, informed consent to participate in the study was obtained. The informed consent form (Appendix A) was verbally summarized and reviewed with participants prior to obtaining their signature to signify consent. Signed consent forms were stored separately from data collected to maintain confidentiality of the collected data. An unsigned copy was provided to the participant for their records. Next, the participant completed a demographics questionnaire (Appendix B). This self-report form contained questions regarding age, culture, gender, and college year. Participants were then administered measures of cognitive and executive function. Scores retained from the D-KEFS for analysis were on three subtests: Sorting Test, Letter Fluency, and Color-Word. On the Sorting Test, the following scores were retained: Confirmed Correct Sorts, Free Sorting Description (Description), and Sort Recognition Description Score (Recognition). On Letter Fluency the Total Correct score was retained. On Color-Word score was retained. From the WCST, Categories Completed and Perseverative Responses were retained for analysis. Optional Sorting Test scores Repeated Sorts and Set Loss Sorts were retained for exploratory correlational analyses with the WCST scores Perseverative Responses and Failure to Maintain Set.

Following the administration of the measures, participants were thanked for their time and provided a debriefing form (Appendix D). This form provided further information and resources regarding the study, which were verbally summarized for the participant. The consent form, demographics questionnaire, and debriefing forms administered to experimental condition participants closely mirrored those administered to control participants.

Assignment to the control and experimental conditions was made via random assignment. Within the experimental conditions, participants were assigned to one of the research protocols in order, such that a participant was assigned to group one, the next participant to group two, then three, and then four. Assignment would then begin, again, at group one. This assignment method was used to account for confounds in semester attitudes of the participants, minimize the harm to the project if the desired number of participants was not achieved, assure balanced groups, and assure adequate power for each group.

For evaluation of the construct validity of the Sorting Test, the Control condition was analyzed. These same data were next used in part two of this study to serve as a comparison group for the experimental conditions.

Experimental Manipulations

The above referenced research protocols, or groups, dictated which measures the participants were administered and the order of the test delivery, as displayed in Table 5 below. Groups 1 and 2 were under the Minimal Instruction condition and groups 3 and 4 were under the Feedback condition. Both conditions were administered the informed consent, demographics questionnaire, NAART, WCST, Letter Fluency, and Color-Word. Within these two conditions, group membership dictated the order in which the measures were administered. Groups 1 and 3 were administered the informed consent, demographics questionnaire, NAART, and the applicable version of the Sorting Test followed by Letter Fluency, Color-Word, and then the WCST. Groups 2 and 4 were administered the informed consent, demographics questionnaire, NAART, and the WCST followed by Letter Fluency, Color-Word, and the applicable version of the Sorting Test. This counterbalancing was designed to inure the results against order effects. This prevents systematic variation, such as practice effects and boredom effects (Field, 2013). The steps for data collection for the two protocols under the Minimal Instruction condition varied by order of test administration, only. The same was true for the Feedback condition groups.

Table 5

Order of Test Administration in the Four Research Protocols

Minimal Instruction condition		Feedback condition	
Group 1	Group 2	Group 3	Group 4
1. Informed Consent Form 2. Demographics Questionnaire 3. NAART **4. Minimal Instruction Condition** 5. Letter Fluency 6. Color-Word Interference **7. WCST** 8. Debriefing	1. Informed Consent Form 2. Demographics Questionnaire 3. NAART **4. WCST** 5. Letter Fluency 6. Color-Word Interference **7. Minimal Instruction Condition** 8. Debriefing	1. Informed Consent Form 2. Demographics Questionnaire 3. NAART **4. Feedback Condition** 5. Letter Fluency 6. Color-Word Interference **7. WCST** 8. Debriefing	1. Informed Consent Form 2. Demographics Questionnaire 3. NAART **4. WCST** 5. Letter Fluency 6. Color-Word Interference 7. **Feedback** Condition 8. Debriefing

Sorting Test Alternative Administration, Minimal Instruction Condition

The Minimal Instruction condition is an alternative administration format of the Sorting Test developed for this study. It differs from the standard administration in several key ways, as summarized in (Table 6) below and fully detailed in Appendix C. First, the Screening Pretest was removed. Although this eliminated the ability to assess for literacy and familiarity with the words contained on the stimuli cards directly, these were indirectly assessed through numerous other components of the neuropsychological assessment. As the Sorting Test is unlikely to be administered alone, illiteracy would have been apparent through the administration of other measures and did not need to be specifically addressed in the Sorting Test.

Another change in this alternative administration was the removal of the Practice Card Set and all of the related instructions and demonstration items. The removal of these components was designed to minimize the inappropriate amount of instruction and guidance for a task assessing executive functions. Exposure to the task, such as provided in the Practice Card Set, may decrease the task's ability to assess the latent construct of executive function due to loss of task novelty. Removal of this section resulted in more onerous demands on the volition, planning, and problem-solving behavior of the participant.

Changes to the Description condition included the removal of the specific instruction to place three cards in each group, which was replaced with the instruction to use all of the cards each time. In the standard administration, participants are instructed to make only two groups with three cards in each group. This degree of specificity inappropriately minimizes the participant's use of planning, thereby decreasing the task's ability to tap executive functioning. The replacement of this instruction with one indicating the need to use all of the cards each time requires the participant to engage in planning how to meet this demand while still making only two groups in which all of the cards in each group are the same in some way. This replacement instruction also serves to maintain the same scoring protocol so that additional components of the Sorting Test need not be modified to acquire the popular scoring category Confirmed Correct Sorts.

The repetition of instructions for the Description condition Card Set 2 was also removed, although Card Set 2 was still administered. This enabled assessment of flexibility and switching, as the participant needed to adapt to the changed stimuli without guidance from the administrator. The repeated instructions were also removed to minimize the task becoming rote and therein failing to assess executive functions.

Finally, pages containing rule summaries for the participant's viewing during the task were also removed. This was done to increase the demand on the participant and better align with other sorting tests and real-life situations. In other sorting tests and in real-life situations, executive functions are necessary in the absence of clear task instruction. Removal of the summary pages containing relevant task rules better aligns the task demand with other sorting tests and real-life situations. Similarly, Recognition Condition instructions prefacing the introduction of Card Set 2 were removed, as well.

Sorting Test Alternative Administration, Feedback Condition

The Feedback Condition is an alternative administration format of the Sorting Test developed for this study, as summarized in (Table 6) below and fully detailed in Appendix C. The standard administration instructions were retained with an additional line included that indicated that the administrator would be providing feedback to the participant about their attempted sort. For ease of comparison and to minimize participant confusion, this line was delivered with the same wording as that of the one included in the standard administration of the WCST. It was expected that this alternative administration would provide additional information about executive functioning, specifically the ability of the participant to integrate feedback from the environment to modify behavior and achieve a desired result. The standard administration of the Sorting Test is unable to assess this specific function because it lacks the provision of feedback. The addition of feedback to this task was intended to broaden the scope of executive functions assessed by the Sorting Test as well as make the test more similar to other sorting tasks such that research in the area of sorting tests in general would be more applicable to this specific test. Positive feedback was delivered only in response to completed target sorts, although there exist other possible sorting configurations that meet the conditions of the task as presented to the participant. This is in alignment with the native scoring protocol that scores target and non-target sorts separately, even if the sort and description is accurate. Negative feedback was delivered in response to any sort that was not a target sort, including failures to use all of the cards or to make two groups. Feedback was additionally administered during the Recognition Conditions. Incorporation of feedback to alter behavior was applicable to the prospective planning of this task condition in that confirmed descriptions should not be repeated. This nuanced category of repetition was intended to add another dimension to the test's ability to assess planning.

Table 6

Differences Between Sorting Test Administration in Each Condition

D-KEFS Sorting Test	Control condition	Minimal Instruction Condition	Feedback Condition
Screening Pretest	Asked participant to read a list of words aloud, identify words of unknown meaning, and provided definition of unknown words.	Removed	Retained in standard format
Practice Card Set	Modeled the task twice and answered questions	Removed	Retained in standard format
Free Sorting Condition, Card Set 1	Explained sorting rules, including number of sorting groups, number of cards in each group, and type and amount of description required. Provided a page summarizing rules to be available throughout the task	Removed reference to practice card set Replaced instruction to use three cards in each group with instruction to use all the cards Removed rule summary page and associated instructions	Retained with additional instruction: I cannot tell you how to sort the cards but I will tell you each time whether you are right or wrong.
Free Sorting Condition, Card Set 2	Explained sorting rules, including number of sorting groups, number of cards in each group, and type and amount of description required. Provided a page summarizing rules to be available throughout the task	Removed instructions, stimuli switched without further guidance	Retained with additional instruction: I cannot tell you how to sort the cards but I will tell you each time whether you are right or wrong.
Sort Recognition Condition, Card Set 1	Explained that the examiner would next complete sorts for participant to identify Informed participant of number of sort groups, number of cards in each group, and type and amount of description required	Retained in standard format	Retained with additional instruction: I cannot tell you how to sort the cards but I will tell you each time whether you are right or wrong.
Sort Recognition, Card Set 2	Explained that the examiner would next complete sorts for participant to identify. Informed participant of number of sort groups, number of cards in each group, and type and amount of description required.	Removed instructions, stimuli switched without further guidance	Retained with additional instruction: I cannot tell you how to sort the cards but I will tell you each time whether you are right or wrong.

Note. For verbatim comparison of D-KEFS Sorting Test instructions by condition see Appendix C.

Recruitment

For this study, a total of 171 participants were recruited. The Control condition was comprised of 58 participants, Minimal Instruction condition of 56 participants, and Feedback condition of 57 participants. Frequencies of participant demographic characteristics for each condition are reflected in Table 7 below. Conditions did not differ significantly on culture group, current college class, or IQ (see Participant Characteristics above). Conditions 1 and 2 significantly differed from the Control condition on gender and age, as reported in Participant Characteristics above.

The demographic characteristics on which conditions significantly differed were analyzed for impact on response variables. Although conditions 1 and 2 were found to significantly differ from the Control condition across gender, gender was not found to significantly impact performance in this study. Independent t-tests of response variables yielded no significant difference in performance for women as compared to men, all $ps > .05$.

Similarly, age was not found to significantly impact most performance variables in this study. Assessing this required that one case be excluded listwise due to age group 24-26 having only one case. With this case excluded, data were confirmed to meet the assumption of homogeneity of variance using Levene's test. Two variables violated this assumption, Categories Completed, $F(3, 145) = 5.08$, $p < .01$, and Failure to Maintain Set, $F(3, 145) = 6.07$, $p < .01$. As transformations of the data to achieve homogeneity of variance were not successful (see detailed review in Assumptions) and not expected to help (Field, 2013), Welch's F-ratio (1951) was used to adjust F and df_r. Accounting for sample size and variance for each age group, in order to minimize the impact of violating the assumption of homogeneity of variance for these two response variables. Welch's F-ratio was chosen as it has been reviewed along with other techniques (Tomarken & Serlin, 1986) and deemed to control for Type I error while retaining more power than the Brown and Forsythe F-ratio (1974). One-way independent ANOVA on all other response variables by age indicated nonsignificant difference between groups. Using the Welch F-ratio adjustment, there was a significant effect of age on Categories Completed, $F(3, 44.02) = 6.08$, $p<.01$, but not on Failure to Maintain Set, $F(3, 45.16) = 1.40$, $p>.05$.

To further inspect the effect of age on Categories Completed, post hoc tests were analyzed. For Categories Completed, the REGWQ test revealed that the nonhomogeneous group mean on this variable was from 18 year old participants. Using the Games-Howell post hoc test due to heterogeneity of variance, 18-year-old participants were found to significantly differ from 19-23 year old participants on Categories Completed (all ps<.05). Where Categories Completed was included in primary study analyses (see Convergent Validity below), results included an additional analysis partialling out the effect of age (i.e., Tables 15, 16, and 17). Although age was found to significantly impact Categories Completed, accounting for age in this way did not yield a change in the significance findings in any of the conditions. Partialling out the effect of age on Categories Completed after excluding the 1 case in age group 24-26, did change the significance findings of Categories Completed as correlated with Confirmed Correct Sorts in Minimal Instruction condition, (Categories Completed), $r= -.31$, $p<.05$ became (Categories Completed), $r= -.28$, $p>.05$, *NS.* This change was negligible to the findings of this study, however, as the correlation between Categories Completed and Confirmed Correct Sorts in Minimal Instruction condition, while statistically significant, was deemed nonsignificant when evaluated at the Bonferroni-adjusted significance level for that analysis (.05/11), p>.005.

All experimental participants were administered both the Sorting Test and the WCST. This made it possible to study correlation between the Sorting Test and the WCST (see Correlation with the "Gold Standard" and CURRENT INVESTIGATION above for complete rationale). The present study evaluated the impact of non-standard theoretically derived administration of the Sorting Test, as used in the experimental conditions of this study, on the test's correlation with other measures of executive function (i.e., WCST, Letter Fluency, Color-Word). This was compared to the impact of the standard atheoretical administration of the Sorting Test, as used in the Control condition of this study, on the same additional measures of executive function. To avoid between subjects variability impacting the comparisons of performance on these measures, all study participants were administered both the WCST and a version of the Sorting Test. This served to hold between subjects variability constant across test measures.

The D-KEFS manual reports standard deviations of test-retest scores on some tests that are larger at second testing. This may suggest weak reliability but is not expected to impact the findings of this study. Although research demonstrating negligible effects of order of test presentation is available (Cassel, Johnson, & Burns, 1962), counterbalancing was conducted to inure against such effects given the similarity between the Sorting Test and the WCST. Both Minimal Instruction condition and Feedback condition were comprised of two groups, one that received the Sorting Test before the WCST and one that received the WCST before the Sorting Test. This parallel construction of the two experimental conditions with groups counterbalanced for test presentation order enabled the groups to be combined to create the two experimental conditions without necessitating group by group by condition comparison. Additionally, t-tests for order effects of group 1 compared to group 2 as well as group 3 compared to group 4 did not yield significant mean differences in primary variables after controlling for family wise error, $p>.005$ (Table 7). Although the control data did not reflect test presentation order, these comparisons suggest that the control data are unlikely to be significantly impacted by order of test presentation.

Table 7

Percentage of Participant Demographics in Each Condition

			% of Cond. 1 *n*=56	% of Cond. 2 *n*=57	% of control *n*=58
Age	**18**		25	19.30	**53.40**
	19		**44.60**	**47.40**	27.60
	20		10.70	17.50	13.80
	21-23		19.60	14	5.20
	24-26		0	1.80	0
		Total	100	100	100
Culture	Asian/ Pacific Isl.		5.40	7	1.70
	Black		19.60	17.50	19
	White, non-Hisp		**73.20**	**70.20**	**70.70**
	Hispanic/ Latino		1.80	5.30	5.20
	Biracial		0	0	3.40
		Total	100	100	100
Gender	F		51.80	45.60	79.30
	M		48.20	54.40	20.70
		Total	100	100	100
Class	**Fresh**		**76.80**	**75.40**	**72.40**
	Soph		12.50	15.80	27.60
	Jun		3.60	5.30	0
	Sen		7.10	3.50	0
		Total	100	100	100
Hx Tested	**No**		**92.90**	**93**	--
	Yes		7.10	7	--
		Total	100	100	--
TBI	**No**		**83.90**	**77.20**	**82.80**
	Yes		16.10	22.80	17.20
		Total	100	100	100
ESL	**No**		**94.60**	**93**	--
	Yes		5.40	7	--
		Total	100	100	--
Misc	**No**		**96.40**	**93**	--
	Yes		3.60	7	--
		Total	100	100	--

Note. -- = factor not assessed in Control condition; % = percentage; Cond. = condition; Hisp. = Hispanic; Pac. Isl. = Pacific Islander; Hx test = previous exposure to a sorting test; TBI = traumatic brain injury; ESL = English as second language; Misc.= other condition of possible relevance but of low frequency in the study sample (e.g., low reading level).

Due to administration, scoring, and participant specific issues, the number of responses used for analyses in this study varies along variables. For example, 111 of 113 experimental participants and 55 of 58 control participants yielded useable scores on the Sorting Test variable Confirmed Correct Sorts. These variations in sample size occurred when known errors of administration or scoring protocol were present. For such cases, responses were excluded from analyses pairwise. Number of responses on each variable of interest by condition are provided in the table below.

Table 8

Number of Responses in Each Condition for Each Variable With T Tests That Supported Collapsing Research Protocol Groups Within the Conditions

Variable	Minimal Instruction condition	Feedback condition	Control
CCS	56 t(54) = -2.01, p>.005	54 t(52) = -.62, p>.005	55
Free	56 t(54) = -2.25, p>.005	54 t(52) = -.05, p>.005	55
Rec	56 t(54) = -.52, p>.005	51 t(48) = .86, p>.005	58
Rep	56 t(54) = 1.39, p>.005	54 t(41.008) = 2.77, p>.005	0
LF	56 t(54) = .38, p>.005	55 t(52) = .37, p>.005	58
CWI	56 t(54) = .19, p>.005	55 t(52) = -.16, p>.005	57
Set-Loss	56 t(54) = -.05, p>.005	55 t(26) = 1, p>.005	0
Cat	53 t(51) = 1.06, p>.005	50 t(47) = .81, p>.005	54
PR	53 t(51) = -.49, p>.005	50 t(37.137) = -1.36, p>.005	54
WFMS	53 t(51) = -.84, p>.005	50 t(47) = -.072, p>.005	54

Note. CCS = confirmed correct sorts; Free = free sorting description score; Rec = sort recognition description score; Rep = repeated sorts, LF = letter fluency total correct index; CWI = color-word interference inhibition condition; Set-Loss = set-loss sorts; Cat= categories completed; PR = perseverative responses; WFMS = WCST failure to maintain set.
[a]significant at p<.005, Bonferroni-adjusted significance level

Statistics and Data Analyses

Data analyzed in this study were obtained by scoring the responses from each participant in accordance with published test standards and deriving standardized scores based on normative data. For the D-KEFS, raw scores were converted to scaled scores with a mean of 10 and a standard deviation of 3 using the D-KEFS Scoring Assistant Software. Scaled scores were produced by the software based on normative data from a stratified standardization sample. The standardization sample included 1,750 individuals stratified on age, sex, race, education, and geographic region to approach target values generated from the 2000 U.S. Census figures. Select process measure raw scores were computed to cumulative percentile ranks, described in further detail below. For sections of the D-KEFS in which high, rather than low, raw scores reflect poor performance, the scoring software incorporated the inversion necessary to produce intuitively oriented scaled scores. For the WCST, computer-scoring software was used to compute raw scores from participant responses. Next, z-scores were derived from the WCST raw scores. Normative data stratified by age and education for the 64-card version of the WCST were obtained from A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary (Strauss et al., 2006) and used to calculate z-scores.

Assumptions

In preparation for statistical analyses, each variable response set was analyzed for compliance with the assumptions of parametric data. Variables were evaluated by condition because the intended analyses for investigating the study hypotheses involved comparing an experimental condition to the Control condition, making the distribution in each condition more important than that of the overall distribution of each variable. Cases with missing values were excluded pairwise to maximize use of the available data.

Interval Level Measurement and Independence

Variables investigated were measured at the interval level, with one exception. Scores along most variables were numerals of known order values and known difference values between numerals. This identifies the variables as measured at the interval level required of the parametric data assumptions. Data from different participants were independent of one another and all intended analyses were conducted between groups. Variable response sets, then, were interval and independent, as necessary to meet the assumptions of parametric data.

One variable of interest did not meet these standards and had to be recoded as a nominal variable and analyzed separately from the rest of the data. The Set-Loss score was recoded as a nominal value to enable its analysis. As computed with the computer scoring software, the Set-Loss score from the D-KEFS was represented as a cumulative percentile rank (CPR) indicating the percentage of the normative sample that obtained raw scores equal to or worse than the scores generated by the participant (Delis, 2001b). In this study, this error measure yielded only three unique CPR: <1, 1, and 100. These scores could not be used in the planned data analyses due to this limited variability. Additionally, the score <1 could not be entered into the software due to its imprecision. These factors required that the Set-Loss be recoded to enable its analysis. The initial CPR of 100 was generated by the scoring software when zero Set-Loss errors were committed by the participant. The CPR of 1 was generated when 1 Set-Loss error was committed by the participant. The CPR of <1 was generated when more than 1 Set-Loss error was committed by the participant. This is a known output limitation for some of the error and ratio measures in the D-KEFS scoring that results from a limited range of scores in nonclinical populations. This restricted range necessitates that some of the raw process scores be converted to CPR at each age level (Strauss et al., 2006). Given the limited variability and specificity of the scores obtained on the variable Set-Loss in this study, and the inability to obtain alternate score formats due to limitations in the normative sample, the Set-Loss score was recoded for the purposes of this study. The score was recoded as a dichotomous variable denoting the presence or absence of Set-Loss errors in the participant response set as presented in Table 9 below.

Table 9

Sorting Test Set-Loss Sorts Recorded From a Cumulative Percentile Rank to Categorical Variable

CPR	Frequency	Percent of Set-Loss	New Code
<1, 1	3	2.65	1
100	108	95.58	0
Total	111	98.23	Categorical

Note. CPR = cumulative percentile rank; Set-Loss = D-KEFS failure to maintain set

Normal Distribution

Evaluation of the assumption of normally distributed variable data by condition was conducted with a variety of approaches. Data were inspected along variables by condition graphically via histogram with normal distribution plots, q-q plots, and p-p plots. Several variables appeared normally distributed and others appeared to be impacted by outliers. Box plots of the nine variables by condition were analyzed for outliers. Examination of normality via skew and kurtosis, the Kolmogorov-Smirnov test, and the Shapiro-Wilk test was conducted, as detailed below.

SPSS was used to calculate the skew, kurtosis, and standard error of each variable by condition (Table 10). These were then used to compute z-scores for the skew and kurtosis. The z-scores were compared to values expected by chance to evaluate the significance of the sample distribution skew and kurtosis from normal distribution. In Minimal Instruction condition, the variables Repeated Sorts, Color-Word, Categories Completed, Perseverative Responses. Failure to Maintain Set, and Set-Loss demonstrated statistically significant non-normality. In Feedback condition, the variables Repeated Sorts, Categories Completed, Perseverative Responses, Failure to Maintain Set and Set-Loss demonstrated statistically significant non-normality. In the Control condition Color-Word, Categories Completed, Perseverative Responses, and Failure to Maintain Set demonstrated statistically significant non-normality. As deviations above chance were found along several variables in all conditions, further examination of non-normality was pursued.

Table 10

Skew and Kurtosis of Variables in Each Condition

		Minimal Instruction condition			Feedback condition			Control		
			SE	Z-score		SE	Z-score		SE	Z-score
CCS	Skew	-.19	.32	-0.60	-.22	.32	-0.67	.06	.32	0.18
	Kurt	-.31	.63	-0.50	-.34	.63	-0.54	-.83	.63	-1.31
Free	Skew	-.25	.32	-0.78	-.51	.32	-1.59	-.21	.32	-0.64
	Kurt	-.55	.63	-0.88	1.02	.63	1.61	-.60	.63	-0.95
Rec	Skew	-.08	.32	-0.26	-.51	.33	-1.54	-.03	.31	-0.09
	Kurt	.07	.63	0.11	-.04	.66	-0.06	-.51	.62	-0.83
Rep	Skew	-1.47	.32	-4.50***	-1.40	.33	-4.30***			
	Kurt	1.64	.63	2.62***	1.12	.64	1.75			
LF	Skew	-.00	.32	-0.01	.24	.32	0.73	.12	.31	0.37
	Kurt	-.80	.63	-1.28	-.31	.63	-0.49	-.55	.62	-0.89
CWI	Skew	-1.17	.32	-3.65***	-.48	.32	-1.48	-.79	.32	-2.51*
	Kurt	1.95	.63	3.10*	-.38	.63	-0.60	.50	.62	0.81
NAART	Skew	.01	.32	0.03	-.09	.32	-0.29			
	Kurt	-.09	.63	-0.14	-.14	.62	-0.22			
Cat	Skew	-.87	.33	-2.65**	-.09	.34	-0.27	-1.26	.33	-3.88***
	Kurt	.28	.64	0.43	-1.41	.66	-2.13*	.56	.64	0.87
PR	Skew	1.69	.33	5.17***	1.36	.34	4.05***	1.70	.33	5.23***
	Kurt	2.14	.64	3.32**	2.09	.66	3.15**	3.06	.64	4.79***
WFMS	Skew	.83	.33	2.54*	1.20	.34	3.55***	1.93	.33	5.95***
	Kurt	-.27	.64	-0.42	.52	.66	0.78	3.03	.64	4.74***
Set-Loss	Skew	5.14	.32	16.12***	7.42	.32	23.03***			
	Kurt	25.35	.63	40.37***	55.00	.63	86.75***			

Note. M = mean; CCS = confirmed correct sorts; Free = free sorting description score; Rec = sort recognition description score; Rep= repeated sorts; LF = letter fluency total correct index; CWI = color-word interference inhibition condition; Set-Loss = set-loss sorts; Cat= categories completed; PR = perseverative responses; WFMS = WCST failure to maintain set; SE = standard error, NAART = North American adult reading test; kurt = kurtosis
*significant at p<.05
**significant at p<.01
***significant at p<.001

The Kolmogorov-Smirnov test (K-S test) was next used to evaluate the deviations of the variable distributions from normal distributions (Table 11). This test compares the scores in the sample to a normally distributed set of scores with the same mean and standard deviation. In this way, the test provides information useful in determining how well the study data meet the normality assumption of parametric data. This test has the power to detect deviations from the hypothesized distribution but it is also very conservative due to location and shape parameters of the hypothesized distribution being estimated from the data (Lilliefors, 1967). The Kolmogorov-Smirnov test was conducted with Lilliefors significance correction. This correction was developed using Monte Carlo simulations in order to adjust the critical values determining significance for the K-S test, making it less conservative (Lilliefors, 1967). The K-S test with Lilliefors significance correction yielded significant deviations from normality for every variable, excepting the North American Adult Reading Test (NAART) on at least one condition. Similarly, the Shapiro-Wilk test of normality (S-W test) was conducted. This test is also used to compare the sample distribution to a normal distribution but differs from the K-S test in that it was specifically designed to test for normality whereas the K-S test can be used to compare sample distributions to other distributions than normal (Shapiro & Wilk, 1965). Results of the S-W test indicated that Description, Recognition, Repeated Sorts, Color-Word, Categories Completed, Perseverative Responses, Failure to Maintain Set and Set-Loss were significantly deviant from normally distributed data. All condition variable data sets were comprised of greater than 50 scores, however, and SPSS recommends that both the K-S and S-W tests be used only for samples of less than 50. Normality tests such as these have been referred to as supplementary to graphical analysis of normality (Elliott & Woodward, 2007).

Table 11

Normality Tests of Variables in Each Condition

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Minimal Instruction condition	Feedback condition	Control condition	Minimal Instruction condition	Feedback condition	Control condition
CCS	Statistic	.13*	.12	.13*	.97	.97	.96
	df	56	55	55	56	55	55
	Sig.	.017	.053	.028	.159	.162	.062
Free	Statistic	.13*	.09	.14*	.97	.96*	.97
	df	56	55	55	56	55	55
	Sig.	.020	.200[b]	.012	.135	.047	.125
Rec	Statistic	.11	.16**	.14**	.98	.96	.96*
	df	56	51	58	56	51	58
	Sig.	.095	.003	.007	.421	.072	.047
Rep	Statistic	.22***	.29***		.79***	.74***	
	df	56	54		56	54	
	Sig.	.000	.000		.000	.000	
LF	Statistic	.12*	.10	.09	.96	.97	.98
	df	56	55	58	56	55	58
	Sig.	.043	.200[b]	.200[b]	.089	.279	.470
CWI	Statistic	.18***	.13*	.13*	.91***	.96*	.92**
	df	56	55	57	56	55	57
	Sig.	.000	.025	.012	.000	.045	.001
NAART	Statistic	.07	.07		.99	.99	
	df	56	57		56	57	
	Sig.	.200[b]	.200[b]		.808	.781	
Cat	Statistic	.17**	.16**	.26***	.92**	.89***	.77***
	df	53	50	54	53	50	54
	Sig.	.001	.002	.000	.001	.000	.000
PR	Statistic	.19***	.14*	.19***	.77***	.88***	.81***
	df	53	50	54	53	50	54
	Sig.	.000	.020	.000	.000	.000	.000
WFMS	Statistic	.362***	.388***	.468***	.707***	.684***	.528***
	df	53	50	54	53	50	54
	Sig.	.000	.000	.000	.000	.000	.000
Set-Loss	Statistic	.540***	.535***		.184***	.117***	
	df	56	55		56	55	
	Sig.	.000	.000		.000	.000	

Note. CCS = confirmed correct sorts; Free = free sorting description score; Rec = sort recognition description score; Repeated Sorts= repeated sorts, LF = letter fluency total correct index; CWI = color-word interference inhibition condition; Set-Loss = set-loss sorts; Cat= categories completed; PR = perseverative responses; WFMS = WCST failure to maintain set; NAART = North American adult reading test.
*significant at p<.05
**significant at p<.01
***significant at p<.001
[a] Lilliefors Significance Correction
[b] This is a lower bound of the true significance

Given that these investigations indicated non-normality in some of the sample variable distributions by condition, several corrections were explored as a means to reduce the impact of outliers and non-normality. First, outliers were removed and distributions were re-evaluated. This did not result in a greater number of sample distributions being normally distributed. Next, the data were transformed. Each score was transformed to a z-score using the sample mean and sample standard deviation by variable by condition. Normality of sample distributions was also not improved with this approach. It was not feasible to transform the data with other popular methods (e.g., change the outlier scores to the next highest score plus one) due to clusters of outlier responding.

Use of nonparametric analyses was considered, however, experts have found that for larger samples in which responses are greater than 30-40, the sampling distribution tends to be normally distributed regardless of the shape of the data (Elliott & Woodward, 2007; Field, 2013; Ghasemi & Zahediasl, 2012). Additional statistical experts have indicated that the violation of the normality assumption is not expected to cause major problems when using parametric tests on non-normally distributed data with total responses greater than 100(Altman & Bland, 1995; Pallant, 2013). Given that the data from this study met both of those sample size standards, analyses with Pearson correlations were deemed appropriate for this study.

Excluded Groups

Participants were recruited with few initial exclusions. Factors that could confound the findings of the study were reviewed following data collection. Factors evaluated were previous head injury, English as second language, anomalous color vision, low reading level, previous exposure to a sorting test, and IQ. Research regarding the impact of these factors on executive function or the tests used in this study as well as analysis for the presence of such an impact is detailed below.

Traumatic Brain Injury (TBI)

Responses from participants who reported traumatic brain injury (TBI) were evaluated for possible exclusion from data analyses. Research showing that the right dorsolateral frontal-subcortical circuit is critical to performance on the WCST (Lombardi et al., 1999; Stuss et al., 2000) and that Perseverative Responses can discriminate between focal and diffuse cerebral regions (Robinson, Heaton, Lehman, & Stilson, 1980) and between closed head injuries, mixed CNS pathology, malingerers, and controls (Bernard, McGrath, & Houston, 1996), suggests that such an exclusionary factor regarding variables of interest on the WCST must be considered. Similarly, research demonstrating that the Sorting Test can differentiate between post-injury and controls for severe TBI supported the evaluation of previous head injury as a possible exclusionary factor regarding variables of interest on the D-KEFS (Heled, Hoofien, Margalit, Natovich, & Agranov, 2012). Participants were provided a forced-choice self-report for such history on the demographics questionnaire. Of the 171 study participants, 32 (18.71%) reported a TBI. To evaluate if the data were significantly different with the 32 participants with a TBI included, a second sample with their responses excluded listwise was created. These two samples were then compared using an independent t-test. The purpose of this analysis was to identify if the 32 participants with a TBI should be excluded from the study analyses due to statistically significant deviation from participants without history of TBI on executive function performance. Independent t-tests were conducted for each variable. As t-tests require assumption of equal variances, Levene's test for equality of variances was conducted for the DKEFS primary and secondary variables and the WCST variables.

Results of Levene's test indicated non-significant differences for each pair at each variable. With equal variances assumed, t-tests were conducted for each variable. Significance tests (2-tailed) indicated that there was no significant difference between performances on most variables of interest for participants with TBI as compared to participants without TBI. For three variables, Confirmed Correct Sorts, Letter Fluency, and Categories Completed, the t-tests were significant, as reflected in the table below (Table 12).

Table 12

Significant Performance Differences for Subjects With Traumatic Brain Injury

Variable	t	df	Sig. (2-tailed)	Mean difference	Direction of Performance relevant to previous head injury
CCS	-1.98	16	.05	-.99	Positive
LF	-2.24	17	.03	-1.29	Positive
Cat	2.18	16	.03	.49	Negative

Note. The variables Repeated Sorts and Failure to Maintain Set were not included in the control data. Thus, the comparison for these variables was conducted with only the experimental condition participants, n=110 and n=111, respectively. CCS = Confirmed Correct Sorts; LF = Letter Fluency; Cat = Categories Completed.

Accounting for the direction of the performance, on two of the three significant performance differences, participants with a TBI performed significantly better than those without injury history on variables from both the D-KEFS and WCST. On Categories Completed from the WCST, participants with TBI performed worse than those without such history. Overall, those with TBI performed better than those without on six variables where those without TBI performed better on five variables. These findings yield no discernable pattern in the direction of the performance scores to warrant the exclusion of those participants with a previous TBI given the results for both significant and non-significant differences between groups. Findings regarding impact of previous TBI on tests of executive function here are inconsistent overall, a result that is not consistent with the diminished performance on D-KEFS or WCST executive function measures predicted in the literature.

English as Second Language (ESL)

Speaking English as a second language has been shown to significantly impact performance on neuropsychological assessment (Boone, Victor, Wen, Razani, & Pontón, 2007). These findings have been reproduced in research assessing performance on the WCST specifically (Coffey, Marmol, Schock, & Adams, 2005) . This study's ESL participants were East Asian (*n*=7, 100% of ESL). An emerging literature supporting cultural neuroscience has indicated that East Asians process semantic versus functional relationships differently than Americans (Leany, Benuto, & Thaler, 2014). Further, research assessing the use of Western neuropsychological measures or linguistic or cultural adaptations of such in comparison to indigenously developed tests appears to suggest an interaction between neurobiological processes and culture(Chan, Shum, & Cheung, 2003; Leany et al., 2014). Additional to these findings regarding the impact of culture and acculturation on performance on measures of executive functioning, the Sorting Test relies on verbal instruction to provide task clarity and direction, an approach that could place ESL participants at further disadvantage. Moreover, the impact of these instructions on construct validity is at the center of this current study, with the experimental conditions specifically manipulating this component. It is for these reasons that participant data were coded for reported or demonstrated ESL. Of the 113 participants in the experimental conditions, 7 (6.2%) were identified as ESL. Score means were compared using independent t-tests. Levene's test for equality of variances was nonsignificant for all variables, excepting Set-Loss (p=.000). Two-tailed significance testing was conducted at the p<.05 level--accounting for variables of assumed equal mean variances and those without assumed equal mean variances--and three variables were found to differ significantly along ESL, as reflected below (Table 13).

Table 13

Significant Performance Differences for Subjects With English as a Second Language

Variable	t	df	Sig. (2-tailed)	Mean difference	Direction of Performance relevant to ESL
LF	3.51	109	.001	4.05	Negative
CWI	2.80	109	.006	3.15	Negative
NAART	5.69	111	.000	15.91	Negative

Note. ESL = English as second language; LF = letter fluency; CWI = Color-Word Interference, Inhibition condition; NAART = North American Adult Reading Test.

Performance on the NAART is dependent on English language production. It is, therefore, predictable that this score would yield lower performance scores for participants for whom English was a secondary language. Additionally, significant task demand, predicated on language production under time constraint, is present in both Letter Fluency and Color-Word. It is, therefore, predictable that these scores would significantly differ along ESL. Overall performance across variables demonstrated a pattern of poorer performance scores for ESL participants (4 variables with superior performance) as compared to non-ESL participants (7 variables with superior performance). Given that the discrepancy in group means was significant at the p<.01 level and that there appeared to be a pattern of poorer performance across seven variables of interest to this study, the seven ESL participant scores were excluded listwise on all variables when statistical tests were computed.

Anomalous Color Vision and Low Reading Level (Misc)

Participants were also analyzed for possible exclusion based on anomalous color vision (*n*=3) and low reading level (*n*=1). Participant responses were coded categorically on variable Misc, reflecting presence or absence of Misc exclusion factors. Next, performance variables were assessed for equal mean variances with Levene's test. Equal mean variances were confirmed, enabling further examination of group mean differences with an independent t-test. Although color vision and reading are critical to performance on all popular variations of the Stroop test, as well as on many of the other performance scores included in this study, no significant differences were found on factor Misc across all variables. Additionally, no pattern of superior performance was discerned (Misc held advantage on 6 of 11 variables). These results supported the indiscriminate inclusion of Misc scores for analyses in this study.

Previous Exposure to a Sorting Test (Hx Tested)

Eight of 111 participants endorsed previous exposure to sorting tasks. Research has indicated that exposure to the WCST results in practice effects on subsequent WCST performance at 12 months (Basso, Bornstein, & Lang, 1999), so those participants in group Hx Tested were compared to those without group membership across each variable with t-tests. Levene's test indicated equal mean variances. T-tests comparing the two group means were conducted and were not significant for any study variable. Performance of those with previous sorting exposure as compared to those without yielded a 5of 11 f variables indicating superior performance. As a result of the absence of discernable performance directionality patterns, all eight of these participants were included in study analyses.

IQ (North American Adult Reading Test (NAART))

Next, an inspection of the data set was conducted to assess if any variables correlated significantly with IQ (Table 14). Previous research has reported that full scale IQ significantly relates to executive function measures, and to the WCST, Stroop, Color-Word Interference Test Condition 3 Inhibition (Color-Word), and COWAT specifically (Arffa, 2007). In this study, NAART was significantly correlated with Rec, r=.20 p<.05, Cat, r=.27 p<.01, Failure to Maintain Set, r=-.28 p<.01, and CWI, r=.30 p<.01. Only CWI was significantly correlated with NAART after correcting for family wise error at the Bonferroni-adjusted significance level of p<.005, however, partial correlations controlling for NAART were used in this study given the literature suggesting significant correlation between IQ and executive function and the possibility that the Bonferroni correction was too cautious and resulted in a Type II error in this analysis.

Table 14

Correlation of Participant IQ Scores (NAART) With Performance Variables

	Pearson Correlation	Sig. (2-tailed)	N
CCS	.15	.116	105
Free	.19	.057	105
Rec	.20*	.041	101
LF	.04	.656	105
Cat	.27**	.008	96
PR	-.16	.123	96
WFM	-.28**	.005	96
Set-Loss	.04	.662	105
Rep	-.03	.761	104
CWI	.30**[a]	.002	105

Note. CCS = confirmed correct sorts; Free = free sorting description score; Rec = sort recognition description score; Rep= repeated sorts, LF = letter fluency total correct index; CWI = color-word interference inhibition condition; Set-Loss = set-loss sorts; Cat= categories completed; PR = perseverative responses; WFMS = WCST failure to maintain set.
*significant at p<.05
**significant at p<.01
[a]significant at p<.005, Bonferroni-adjusted significance level

CHAPTER 4

RESULTS

Statistical Analyses of Hypothesis 1

Convergent Validity

To evaluate the hypothesis that theoretically derived administration of the Sorting Test would increase the correlation between Sorting Test scores and primary measures from the WCST, convergent validity of the Sorting Test with Letter Fluency Total Correct (Letter Fluency) and the WCST was calculated. Partial correlations were conducted between the primary D-KEFS variables and the primary WCST variables, as depicted below (Tables 15, 16, and 17). For the Minimal Instruction condition, Confirmed Correct Sorts was significantly correlated with Categories Completed, $r= -.312$, $p<.05$ and Description was significantly correlated with Perseverative Responses , $r=.317$, $p<.05$. Recognition was not significantly correlated with Letter Fluency, Categories Completed, or Perseverative Responses. These significance findings account for the confounding effect of IQ, but required multiple comparisons be conducted with the data from Minimal Instruction condition. Multiple comparisons conducted with one sample of data had the potential to increase the family wise error rate. This could have resulted in false positive significance findings. To address this, a Bonferroni-adjusted significance level was calculated ($.05/11=p<.005$). When convergent validity of the Minimal Instruction condition was assessed for significance at this level, none of the primary D-KEFS scores were significantly correlated with Letter Fluency, Categories Completed or Perseverative Responses, all $ps>.005$.

For Feedback condition, Perseverative Responses was significantly correlated with Confirmed Correct Sorts, $r = .32$, $p < .05$ but Description , Recognition and Letter Fluency were not significantly correlated with Categories Completed, Perseverative Responses or each other. When the convergent validity of Feedback condition was assessed with the Bonferroni-adjusted significance level, none of the primary D-KEFS scores were significantly correlated with the primary WCST variables, all $ps > .005$.

For the Control condition, neither Confirmed Correct Sorts nor Description were significantly correlated with the primary measures of the WCST. Recognition was significantly correlated with Categories Repeated $r = .51$, and Perseverative Responses, $r = -.49$ (all $ps < .001$). Letter Fluency was significantly correlated with Description, $r = .28$, and PR, $r = -.31$ (all $ps < .05$). When convergent validity of the Control condition was assessed at the Bonferroni-adjusted significance level, Recognition was significantly correlated with Categories Completed and Perseverative Responses at $p < .005$. In summation, the Sorting Test was not found to significantly correlate with other measures of executive function under either experimental condition after controlling for IQ. Under the Control conditions, Recognition was found to significantly correlate with Perseverative Responses after controlling for IQ, and with Categories Completed after controlling for IQ alone, as well as after controlling for IQ and age.

Table 15

Examination of Convergent Validity Using Correlation of Variables Under the Minimal Instruction Condition After Controlling for Participant IQ

				D-KEFS Variables		
WCST Variables			CCS	Free	Rec	LF
Cat	*r*	-.31*	-.28[3]	-.25	-.15	-.12
	Sig.	.029	.054	.082	.309	.401
PR	*r*	.22		.32*	.12	-.08
	Sig.	.139		.027	.404	.600
LF[b]	*r*	.02		.04	-.17	
	Sig.	.895		.763	.232	

Note. Cat = categories completed; PR = perseverative responses; CCS = confirmed correct sorts; Free = free sorting description score; Rec = sort recognition description score; LF = letter fluency.
*significant at p<.05
[b]D-KEFS variable
[3]new values generated by controlling for age with 1 case excluded listwise

Table 16

Examination of Convergent Validity Using Correlation of Variables Under the Feedback Condition After Controlling for Participant IQ

			D-KEFS Variables		
WCST Variables		CCS	Free	Rec	LF
Cat	r	-.19	-.14	-.04	.11
	Sig.	.201	.372	.786	.476
PR	r	.32*	.27	.12	.01
	Sig.	.031	.075	.458	.945
LF[b]	r	.19	.18	.04	
	Sig.	.173	.200	.811	

Note. Cat = categories completed; PR = perseverative responding; CCS = confirmed correct sorts; Free = free sorting description score; Rec = sort recognition description score; LF = letter fluency.
*significant at p<.05
[b]D-KEFS variable

Table 17

Examination of Convergent Validity Using Correlation of Variables Under the Control Condition

				D-KEFS Variables		
WCST Variables		CCS	Free		Rec	LF
Cat	*r*	.11	.15	.51***[a]	.49***[a 23]	.21
	Sig.	.445	.288	.000	.000	.129
PR	*r*	.10	-.01	-.49***[a]		-.31*
	Sig.	.506	.970	.000		.022
LF[b]	*r*	.14	.28*	.16		
	Sig.	.321	.037	.224		

Note .Cat = categories completed; PR = perseverative responding; CCS = confirmed correct sorts; free = free sorting description score; Rec = sort recognition description score LF = letter fluency.
*significant at p<.05
***significant at p<.001
[a]significant at p<.005, Bonferroni-adjusted significance level
[b]D-KEFS variable
[2]new values generated by controlling for age
[3]new values generated by controlling for age with 1 case excluded listwise

Change in Convergent Validity of the Sorting Test

Next, significance testing was conducted comparing experimental condition correlations to Control condition correlations. This was done by converting correlation coefficients to z-scores and then calculating the difference between the z_r from experimental condition versus control. These difference scores were transformed into z-scores and two-tailed test of significance of the $Z_{Difference}$s was generated, as detailed in the table below. All significant differences found in convergent validity reflected weaker correlations. In the Minimal Instruction condition, the correlation of Categories Completed with Confirmed Correct Sorts, $z=-2.07$ $p<.05$, was significantly different than in the Control condition. The difference between the Minimal Instruction condition and controls for Recognition by Categories Completed $z=-3.45$, and by Perseverative Responses, $z=3.18$, was significant at the $p<.001$ level. When evaluated with a Bonferroni-adjusted significance level, Recognition by Categories Completed and by Perseverative Responses were significantly different in Minimal Instruction condition as compared to the Control condition, $p<.005$.

In Feedback condition, a two-tailed test of significance of the $Z_{Difference}$s indicated significant differences in Recognition by Categories Completed, $z=-2.78$, and by Perseverative Responses, $z=2.99$, at the $p<.01$ level. When evaluated with a Bonferroni-adjusted significance level, Recognition by Perseverative Responses was significantly different in the Feedback condition as compared to the Control condition, $p<.005$. In summation, after controlling for family wise error, this study found the correlation of Recognition with Categories Completed and with Perseverative Responses to be significantly different in the Minimal Instruction condition as compared to the Control condition. Results also yield a significant difference in the Feedback condition as compared to controls for the correlation of Recognition with Perseverative Responses.

Table 18

Significance Testing of the Change in Convergent Validity Under the Minimal Instruction Condition

Correlation	Raw Difference	z	Sig (2-tail)
CCSxCat	-0.42	-2.07	.038*
CCSxPR	0.12	0.59	.555
CCSxLF	-0.12	-0.59	.558
FreexCat	-0.40	-1.96	.050
FreexPR	0.32	1.60	.110
FreexLF	-0.24	-1.23	.220
RecxCat	-0.66	-3.45	.001***[a]
RecxPR	0.61	3.18	.001***[a]
RecxLF	-0.33	-1.68	.093
LFxCat	-0.33	-1.63	.103
LFxPR	0.23	1.19	.235

Note. Sig = significance; CCS = confirmed correct sorts; Free = free sorting description score = sort recognition description score LF = letter fluency total correct index; CWI = color-word interference inhibition condition; Set-Loss = D-KEFS failure to maintain set; Cat = categories completed; PR = perseverative responses; WFMS = WCST failure to maintain set.
*significant at p<.05
***significant at p<.001
[a]significant at p<.005, Bonferroni-adjusted significance level

Table 19

Significance Testing of the Change in Convergent Validity Under the Feedback Condition

Correlation	Raw Difference	z	Sig (2-tail)
CCSxCat	-0.30	-1.43	.153
CCSxPR	0.23	1.11	.267
CCSxLF	0.06	0.30	.768
FreexCat	-0.28	-1.36	.175
FreexPR	0.27	1.31	.190
FreexLF	-0.10	-0.52	.601
RecxCat	-0.55	-2.78	.005**
RecxPR	0.61	2.99	.003**[a]
RecxLF	-0.13	-0.62	.534
LFxCat	-0.10	-0.49	.627
LFxPR	0.32	1.58	.115

Note. CCS= confirmed correct sorts; Free = free sorting description score; Rec = sort recognition description score Rep = repeated sorts, LF = letter fluency total correct index; CWI = color-word interference inhibition condition; Set-Loss = D-KEFS failure to maintain set; Cat = categories completed; PR = perseverative responses; WFMS = WCST failure to maintain set.
** significant at p<.01
[a]significant at p<.005, Bonferroni-adjusted significance level

Statistical Analysis of Hypothesis 2

Divergent Validity

Recent functional brain imaging studies indicate that retrieval, maintenance, and selection of semantic information, such as is needed for object concepts on Color-Word, involve distributed networks of cortical regions additional to the left lateral PFC, such as the anterior temporal cortex (Martin & Chao, 2001). Thus, Color-Word can serve to contrast with the WCST's undisputed cortical region engagement of the DLPFC, as well as with the Sorting Test, which should similarly engage the DLPFC if successfully evaluating the same theoretical construct of executive function.

To evaluate the hypothesis that theoretically derived administration of the Sorting Test would not significantly impact the correlation between Color-Word scores and primary measures from the D-KEFS and the WCST, divergent validity of the Sorting Test, Letter Fluency, and WCST from Color-Word was calculated. For Minimal Instruction condition, Color-Word was significantly correlated with Recognition, $r=-.33$, $p<.05$, but not with the other primary measures of the D-KEFS or the WCST. For Feedback condition, Color-Word was not significantly correlated with the primary measures of the D-KEFS or the WCST. For the Control condition, Color-Word was significantly correlated with Letter Fluency $r=.36$, Cat, $r=.37$, and Perseverative Responses, $r=-.38$ (all ps $<.01$). When possible increase in Type I error was accounted for with a Bonferroni-adjusted significance level ($.05/6=p<.008$), no significant correlations were found between Color-Word and the other D-KEFS and WCST primary variables in the Minimal Instruction condition or the Feedback condition, $p>.008$. For the Control condition, Color-Word was significantly correlated with

Letter Fluency, r= .36, Cat, r=.37, and Perseverative Responses, r=-.38 (all ps <.008). In summation, after controlling for family wise error, Color-Word was not significantly correlated with the Sorting Test or the WCST under either experimental condition. Under the Control conditions, Color-Word was significantly correlated with Letter Fluency, Categories Completed, and Perseverative Responses.

Table 20

Examination of Divergent Validity Using Correlation of Variables to Color-Word Interference

		CWI Cond. 1	CWI Cond. 2	CWI control
CCS	*r*	-.17	.22	-.08
	Sig.	.230	.117	.589
Free	*r*	-.22	.22	-.02
	Sig.	.124	.118	.911
Rec	*r*	-.33*	.22	.04
	Sig.	.016	.145	.783
LF	*r*	-.00	-.01	.36**[a]
	Sig.	.997	.934	.007
Cat	*r*	.05	.03	.37**[a]
	Sig.	.735	.833	.006
PR	*r*	-.16	.01	-.38**[a]
	Sig.	.276	.975	.004

Note. Cond. = condition; CWI = Color Word Interference; CCS = confirmed correct sorts; Free = free sorting description score Rec = sort recognition description score; LF = letter fluency; Cat = categories completed; PR = perseverative responding.
*significant at p<.05
**significant at p<.01
[a]significant at p<.008, Bonferroni-adjusted significance level

Change in Divergent Validity of Color-Word to the Sorting Test and WCST

Next, significance testing was conducted comparing experimental condition correlations with Color-Word to Control condition correlations with Color-Word. This was done by the same method described above. Results are depicted in tables 21 and 22 below. A two-tailed test of significance of the $Z_{Differences}$ indicated no significant difference between participants in the experimental conditions as compared to control participants. Divergence of the Sorting Test and WCST variables from Color-Word was not significantly different in either experimental condition when compared to controls.

Table 21

Significance Testing of the Change in Divergent Validity Under the Minimal Instruction Condition

Correlation	Raw Difference	z	Sig (2-tail)
CCSxCWI	-0.09	-0.47	.637
FreexCWI	-0.20	-1.01	.314
RecxCWI	-0.37	-1.93	.054
CWIxCat	-0.32	-1.66	.098
CWIxPR	0.22	1.18	.237

Note. CCS= confirmed correct sorts; Free = free sorting description score; Rec = sort recognition description score CWI = color-word interference inhibition condition; Cat = categories completed; PR = perseverative responses.

Table 22

Significance Testing of the Change in Divergent Validity Under the Feedback Condition

Correlation	Raw Difference	z	Sig (2-tail)
CCSxCWI	0.30	1.49	.138
FreexCWI	0.24	1.18	.237
RecxCWI	0.18	0.89	.375
CWIxCat	-0.34	-1.70	.089
CWIxPR	-0.30	-1.43	.153

Note. CCS = confirmed correct sorts; Free = free sorting description score; Rec = sort recognition description score; CWI = color-word interference inhibition condition; Cat = categories completed; PR = perseverative responses.

Statistical Analyses of Hypothesis 3

To evaluate the relative value of the optional measures of the Sorting Test, Pearson correlation coefficients were calculated between the variable pairs of interest (Table 23). Results indicated a significant relationship between Repeated Sorts and Perseverative Responses for participants in the Feedback condition, $r=-.32$, $p<.05$. but not in the Minimal Instruction Condition. When assessed at the Bonferroni-adjusted significance level ($.05/2=p<.025$), however, no significant comparisons were obtained, all ps>.025. Statistical analysis of Set-Loss was not possible due to its low frequency of occurrence (Table 24). In the Feedback condition, no scores were generated for Set-Loss. Only two scores were generated in the Minimal Instruction Condition, . This variable, re-coded categorical due to score output, was constant across the Feedback condition, with no participants receiving a 1. Set-Loss was coded as 1 for two participants in the Minimal Instruction Condition, a rate without meaningful predictive power. Combining the experimental conditions yielded no change in the utility of the response variable in this data set. Optional scores could not be obtained for the Control condition because the data did not contain information necessary for computation of these scores.

Lastly, the correlation coefficients from the optional scores were further evaluated for significant differences between experimental conditions. The correlation coefficients were converted to z_r and the differences between these were transformed into z-scores. A two-tailed test of significance of the $Z_{Difference}$s indicated no significant difference between participants in the Minimal Instruction condition as compared to the Feedback Condition on Repeated Sorts by Perseverative Responses (Table 25). In summation, exploratory analyses of optional scores on the Sorting Test yielded no significant correlations with other measures of executive function and these results were not significantly different between experimental conditions.

Table 23

Correlation of Repeated Sorts With Perseverative Responses

		Minimal Instruction condition	Feedback condition
Variable		PR	PR
Rep	*r*	-.17	-.32*
	Sig.	.249	.031
	N	50	45

Note. Rep = repeated sorts; PR = Perseverative Responses.
*significant at p<.05

Table 24

Correlation of Set-Loss Sorts With Failure to Maintain Set

Variable	Condition 1	Condition 2	Total
Set-Loss	2	0	2

Table 25

Significance Testing of the Change in Correlation of Repeated Sorts With Perseverative Responses

Correlation	Raw Difference	z	Sig (2-tail)
RepxPR	.16	.79	.430

Note. Rep = repeated sorts; PR = perseverative responses.

Results of Analyses at a Less Conservative Significance Threshold

Results were first analyzed with Bonferroni-adjusted significance levels to decrease the probability of a Type I error; however, this may have resulted in an overcorrection such that Type II error was unnecessarily inflated. Results were next analyzed at ~~a~~ the less conservative significance threshold of .01.

There were no significant correlations between the Sorting Test and the WCST in either experimental condition. Recognition was significantly correlated with Categories Completed and Perseverative Responses, *rs*=.49, ps<.01, in the Control condition. Under Minimal Instruction condition and Feedback condition, the difference between the correlations was significantly different (lower) than in the Control condition, Recognition by Categories Completed, z=-3.45, and by Perseverative Responses, z=3.18, all ps<.01.

Divergent validity of the Sorting Test from Color-Word was supported with nonsignificant correlations of test variables with Color-Word. Significant correlations were observed in the Control condition between Color-Word with Letter Fluency, r=.36, Categories Completed, r=.37, and Perseverative Responses, r=.38, all ps<.01. No significant difference was found in the correlation of variables under experimental conditions as compared to the Control condition.

Optional score Repeated Sorts was not found to significantly correlate with Perseverative Responses under any study condition. The difference between these results was not significant when experimental conditions were compared to the Control condition. Correlation of Set-Loss Sorts with Failure to Maintain Set was not possible due to low occurrence in the study sample.

CHAPTER 5

DISCUSSION

This study examined the Sorting Test and had three purposes. It first strove to examine the construct validity of the Sorting Test. This was achieved by investigating convergent and divergent validity with other executive function measures. It next strove to evaluate the impact of the atheoretical design of the Sorting Test on convergent and divergent validity with other executive function measures. This was achieved by designing and administering two theoretically derived alternative administration protocols. The resulting performances for each of the three formats were then compared on their convergent and divergent properties. Lastly, optional scores of the Sorting Test were explored for convergence with other executive function measures.

The Sorting Test was not significantly correlated with other measures of executive function under theoretically derived non-standard administration. Under the Control condition, only Recognition was significantly correlated with other measures after controlling for IQ. Similarly, comparison of the conditions yielded weaker and statistically significant differences in the correlation of the Sorting Test with other measures of executive function only for Recognition by Categories Completed and by Perseverative Responses. Neither the Sorting Test nor the WCST were significantly correlated with measures not assessing the DLPFC under non-standard administration but WCST and Letter Fluency Total Correct (Letter Fluency) were under the Control condition. Comparison of the conditions was non-significant for change in variable correlations with the non-DLPFC measure. Optional scores from the Sorting Test were not significantly correlated with other measures of executive function, regardless of which non-standard administration was provided. Results of this study derived from Bonferroni-adjusted significance levels were upheld without deviation under a less conservative $p<.01$ threshold.

Executive functions have been defined as paramount to a healthy lifestyle, surpassing cognitive functioning in their regard as quintessential higher-level cortical processing (Eslinger et al., 1995). The impact of executive functioning on the life of a patient is important to address with modern neuropsychological assessment (Manchester, Priestley, & Jackson, 2004). Results of neuropsychological assessment have a critical role in evaluating the patient's strengths and weaknesses and in developing an accurate prognosis with treatment planning recommendations (Gioia et al., 2002).

Neuropsychological assessment of a clinical population requires powerful and precise measurement tools. Evaluation of measurement tools intended for neuropsychological assessment is often conducted via analysis of construct validity (Bagozzi, Yi, & Phillips, 1991). Construct validity in this sense requires that the successful measurement tool demonstrate convergent validity with other measures assessing the same construct, and divergent validity with measures that assess adjacent but non-overlapping cognitive constructs (Carmines & Zeller, 1979; O'Leary-Kelly & Vokurka, 1998).

When a new measurement tool becomes available, it is validated through research. In this way, new measures are comprehensively evaluated and clinical professionals make a decision as to whether such measures shall be incorporated into their common battery of tests. By this standard, the WCST has been determined by the field at-large to evaluate executive functioning, specifically as processed in the dorsolateral prefrontal cortex (Diamond, 1991; Sumitani et al., 2006). Developing out of a long tradition of sorting tests as measurement tools of higher-order cognitive functioning, previous research has demonstrated that the WCST activates the DLPFC as observed with multichannel near-infrared spectroscopy (Sumitani et al., 2006), MRI and rCBF (Berman et al., 1992; Weinberger et al., 1992), and positron emission tomography (Kirkby et al., 1996). The WCST is predicated on executive function as a construct requiring novelty and incorporation of environmental feedback. It has been found to have high stability and reliability with a population of diverse age, structural and functional changes in the frontal lobes, and psychological disorders (Ettenhofer et al., 2006; McCabe et al., 2010). The administration of this measure provides for clinical observation of a patient under emotionally activating test procedures (Schneider, Schumann-Hengsteler, & Sodian, 2014; Zelazo & Carlson, 2012). Successful completion of the WCST task requires that patients make and maintain mental shifts and incorporates within the measure a means to assess perseveration and rigidity, disinhibition, overreactivity, and deficits in start/stop.

One thing contributing to the clinical utility of the WCST is its large research base. There are available multiple sets of normative data (Abe et al., 2004; Chelune & Baer, 1986; Mitrushina, Boone, Razani, & D'Elia, 2005), such that a clinician is able to compare performance of a patient on this measure to peers as determined along a variety of demographic variables, such as age, culture, educational attainment, and cognitive functioning, among others. This large normative data literature is a result of the decades of development and refinement of sorting tests for the purpose of evaluating executive function.

New measures of executive function lack a relative literature base. Where research validating new measures is lacking, clinicians are left to determinate whether to include the measure in their traditional battery based on unknown or imprecise heuristics (Churchill, 1979). This is particularly true for omnibus measures designed to provide standalone subtests rather than a composite performance score. One such measure, the Delis-Kaplan Executive Function System (D-KEFS) modifies several existing public domain measures; with the benefit that each of these subtests has been evaluated using the same normative data set. Research literature evaluating D-KEFS subtests as appropriate analogues of their original source material has addressed the merits of such a relationship (Crawford, Sutherland, & Garthwaite, 2008; Delis, Kramer, Kaplan, & Holdnack, 2004; Homack, Lee, & Riccio, 2005; Latzman & Markon, 2010). This literature is comparatively sparse in examination of the D-KEFS Sorting Test (Sorting Test) as an analogue of the WCST. The results of this study address this deficit, providing research regarding the construct validity of the Sorting Test as well as evaluating the merits of its atheoretical test construction.

The construct validity of the Sorting Test failed to meet the gold standard set by the WCST. Specifically, the Sorting Test failed to demonstrate convergent validity with the WCST or Letter Fluency. Most Sorting Test variables were not significantly correlated with variables on these other measures of executive function after controlling for IQ. Only Recognition was found to significantly correlate with scores on the WCST after controlling for IQ. These underwhelming results on convergent validity demonstrate mixed consistency with other research. For instance, the D-KEFS manual reports coefficients ranging from .30 to .60 for Categories Completed and .20 to .71 for Perseverative Responses whereas other research reports correlation coefficients between the Sorting Test and Categories Completed of .64 and of .15 between the Sorting Test and Perseverative Responses (Beatty & Monson, 1996). These correlations between the Sorting Test and the WCST are consistent with results obtained here, in which Recognition was significantly correlated with both WCST scores, Categories Completed, $r=.51$, and Perseverative Responses, $r=-.49$ (all $ps <.005$), but inconsistent with the absence of correlation between the WCST and Confirmed Correct Sorts. Factor analysis of the WCST has supported the use of Categories Completed and Perseverative Responses and the measure has been found to provide high clinical utility (Robinson et al., 1980). As such, the failure in this study of Sorting Test cognates Confirmed Correct Sorts and Description to correlate with their WCST score counterparts suggests that the actual value of the Sorting Test scores relative to one another is inconsistent with the description presented in the D-KEFS materials (i.e., Confirmed Correct Sorts appears to be of limited overall utility). Obtaining the Recognition score necessitates the administration of only the Recognition stage of the subtest. Experientially, this is the least complicated test component for which to maintain standardized administration. Removal of the lengthy first stage of this subtest, along with the higher task demand for the administrator, may be warranted given the failure of the first stage scores to significantly correlate with other measures of executive function. Further, the results of this study indicate that incorporating theoretically derived administration components into the Sorting Test—novelty and feedback, those same elements as seen in the WCST—failed to improve the correlation between these two measures. These nonstandard administrations of the Sorting Test resulted in statistically significant decreases in correlations between the Sorting Test and WCST.

The experimental Minimal Instruction condition consisted of non-standard Sorting Test administration with minimal instruction, designed to introduce into this task the novelty present in the WCST. This component better aligns the Sorting Test with Norman and Shallice's executive function construct as developed from Luria's model. The results of this study show statistically significant decreases in correlation between this administration with controls. Although novelty is a key component of the WCST, these findings indicate that the Sorting Test is not improved with this addition, insofar as it correlates with other measures of executive function. Functional imaging research investigating the localized component of action during executive function test administration helps to explain these results. Research analyzing large neuroimaging studies (Fine et al., 2009) using BRAINS2 image analysis software (Brain Research: Analysis of Images, Networks, and Systems) has demonstrated a singular left frontal lobe contribution to performance on the Sorting Test. Novelty, as introduced into the Sorting Test in this study, is understood to activate right frontal systems, as demonstrated in research on functional lateralization (Goldberg, Podell, & Lovell, 1994). The results of this study are consistent with this research, such that a right frontal construct component (novelty) was introduced to the Sorting Test, which recent research now suggests activates the left frontal lobe to the exclusion of right frontal systems.

Introduction of feedback to the Sorting Test administration, another component present in the WCST, better aligned the Sorting Test with Fuster and Lezak's definitions of executive function. This alternative administration, too, failed to improve correlation of the D-KEFS measure with other measures of executive function. As before, non-standard administration of the Sorting Test incorporating a theoretically derived administration style was found to decrease the correlation of these measures with one another. This study is not able to identify the mechanism by which the Sorting Test engages the executive functioning of test takers. Task novelty and incorporation of feedback to facilitate successful problem solving under changing conditions of success are shown here not to contribute to executive function activation as measured by the Sorting Test.

In this study, the Sorting Test was found to demonstrate expected divergence from Color-Word, chosen for its purported endophenotype properties discrepant from the WCST endophenotype (Taylor, 2007). None of the primary measures of the Sorting Test were significantly correlated with Color-Word. Color-Word was, however, weakly correlated with Letter Fluency $r=.36$, and both WCST scores, Categories Completed $r=.37$, Perseverative Responses $r=-.38$ (all ps $<.008$). Color-Word, as a variant on Stroop, is a conflict task and thus is purported to load on different factors of executive function than Letter Fluency and WCST. Significant correlation of Color-Word with these other scores but not with the Sorting Test, suggests that Letter Fluency and WCST are measuring a construct that is in some way different than that being measured by the Sorting Test. Another explanation may be the confounding impact of IQ on Color-Word. Given that this study found that IQ significantly correlated with Color-Word, $r=.30$, $p<.005$, the inability to partial out the effect of IQ when conducting comparisons with control group data may contribute to the lack of divergence of Letter Fluency and WCST from the Color-Word that is inconsistent with published research. The significant correlations between the WCST and Color-Word were absent under both experimental conditions. This suggests that the theoretically derived administrations improved divergent validity of the Sorting Test and WCST with the Color-Word. This finding should not be over interpreted, however, as the magnitude of these differences were not found to be statistically significant. At its least, this result provides some support for the experimental manipulations functioning as intended.

This study identified two optional scores from the Sorting Test to compare to two scores on the WCST whose nomenclature indicated were possibly assessing related construct components. Exploratory correlations found a weak correlation between Repeated Sorts and Perseverative Responses under Feedback condition, $r=-.32$ $p<.05$, but not Minimal Instruction condition. No significant correlations were found between Set-Loss and Failure to Maintain Set either experimental condition. No significant difference was found when comparing the correlations obtained in either of the two experimental conditions. This is consistent with other research comparing Set-Loss and Failure to Maintain Set which found no evidence of relation between the two variables, $r=.15$ (Beatty & Monson, 1996). These findings contradict the premise that failure to maintain set, as measured on the Sort Test, is analogous to FMS as measured on the WCST. Perseverative responding, as measured by Repeated Sorts on the Sort Test is additionally demonstrated to not be analogous to Perseverative Responses as measured on the WCST. The introduction of higher task novelty was not demonstrated to improve the correlation of these two variable pairs. The introduction of feedback was demonstrated to generate a weak correlation between Repeated Sorts and Perseverative Responses. The findings of this study add to the current literature that has yet to justify the 29 additional scores that are generated along with the five primary scores for the Sorting Test.

The two-part design of this study introduced limitations to the interpretation of these findings. In the second phase of this study, the experimental stage, IQ was found to significantly correlate with Categories Completed from the WCST but not with Confirmed Correct Sorts from the Sort Test, suggesting some shared variance of IQ and executive function that was not uniformly present across measures. Only participants in phase two of this study were administered a proxy measure of intelligence. IQ of experimental participants was normally distributed with a mean of 102, a minimum of 86, and a maximum of 121. Executive function performance has been shown to produce group differences at different levels of IQ and to have a multivariate effect (i.e., average, high average, superior) on executive functions (Mahone et al., 2002). For this study, it was possible to partial out the effect of IQ on executive function performance only for stage two participants. Given that these participants demonstrated a significant correlation between IQ and Categories Completed, and that this effect could not be partialled out of control subject performance, differences between the experimental conditions and controls investigated in phase two of this study may be attenuated. Similarly, the inability to compute partial correlations with the control data may have contributed to the non-significant correlation between Confirmed Correct Sorts and Categories Completed in phase one of this study.

All three conditions of this study sampled from a largely homogenous group restricted in range on age, culture, and educational attainment. Although the majority of demographic characteristics analyzed here were demonstrated to not significant impact test performance, homogenous sample characteristics may limit the generalizability of these findings and produce inconsistency with other research that sampled from a different or heterogeneous group. Additionally, use of undergraduate study participants is unlikely to include a normal distribution of executive function performance given educational attainment and current trajectory. Demonstrated mean average IQ of measurable study participants and restricted range of executive abilities increase the discriminative validity and precision necessary for a measure to adequately tap executive function. The D-KEFS manual defines one aim of the measure as differentiating among clinical presentations of impairment in frontal lobe functioning. Results of this study may under predict the correlation of the Sorting Test with the WCST if the D-KEFS has the discriminative sensitivity to identify deficits in participants at the test ceiling, as has been suggested in some research (Jak et al., 2009; Manchester et al., 2004).

Previous research has suggested a significant gender effect on executive function performance measures (Fuster, 2008). The findings of this study do not support such an effect but they do suggest age as a source of performance variability. Although it is not singular in its deviation from researching executive functions in older age (e.g., Huizinga, Dolan, & van der Molen, 2006), this study is unique in its use of young adult non-clinical participants to examine executive functioning. Executive functioning of 18 year olds was herein found to significantly differ from older participants, including significant difference from participants just one year older at 19 years old. These findings are consistent with growing research on executive function from a lifespan perspective (Zelazo et al., 2004). These results are particularly supported by the research of MacPherson and colleagues that has demonstrated age related differences in executive function dependent on dorsolateral prefrontal function but not in executive function dependent on ventromedial prefrontal function (MacPherson, Phillips, & Della Sala, 2002).

In conclusion, this study found mixed evidence for the convergent and divergent validity of the Sorting Test as compared to other measures of executive function. Findings were consistent with other research that cannot confirm the utility of the optional scores on the Sorting Test. These results suggest that further investigation into the construct validity of the Sorting Test may be warranted and that this measure is not an adequate substitution for the WCST.

Limitations and Future Research

The study hypotheses were largely not supported by the outcomes of this study. The core investigations of this study required testing for significant differences in correlations between experimental conditions and a control. Multiple elements may have contributed to study hypotheses being refuted, as explored below.

The findings in this study are likely to reflect diminished returns as experimental condition correlations partialled out the variance accounted for by IQ, whereas the comparison group correlations were unable to account for IQ due to absence of such data in the Control condition. As such, it is possible that the findings of this study underestimate the differences between the correlations from the experimental condition compared to the

Control condition. As the difference findings in this study suggested that the experimental manipulations decreased the correlation of the Sorting Test with the WCST, underestimation of differences due to mixed control for IQ is not expected to support the hypotheses of this study.

Sample size may have been insufficient to capture the variability of executive function in the population. Examination of study data for normality of response distribution revealed apparent clusters of outlier scores. These outliers were not removed from the study due to their clustered distributions and also the high number of identified scores. A larger sample size may have filled the distribution in such a way that these clusters would have become interpretable as part of the normal curve. This would have allowed for transformation or exclusion of true outliers and cleaner data analyses.

The restricted age of the sample around the bimodal development of executive function introduced variance that was difficult to fully explicate. Neuropsychological assessment can account for developmental changes in executive function across the lifespan by using normative data from similarly aged peers; however, there is variability in development that cannot be accounted for in this study design. It is not possible to identify which participants were operating at peak executive function development versus those before and after bimodal peak. The sample of this study may have been insufficient to account for age-related variability in the development of executive function.

Precision of this study was hampered by its implementation. Multiple clinicians were used to administer protocols in the experimental and the Control conditions. Several cases were excluded pairwise to retain study power when spoiled data were discovered.

Inconsistent administration and scoring was addressed through rescoring each subject performance twice and logging all data point issues for outlier assessment. Oversight of this type cannot fully mitigate diminished standardization in neuropsychological test administration, however, and the Sorting Test requires a more complex understanding of standardized administration than is easily achieved by multiple clinicians. Thus, additional research is needed to corroborate the findings of this study.

The study design relied on comparative measures to make inferences about the measure of interest, the Sorting Test. The choice as to which measures to use for such comparison is relevant to the confidence with which these inferences may be drawn. This study relied on the Color-Word Interference Test from the D-KEFS as a proxy of the Stroop test. The Stroop was to serve as a comparison from which the Sorting Test was expected to diverge. Neuroimaging research evaluating the classic interference task under specialized conditions permitting functional MRI during task completion (i.e., Counting-Stroop) indicates that significant anterior cingulate cortex activity is present during the Stroop (Bush et al., 1998). Although this supported the measures' use as a non-DLPFC test, the primary functions engaged by this task are mediation of response selection and allocation of attentional resources. The inhibition necessary for this task may be too similar to the inhibition necessary for the Sorting Test, making it a less precise comparison tool than some other measures.

The effect of administering both the WCST and the Sorting Test to each subject was not calculated in this study. Although counterbalancing was used to account for the order in which these two measures were administered, the impact of the first sorting test on performance on the second sorting test was not known. Counterbalancing served to ensure that the same impacts were present in both research protocols so that groups within conditions could be collapsed. It is possible that these order effects were not equal, such that the effect of being administered the WCST first was greater than the effect of being administered the Sorting Test first. In such a scenario, the comparison of the two measures would still be viable, as the conditions contain roughly equal number of participants receiving each test first, but the unequal order effect would not be identified and could represent a meaningful piece of information not considered in this study.

The manipulations created for use in this study may not have acted as intended. Introduction of novelty, as in the Minimal Instruction condition, and feedback, as in the Feedback condition, were intended to activate the DLPFC similar to how the WCST would activate this region of the frontal lobes. The alternative administrations used in this study were determined to function as intended because they fulfilled the hypothesis that they would generate no significant change in correlation between the Sorting Test and Color-Word. Although comparisons such as this are an appropriate method for measurement validation, it may be that the utility of this validation is diminished because it relied on the same data as were used for the study at-large. Use of a different sample to validate that the alternative administrations acted as expected would have strengthened the findings of this study.

Alternatively, the experimental manipulations may have activated brain structures other than the DLPFC. The introduction of novelty and performance feedback was expected to activate functions termed “cold” components of executive function (Grafman &

Litvan, 1999), but may have unintentionally activated “hot” components (Burgess, Veitch, de Lacy Costello, & Shallice, 2000). Cold executive function tasks are those that are associated with the DLPFC, such as planning and problem solving (Chan et al., 2008). Hot executive function tasks are those associated with ventromedial or orbitofrontal prefrontal cortex, such as interpersonal behavior and interpretation of emotions in social contexts (Chan et al., 2008). Activation may have been broader than anticipated, with activation of the limbic system or other non frontal lobe systems. Additional examination of the validity of the alternative test administrations may have better informed the interpretation of this study’s results.

Due to the population sampled and the self-selection inherent in the recruitment of participants, this study lacks generalizability of results to the larger national population. Study participants were between the ages of 18 and 26, 74.40% White, non-Hispanic, and were all enrolled in an undergraduate program at a state university in rural Western Pennsylvania. Future research is needed to assess the stability of these findings across more diverse age and cultural samples.

Participants were not excluded after self-report of previous TBI. Results indicated that these participants did not perform significantly different from those participants without a TBI history. Research on outcome from TBI has identified a threshold at which rehabilitation may be expected to have manifested. Those participants who experienced a TBI within three months of participating in this study would be likely to perform differently due to their injury and inadequate amount of time for full recovery to occur. As this study did not gather data regarding recency of TBI reported by participants, it is possible that some of the TBI participants were included that should have been excluded from the study. Depending on the severity of the injury, some research points to continued functional deficits at five -year follow-up (Olver, Ponsford, & Curran, 1996)

Recruitment efforts that excluded those subjects that were not available for scheduling following three outreach attempts may have inadvertently limited the inclusion of those undergraduates with lower executive function. Enrollment in the subject pool was a requirement for students in introductory psychology courses. Students who successfully enrolled in this the subject pool, signed up for this study, and attended the research session as scheduled may have possessed particularly high executive function. Sampling may have been unequal across the semester, allowing for a greater number of participants from one point in the semester over others. For example, those with better foresight abilities may have scheduled early in the semester, anticipating lack of time at the end of the semester. This could also effect the distribution of executive function observed in the sample if semester attitudes and executive function lead to uneven distribution of executive function range across the semester.

As demonstrated in this study and in previously referenced research literature, executive functions correlate with IQ such that this study's results may not reflect results that would be obtained from individuals with IQ different than this sample. Proxy IQ scores were obtained for 106 of the experimental condition participants (M=101.70, SD=7.29, min=85.58, max=120.78). Scores indicate that the sample was normally distributed around an average IQ of 100, however, so the sample for this study may be less restricted than enrollment in university may initially suggest.

The sample studied was from a non-clinical population. The Sorting Test is designed to identify executive dysfunction resulting from frontal lobe damage in clinical settings. Use of a sample that was inconsistent with the population the test was designed to assess may have hampered the utility of the test. It may be that the experimental manipulations undertaken in this study would have demonstrated the hypothesized effects if the manipulations were undertaken with a clinical population. One study investigating the effort of non-clinical undergraduate research subjects on neuropsychological tests found that suboptimal effort in this population is between 30.80% and 55.60% (An, Zakzanis, & Joordens, 2012). Given that the base rate for poor effort may be quite high with the sample used for this study and that the Sorting Test was designed for a different population, replication of this study with a clinical population would be an appropriate next step in corroborating the results of this study.

Issues related to range restriction and attenuation effects were unavoidable in data collection for this study. Due to this study's design, indirect restriction of range and Missing Not At Random data were generated, such that restriction and attenuation effects could not be uniformly corrected across conditions (Hedeker & Gibbons, 2006; Little, 1992; Little & Rubin, 2014) As the impact of these methodological problems is expected to result in underestimation of population correlations (Thorndike, 1949), it stands to reason that corrections of these statistical artifacts would serve only to further buttress the findings already discussed herein. Further, attenuation effects were near to intrinsic in the research questions, such that neither traditional Spearman double correction or latent variable modeling approaches were undertaken in the interpretation of this study's data. Future research that applies Thorndike's Case II range restriction correction formula or maximum likelihood estimates obtained from the expectation maximization algorithm may further elucidate the magnitude of the between group differences of interest to this study's hypotheses (Hunter, Schmidt, & Le, 2006).

Research examining the factor structure of the D-KEFS has described a three-factor model that may indicate a weakness in the comparison design of this study. In these models, the factor Inhibition is captured by performance on measures like Color-Word, Conceptual Flexibility is reflected in scores from the Sorting Test, and Shifting or Monitoring is reflected in WCST and other category shifting measures, such as Category Switching on the D-KEFS subtest Verbal Fluency (Latzman & Markon, 2010; Miyake et al., 2000). With the three primary Sorting Test scores retained for analysis in this study demonstrating load on a factor that is different from that which WCST has been shown to load on, it may be that the WCST was not the optimal measure to use as a comparison standard. A variant of this study that might improve the comparison design could use the three-factor structure rather than theoretical construct definitions to generate the experimental manipulations undertaken to improve the correlation between the Sorting Test and the WCST. Research has suggested that a variant factor structure is appropriate when accounting for some age-related changes in executive function (Latzman & Markon, 2010; Miyake et al., 2000). These two-factor variants have observed that Updating and Shifting, but not Inhibition, are found in the factor structure of executive function in children and adolescents (Latzman & Markon, 2010; Miyake et al., 2000). Given the age range of the sample studied and the finding that 18 year old participants performed significantly different from older participants, it is possible that developmental trajectory of executive function introduced unexpected variance into this study.

Implications

The single-use design of the WCST makes finding an analogue of the measure clinically relevant. The results of this study indicate that the Sorting Test is an insufficient substitute for the WCST. At a time when neuropsychological testing is being impacted by financial considerations of both the practitioner and patient, findings regarding duplicating assessment tools can be fiscally useful. This study provides preliminary evidence that the D-KEFS Sorting Test (Sorting Test) is insufficient to replace the WCST due to deviations in the executive function construct components assessed by each measure, as reflected in lower than expected convergent validity. Moreover, these findings do not support the use of optional scores obtained with the Sorting Test to approximate similarly named scores from WCST, as they failed to demonstrate substantive equality. Incorporation of novelty or feedback to enhance the correlation of the Sorting Test with the WCST was not supported by the findings of this study.

As this study investigated measures of executive function that are well known and often used in varieties of settings and contexts, findings regarding construct validity of these measures are significant to the practice of clinical neuropsychology. The findings of this study indicate that neuropsychologists assessing executive function should not equate the Sorting Test and WCST, as these two measures have not demonstrated consistent construct validity. Given the multi-source multi-method approach of ethical neuropsychological assessment, these findings may not require significant modification of assessment procedures. Conflict-laden or litigious areas of practice, such as forensic assessment, may dictate a prophylactic removal of the Sorting Test from a clinician's battery measuring executive function. These findings do not directly reflect on the general overall utility of the Sorting Test. Although these findings indicate that the Sorting Test does not function as an analogue of the WCST in the population sampled, it may function as a measure of a different component of executive function. To investigate beyond the hypotheses of this study and evaluate the clinical utility of the Sorting Test, use of regression analysis and a larger set of input sources would be required. This question may be best considered in light of research that is questioning the present methods of neuropsychological assessment. Researchers evaluating the discriminant and ecological validity of tests such as the Sorting Test (Manchester et al., 2004) may soon be evaluating this question of clinical utility from a very different angle than pure regression analysis.

References

Abe, M., Suzuki, K., Okada, K., Miura, R., Fujii, T., Etsurou, M., & Yamadori, A. (2004). Normative data on tests for frontal lobe functions: Trail Making Test, Verbal fluency, Wisconsin Card Sorting Test (Keio version). *Brain and nerve*, *56*(7), 567–574.

Adams, N. C., & Jarrold, C. (2009). Inhibition and validity of the Stroop task for children with autism. *Journal of Autism and Developmental Disorders*, *39*(8), 1112–1121. doi:10.1007/s10803-009-0721-8

Alevriadou, A., Katsarou, Z., Bostantijopoulou, S., Kiosseoglou, G., & Mentenopoulos, G. (1999). Wisconsin Card Sorting Test variables in relation to motor symptoms in Parkinson's disease. *Perceptual and Motor Skills*, *89*(3), 824–830.

Altman, D. G., & Bland, J. M. (1995). Statistics notes: The normal distribution. *The BMJ*, *310*(6975), 298–298. doi:10.1136/bmj.310.6975.298

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing (US). (1999). *Standards for educational and psychological testing*. Amer Educational Research Assn.

Amunts, K., Schleicher, A., & Zilles, K. (2007). Cytoarchitecture of the cerebral cortex—More than localization. *NeuroImage*, *37*(4), 1061–1065. doi10.1016/j.neuroimage.2007.02.037

Amunts, K., & Zilles, K. (2001). Advances in cytoarchitectonic mapping of the human cerebral cortex. *Neuroimaging Clinics of North America*, *11*(2), 151–69, vii.

Anderson, S. W., Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature Neuroscience*, *2*(11), 1032–1037. doi:10.1038/14833

An, K. Y., Zakzanis, K. K., & Joordens, S. (2012). Conducting research with non-clinical healthy undergraduates: Does effort play a role in neuropsychological test performance? *Archives of Clinical Neuropsychology*, acs085. doi:10.1093/arclin/acs085

Archibald, S. J., Mateer, C. A., & Kerns, K. A. (2001). Utilization behavior: Clinical manifestations and neurological mechanisms. *Neuropsychology Review*, *11*(3), 117–130. doi:10.1023/A:1016673807158

Arffa, S. (2007). The relationship of intelligence to executive function and non-executive function measures in a sample of average, above average, and gifted youth. *Archives of Clinical Neuropsychology*, *22*(8), 969–978. doi:10.1016/j.acn.2007.08.001

Armengol, C. G. (2002). Stroop test in Spanish: Children's norms. *The Clinical Neuropsychologist*, *16*(1), 67–80. doi:10.1076/clin.16.1.67.8337

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559. doi:10.1126/science.1736359

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative Science Quarterly*, *36*(3), 421–458. doi:10.2307/2393203

Baldo, J. V., Shimamura, A. P., Delis, D. C., Kramer, J., & Kaplan, E. (2001). Verbal and design fluency in patients with frontal lobe lesions. *Journal of the International Neuropsychological Society*, *7*(5), 586–596. doi:10.1017/S1355617701755063

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin*, *121*(1), 65–94. doi:10.1037/0033-2909.121.1.65

Basso, M. R., Bornstein, R. A., & Lang, J. M. (1999). Practice effects on commonly used measures of executive function across twelve months. *The Clinical Neuropsychologist*, *13*(3), 283–292. doi:10.1076/clin.13.3.283.1743

Beatty, W. W., Jocic, Z., Monson, N., & Katzung, V. M. (1994). Problem solving by schizophrenic and schizoaffective patients on the Wisconsin and California Card Sorting Tests. *Neuropsychology*, *8*(1), 49–54. doi:10.1037/0894-4105.8.1.49

Beatty, W. W., Katzung, V. M., Nixon, S. J., & Moreland, V. J. (1993). Problem-solving deficits in alcoholics: Evidence from the California Card Sorting Test. *Journal of Studies on Alcohol*, *54*(6), 687–692.

Beatty, W. W., & Monson, N. (1990). Problem solving in Parkinson's disease: Comparison of performance on the Wisconsin and California Card Sorting Tests. *Journal of Geriatric Psychiatry and Neurology*, *3*(3), 163–171. doi:10.1177/089198879000300308

Beatty, W. W., & Monson, N. (1996). Problem solving by patients with multiple sclerosis: Comparison of performance on the Wisconsin and California Card Sorting Tests. *Journal of the International Neuropsychological Society*, *2*(2), 134–140. doi:10.1017/S1355617700000989

Beauvais, J. E., Woods, S. P., Delaney, R. C., & Fein, D. (2004). Development of a tactile Wisconsin Card Sorting Test. *Rehabilitation Psychology*, *49*(4), 282–287. doi:10.1037/0090-5550.49.4.282

Bechara, A. (2004). The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage. *Brain and Cognition*, *55*(1), 30–40. doi:10.1016/j.bandc.2003.04.001

Benton, A. L., Hamsher, K., & Sivan, A. B. (1994). Multilingual Aphasia Examination: Manual of instructions. Iowa City, IA: AJA Associates.

Berman, K. F., Torrey, E. F., Daniel, D. G., & Weinberger, D. R. (1992). Regional cerebral blood flow in monozygotic twins discordant and concordant for schizophrenia. *Archives of General Psychiatry*, *49*(12), 927–934. doi:10.1001/archpsyc.1992.01820120015004

Bernard, L. C., McGrath, M. J., & Houston, W. (1996). The differential effects of simulating malingering, closed head injury, and other CNS pathology on the Wisconsin Card Sorting Test: Support for the "pattern of performance" hypothesis. *Archives of Clinical Neuropsychology*, *11*(3), 231–245. doi:10.1093/arclin/11.3.231

Binet, A., & Simon, T. (1948). The development of the Binet-Simon scale, 1905-1908. In Dennis, W. (Ed.), *Readings in the history of psychology* (pp. 412–424). East Norwalk, CT, Appleton-Century-Crofts.

Blair, J. R., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test. *Clinical Neuropsychologist*, *3*(2), 129–136. doi:10.1080/13854048908403285

Boone, K. B., Victor, T. L., Wen, J., Razani, J., & Pontón, M. (2007). The association between neuropsychological scores and ethnicity, language, and acculturation variables in a large patient population. *Archives of Clinical Neuropsychology*, *22*(3), 355–365. doi:10.1016/j.acn.2007.01.010

Brodmann, K. (2006). *Brodmann's localization in the cerebral cortex: the principles of comparative localisation in the cerebral cortex based on cytoarchitectonics*. New York, NY: Springer.

Brooks, B., Strauss, E., Sherman, E., Iverson, G., & Slick, D. (2009). Developments in neuropsychological assessment: Refining psychometric and clinical interpretive methods. *Canadian Psychology*, *50*(3), 196. doi:10.1037/a0016066

Brown, M. B., & Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, *69*(346), 364–367. doi:10.1080/01621459.1974.10482955

Burgess, P. W., Veitch, E., de Lacy Costello, A., & Shallice, T. (2000). The cognitive and neuroanatomical correlates of multitasking. *Neuropsychologia*, *38*(6), 848–863. doi:10.1016/S0028-3932(99)00134-7

Bush, G., Whalen, P. J., Rosen, B. R., Jenike, M. A., McInerney, S. C., & Rauch, S. L. (1998). The counting Stroop: an interference task specialized for functional neuroimaging--validation study with functional MRI. *Human Brain Mapping*, *6*(4), 270–282.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. SAGE Publications.

Cassel, R., Johnson, A., & Burns, W. (1962). The order of tests in the battery. *Journal of Clinical Psychology*, *18*(4), 464–465. doi:10.1002/1097-4679(196210)18:4<464::AID-JCLP2270180420>3.0.CO;2-W

Chan, R. C. K., Shum, D., Toulopoulou, T., & Chen, E. Y. H. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology*, *23*(2), 201–216. doi:10.1016/j.acn.2007.08.010

Chan, R. C. K., Shum, D., & Cheung, R. (2003). Recent development of cognitive and neuropsychological assessment in Asian countries. *Psychological Assessment*, *15*(3), 257–267. doi:10.1037/1040-3590.15.3.257

Chelune, G. J., & Baer, R. A. (1986). Developmental norms for the Wisconsin Card Sorting Test. *Journal of Clinical and Experimental Neuropsychology*, *8*(3), 219–228. doi:10.1080/01688638608401314

Churchill, G. A., Jr. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, *16*(1), 64–73. doi:10.2307/3150876

Clark, L., Bechara, A., Damasio, H., Aitken, M. R. F., Sahakian, B. J., & Robbins, T. W. (2008). Differential effects of insular and ventromedial prefrontal cortex lesions on risky decision-making. *Brain*, *131*(5), 1311–1322. doi:10.1093/brain/awn066

Coffey, D. M., Marmol, L., Schock, L., & Adams, W. (2005). The influence of acculturation on the Wisconsin Card Sorting Test by Mexican Americans. *Archives of Clinical Neuropsychology*, *20*(6), 795–803. doi:10.1016/j.acn.2005.04.009

Cohen, M. J., & Stanczak, D. E. (2000). On the reliability, validity, and cognitive structure of the Thurstone Word Fluency Test. *Archives of Clinical Neuropsychology*, *15*(3), 267–279. doi:10.1016/S0887-6177(99)00017-7

Coolidge, F. L., & Wynn, T. (2001). Executive functions of the frontal lobes and the evolutionary ascendancy of Homo Sapiens. *Cambridge Archaeological Journal*, *11*(02), 255–260. doi:10.1017/S0959774301000142

Crawford, J. R., Allan, K. M., Cochrane, R. H. B., & Parker, D. M. (1990). Assessing the validity of NART-estimated premorbid IQs in the individual case. *British Journal of Clinical Psychology*, *29*(4), 435–436. doi:10.1111/j.2044-8260.1990.tb00908.x

Crawford, J. R., Besson, J. A., Bremner, M., Ebmeier, K. P., Cochrane, R. H., & Kirkwood, K. (1992). Estimation of premorbid intelligence in schizophrenia. *The British Journal of Psychiatry*, *161*(1), 69–74. doi:10.1192/bjp.161.1.69

Crawford, J. R., Sutherland, J., & Garthwaite, P.H. (2008). On the reliability and standard errors of measurement of contrast measures from the D-KEFS. *Journal of the International Neuropsychological Society*, *14*(06), 1069–1073. doi:10.1017/S1355617708081228

Crouch, J. A., Greve, K. W., & Brooks, J. (1996). The California Card Sorting Test may dissociate verbal and non-verbal concept formation abilities. *British Journal of Clinical Psychology*, *35*(3), 431–434. doi:10.1111/j.2044-8260.1996.tb01196.x

Daffner, K. R., Mesulam, M. M., Scinto, L. F. M., Acar, D., Calvo, V., Faust, R., … Holcomb, P. (2000). The central role of the prefrontal cortex in directing attention to novel events. *Brain*, *123*(5), 927–939. doi:10.1093/brain/123.5.927

Damasio, H. (1991). Neuroanatomy of frontal lobe in vivo: A comment on methodology. In *Frontal lobe function and dysfunction* (pp. 92–121). Oxford: Oxford University Press.

Delis, D. C. (2001). *Delis-Kaplan Executive Function System (D-KEFS) examiner's manual*. San Antonio, TX: The Psychological Corporation.

Delis, D. C. (2010). Edith Kaplan (1924–2009). *American Psychologist*, *65*(2), 127–128. doi:10.1037/a0018250

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System (D-KEFS)*. Psychological Corporation.

Delis, D. C., Kramer, J. H., Kaplan, E., & Holdnack, J. (2004). Reliability and validity of the Delis-Kaplan Executive Function System: An update. *Journal of the International Neuropsychological Society*, *10*(02), 301–303. doi:10.1017/S1355617704102191

Delis, D. C., Squire, L. R., Bihrle, A., & Massman, P. (1992). Componential analysis of problem-solving ability: Performance of patients with frontal lobe damage and amnesic patients on a new sorting test. *Neuropsychologia*, *30*(8), 683–697. doi:10.1016/0028-3932(92)90039-O

Demakis, G. J. (2003). A meta-analytic review of the sensitivity of the Wisconsin Card Sorting Test to frontal and lateralized frontal brain damage. *Neuropsychology*, *17*(2), 255–264. doi:10.1037/0894-4105.17.2.255

Diamond, A. (1991). Guidelines for the study of brain-behavior relationships during development. In *Frontal lobe function and dysfunction* (pp. 339–378). Oxford: Oxford University Press.

Dimitrov, M., Grafman, J., & Hollnagel, C. (1996). The Effects of frontal lobe damage on everyday problem solving. *Cortex*, *32*(2), 357–366. doi:10.1016/S0010-9452(96)80057-8

Dimitrov, M., Grafman, J., Soares, A. H. R., & Clark, K. (1999). Concept formation and concept shifting in frontal lesion and Parkinson's disease patients assessed with the California Card Sorting Test. *Neuropsychology*, *13*(1), 135–143. doi:10.1037/0894-4105.13.1.135

Eling, P., Derckx, K., & Maes, R. (2008). On the historical and conceptual background of the Wisconsin Card Sorting Test. *Brain and Cognition*, *67*(3), 247–253. doi:10.1016/j.bandc.2008.01.006

Elliott, A. C., & Woodward, W. A. (2007). *Statistical analysis quick reference guidebook: With SPSS examples*. SAGE.

Eslinger, P.J. (1996). Conceptualizing, describing, and measuring components of executive function: A summary. In Lyon & Krasnegor (Eds.), *Attention, memory, and executive function* (pp. 367–395). Baltimore, MD: Paul H. Brookes Publishing.

Eslinger, P.J. (1999). Orbital frontal cortex: Historical and contemporary views about its behavioral and physiological significance. An introduction to special topic papers: Part I. *Neurocase*, *5*(3), 225–229. doi:10.1080/13554799908402727

Eslinger, P.J., Grattan, L.M., & Geder, L. (1995). Impact of frontal lobe lesions on rehabilitation and recovery from acute brain injury. *NeuroRehabilitation*, *5*(2), 161–182. doi:10.3233/NRE-1995-5206

Ettenhofer, M. L., Hambrick, D. Z., & Abeles, N. (2006). Reliability and stability of executive functioning in older adults. *Neuropsychology*, *20*(5), 607–613. doi:10.1037/0894-4105.20.5.607

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. SAGE.

Fine, E. M., Delis, D. C., Dean, D., Beckman, V., Miller, B. L., Rosen, H. J., & Kramer, J. H. (2009). Left frontal lobe contributions to concept formation: A quantitative MRI study of performance on the Delis–Kaplan Executive Function System Sorting Test. *Journal of Clinical and Experimental Neuropsychology*, *31*(5), 624–631. doi:10.1080/13803390802419017

Franzen, M. D., Tishelman, A. C., Sharp, B. H., & Friedman, A. G. (1987). An investigation of the test-retest reliability of the Stroop Color-Word Test across two intervals. *Archives of Clinical Neuropsychology*, *2*(3), 265–272. doi:10.1016/0887-6177(87)90014-X

Frederiksen, N. (1986). Toward a broader conception of human intelligence. *American Psychologist*, *41*(4), 445–452. doi:10.1037/0003-066X.41.4.445

Fuster, J. M. (1991). Role of prefrontal cortex in delay tasks: Evidence from reversible lesion and unit recording in the monkey. In *Frontal lobe function and dysfunction* (pp. 59–71). Oxford: Oxford University Press.

Fuster, J. M. (2000). Prefrontal neurons in networks of executive memory. *Brain Research Bulletin*, *52*(5), 331–336. doi:10.1016/S0361-9230(99)00258-0

Fuster, J. M. (2001). The prefrontal cortex—an update: Time is of the essence. *Neuron*, *30*(2), 319–333. doi:10.1016/S0896-6273(01)00285-9

Fuster, J. M. (2004). Upper processing stages of the perception–action cycle. *Trends in Cognitive Sciences*, *8*(4), 143–145. doi:10.1016/j.tics.2004.02.004

Fuster, J. M. (2008). *The Prefrontal Cortex* (Fourth). Amsterdam: Elsevier.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, *10*(2), 486–489. doi:10.5812/ijem.3505

Gioia, G. A., Isquith, P. K., Kenworthy, L., & Barton, R. M. (2002). Profiles of everyday executive function in acquired and developmental disorders. *Child Neuropsychology*, *8*(2), 121–137. doi:10.1076/chin.8.2.121.8727

Goldberg, E., Podell, K., & Lovell, M. (1994). Lateralization of frontal lobe functions and cognitive novelty. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *6*(4), 371–378.

Golden, C. J., Freshwater, S. M., & Golden, Z. (2003). Stroop Color and Word Test [Adult and Children's Versions, Revised]. Stoelting Co., 620 Wheat Lane, Wood Dale, IL 60191; Telephone: 630-860-9700; FAX: 630-860-9775; E-mail: psychtests@stoeltingco.com; Web: www.stoeltingco.com/tests.

Golden, C., Freshwater, S., & Golden, Z. (1985). Stroop Color and Word Test adult and children's versions revised. *Mental Measurement Yearbook Wood Dale, Illinois: Stoeltin*

Goldman-Rakic, P., & Friedman, H. (1991). The circuitry of working memory revealed by anatomy and metabolic imaging. In *Frontal lobe function and dysfunction* (pp. 72–91). Oxford: Oxford University Press.

Goldstein, K., Scheerer, M., & Hanfmann, E. (1953). Tests of abstract and concrete thinking. In *Contributions toward medical psychology: Theory and psychodiagnostic methods, Vol II* (pp. 702–740). New York, NY, Ronald Press Company.

Goldstein, S., & Naglieri, J. A. (2013). *Handbook of Executive Functioning*. Springer Science & Business Media.

Gottfredson, L., & Saklofske, D. (2009). Intelligence: Foundations and issues in assessment. *Canadian Psychology*, *50*(3), 183. doi:10.1037/a0016641

Grafman, J., & Litvan, I. (1999). Importance of deficits in executive functions. *The Lancet*, *354*(9194), 1921–1923. doi:10.1016/S0140-6736(99)90438-5

Grant, I., & Adams, K. (2009). *Neuropsychological Assessment of Neuropsychiatric and Neuromedical Disorders*. Oxford: Oxford University Press.

Greve, K. W., Farrell, J. F., Besson, P. S., & Crouch, J. A. (1995). A psychometric analysis of the California Card Sorting Test. *Archives of Clinical Neuropsychology*, *10*(3), 265–278. doi:10.1016/0887-6177(94)00046-S

Hakoda, Y., & Sasaki, M. (1990). Group version of the Stroop and reverse-Stroop Test: The effects of reaction mode, order and practice. *Japanese Journal of Educational Psychology*, *38*(4), 389–394.

Harding, I. H., Harrison, B. J., Breakspear, M., Pantelis, C., & Yücel, M. (2014). Cortical representations of cognitive control and working memory are dependent yet non-interacting. *Cerebral Cortex*, bhu208. doi:10.1093/cercor/bhu208

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal Data Analysis*. John Wiley & Sons.

Heflin, L. H., Laluz, V., Jang, J., Ketelle, R., Miller, B. L., & Kramer, J. H. (2011). Let's inhibit our excitement: The relationships between Stroop, behavioral disinhibition, and the frontal lobes. *Neuropsychology*, *25*(5), 655–665. doi:10.1037/a0023863

Heled, E., Hoofien, D., Margalit, D., Natovich, R., & Agranov, E. (2012). The Delis–Kaplan Executive Function System Sorting Test as an evaluative tool for executive functions after severe traumatic brain injury: A comparative study. *Journal of Clinical and Experimental Neuropsychology*, *34*(2), 151–159. doi:10.1080/13803395.2011.625351

Homack, S., Lee, D., & Riccio, C. A. (2005). Test Review: Delis-Kaplan Executive Function System. *Journal of Clinical and Experimental Neuropsychology*, *27*(5), 599–609. doi:10.1080/13803390490918444

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*(11), 2017–2036. doi:10.1016/j.neuropsychologia.2006.01.010

Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, *91*(3), 594–612. doi:10.1037/0021-9010.91.3.594

Jak, A. J., Bondi, M. W., Delano-Wood, L., Wierenga, C., Corey-Bloom, J., Salmon, D. P., & Delis, D. C. (2009). Quantification of five neuropsychological approaches to defining mild cognitive impairment. *The American Journal of Geriatric Psychiatry*, *17*(5), 368–375. doi:10.1097/JGP.0b013e31819431d5

Johnstone, B., Callahan, C. D., Kapila, C. J., & Bouman, D. E. (1996). The comparability of the WRAT-R reading test and NAART as estimates of premorbid intelligence in neurologically impaired patients. *Archives of Clinical Neuropsychology*, *11*(6), 513–519. doi:10.1093/arclin/11.6.513

Jones-Gotman, M. (1990). Localization of lesions by neuropsychological testing. *Epilepsia*, *32 (5)*, S41–52.

Kirkby, B. S., van Horn, J. D., Ostrem, J. L., & Weinberger, D. R. (1996). Cognitive activation during PET: A case study of monozygotic twins discordant for closed head injury. *Neuropsychologia*, *34*(7), 689–697. doi:10.1016/0028-3932(95)00147-6

Knight, R. (1991). Evoked potential studies of attention capacity in human frontal lobe lesions. In *Frontal lobe function and dysfunction* (pp. 139–153). Oxford: Oxford University Press.

Koechlin, E., Basso, G., Pietrini, P., Panzer, S., & Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. *Nature*, *399*(6732), 148–151. doi:10.1038/20178

Koechlin, E., Ody, C., & Kounelher, F. (2003). The Architecture of cognitive control in the Human Prefrontal Cortex. *Science*, *302*(5648), 1181–1185. doi:10.1126/science.1088545

Kongs, S. K., Thompson, L. L., Iverson, G. L., & Heaton, R. K. (2000). Wisconsin Card Sorting Test-64 card version (WCST-64). *Odessa, FL: Psychological Assessment Resources*.

Lamm, C., Zelazo, P. D., & Lewis, M. D. (2006). Neural correlates of cognitive control in childhood and adolescence: Disentangling the contributions of age and executive function. *Neuropsychologia*, *44*(11), 2139–2148. doi:10.1016/j.neuropsychologia.2005.10.013

Latzman, R. D., & Markon, K. E. (2010). The factor structure and age-related factorial invariance of the Delis-Kaplan Executive Function System (D-KEFS). *Assessment*, *17*(2), 172–184. doi:10.1177/1073191109356254

Leany, B. D., Benuto, L. T., & Thaler, N. S. (2014). Neuropsychological assessment with Asian American clients. In L. T. Benuto, N. S. Thaler, & B. D. Leany (Eds.), *Guide to Psychological Assessment with Asians* (pp. 427–441). New York: Springer. doi:10.1007/978-1-4939-0796-0_27

Levine, A. J., Miller, E. N., Becker, J. T., Selnes, O. A., & Cohen, B. A. (2004). Normative data for determining significance of test-retest differences on eight common neuropsychological instruments. *The Clinical Neuropsychologist*, *18*(3), 373–384. doi:10.1080/1385404049052420

Levin, H. S., Benton, A. L., & Grossman, R. G. (1982). *Neurobehavioral Consequences of Closed Head Injury*. Oxford: Oxford University Press.

Lezak, M., Howieson, D., Bigler, E., & Tranel, D. (2012). *Neuropsychological Assessment* (Fifth). Oxford: Oxford University Press.

Lezak, M., Howieson, D., Loring, D., Hannay, H., & Fischer, J. (2004). *Neuropsychological Assessment* (Fourth). Oxford: Oxford University Press.

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov Test for Normality with mean and variance unknown. *Journal of the American Statistical Association*, *62*(318), 399–402. doi:10.1080/01621459.1967.10482916

Little, R. J. A. (1992). Regression with missing x's: A review. *Journal of the American Statistical Association*, *87*(420), 1227–1237. doi:10.1080/01621459.1992.10476282

Little, R. J. A., & Rubin, D. B. (2014). *Statistical analysis with missing data*. John Wiley & Sons.

Lombardi, W. J., Andreason, P. J., Sirocco, K. Y., Rio, D. E., Gross, R. E., Umhau, J. C., & Hommer, D. W. (1999). Wisconsin Card Sorting Test performance following head injury: Dorsolateral Fronto-Striatal Circuit Activity Predicts Perseveration. *Journal of Clinical and Experimental Neuropsychology*, *21*(1), 2–16. doi:10.1076/jcen.21.1.2.940

Luca, C. R. D., Wood, S. J., Anderson, V., Buchanan, J.-A., Proffitt, T. M., Mahony, K., & Pantelis, C. (2003). Normative data from the CANTAB. I: Development of executive function over the lifespan. *Journal of Clinical and Experimental Neuropsychology*, *25*(2), 242–254. doi:10.1076/jcen.25.2.242.13639

MacPherson, S. E., Phillips, L. H., & Della Sala, S. (2002). Age, executive function and social decision making: A dorsolateral prefrontal theory of cognitive aging. *Psychology and Aging*, *17*(4), 598–609. doi:10.1037/0882-7974.17.4.598

Mahone, E. M., Hagelthorn, K. M., Cutting, L. E., Schuerholz, L. J., Pelletier, S. F., Rawlins, C., … Denckla, M. B. (2002). Effects of IQ on executive function measures in children with ADHD. *Child Neuropsychology*, *8*(1), 52–65. doi:10.1076/chin.8.1.52.8719

Malloy, P., Bihrle, A., Duffy, J., & Cimino, C. (1993). The orbitomedial frontal syndrome. *Archives of Clinical Neuropsychology*, *8*(3), 185–201. doi:10.1093/arclin/8.3.185

Manchester, D., Priestley, N., & Jackson, H. (2004). The assessment of executive functions: Coming out of the office. *Brain Injury*, *18*(11), 1067–1081. doi:10.1080/02699050410001672387

Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: Structure and processes. *Current Opinion in Neurobiology*, *11*(2), 194–201. doi:10.1016/S0959-4388(00)00196-3

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*, *24*(2), 222–243. doi:10.1037/a0017619

McDonald, C. R., Delis, D. C., Norman, M. A., Tecoma, E. S., & Iragui, V. J. (2005). Discriminating patients with frontal-lobe epilepsy and temporal-lobe epilepsy: Utility of a multilevel Design Fluency Test. *Neuropsychology*, *19*(6), 806–813. doi:10.1037/0894-4105.19.6.806

McGurn, B., Starr, J. M., Topfer, J. A., Pattie, A., Whiteman, M. C., Lemmon, H. A., … Deary, I. J. (2004). Pronunciation of irregular words is preserved in dementia, validating premorbid IQ estimation. *Neurology*, *62*(7), 1184–1186. doi:10.1212/01.WNL.0000103169.80910.8B

Middleton, F. A., & Strick, P. L. (2000). Basal ganglia and cerebellar loops: Motor and cognitive circuits. *Brain Research Reviews*, *31*(2–3), 236–250. doi:10.1016/S0165-0173(99)00040-5

Miller, B. L. (2007). The human frontal lobes: An introduction. In *The Human Frontal Lobes Functions and Disorders* (Second Edition). New York: The Guilford Press.

Mitrushina, M., Boone, K. B., Razani, J., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment*. Oxford: Oxford University Press.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. doi:10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. doi:10.1006/cogp.1999.0734

Nelson, H. E., & O'Connell, A. (1978). Dementia: The estimation of premorbid intelligence levels using the new adult reading test. *Cortex*, *14*(2), 234–244. doi:10.1016/S0010-9452(78)80049-5

Nelson, H., & Willison, J. (1991). *The revised National Adult Reading Test (NART): Test manual* (2nd ed.). Windsor, UK: NFER Nelson.

Nutter-Upham, K. E., Saykin, A. J., Rabin, L. A., Roth, R. M., Wishart, H. A., Pare, N., & Flashman, L. A. (2008). Verbal fluency performance in amnestic MCI and older adults with cognitive complaints. *Archives of Clinical Neuropsychology*, *23*(3), 229–241. doi:10.1016/j.acn.2008.01.005

Nyhus, E., & Barceló, F. (2009). The Wisconsin Card Sorting Test and the cognitive assessment of prefrontal executive functions: A critical update. *Brain and Cognition*, *71*(3), 437–451. doi:10.1016/j.bandc.2009.03.005

O'Leary-Kelly, S. W., & Vokurka, R. (1998). The empirical assessment of construct validity. *Journal of Operations Management*, *16*(4), 387–405. doi:10.1016/S0272-6963(98)00020-5

Olver, J. H., Ponsford, J. L., & Curran, C. A. (1996). Outcome following traumatic brain injury: a comparison between 2 and 5 years after injury. *Brain Injury*, *10*(11), 841–848. doi:10.1080/026990596123945

Owen, A. M. (1997). The functional organization of working memory processes within human lateral frontal cortex: The contribution of functional neuroimaging. *European Journal of Neuroscience*, *9*(7), 1329–1339. doi:10.1111/j.1460-9568.1997.tb01487.x

Ownsworth, T., & Shum, D. (2008). Relationship between executive functions and productivity outcomes following stroke. *Disability and Rehabilitation: An International, Multidisciplinary Journal*, *30*(7), 531–540. doi:10.1080/09638280701355694

Ozonoff, S. (1995). Reliability and validity of the Wisconsin Card Sorting Test in studies of autism. *Neuropsychology*, *9*(4), 491–500. doi: 10.1037/0894-4105.9.4.491

Pallant, J. (2013). *Spss survival manual*. UK: McGraw-Hill Education.

Parmenter, B. A., Zivadinov, R., Kerenyi, L., Gavett, R., Weinstock-Guttman, B., Dwyer, M. G., … Benedict, R. H. B. (2007). Validity of the Wisconsin Card Sorting and Delis-Kaplan

Executive Function System (DKEFS) Sorting Tests in multiple sclerosis. *Journal Of Clinical And Experimental Neuropsychology*, *29*(2), 215–223.

Pavawalla, S. P., Schmitter-Edgecombe, M., & Smith, R. E. (2012). Prospective memory after moderate-to-severe traumatic brain injury: A multinomial modeling approach. *Neuropsychology*, *26*(1), 91–101. doi:10.1037/a0025866

Pearson Education, Inc. (2012). Delis-Kaplan Executive Function System (D-KEFS) assessment and information.

Possin, K. L., Chester, S. K., Laluz, V., Bostrom, A., Rosen, H. J., Miller, B. L., & Kramer, J. H. (2012). The frontal-anatomic specificity of Design Fluency repetitions and their diagnostic relevance for behavioral variant frontotemporal dementia. *Journal of the International Neuropsychological Society*, *18*(05), 834–844. doi:10.1017/S1355617712000604

Priebe, N., & Ferster, D. (2010). Each synapse to its own. *Nature*, *464*, 1290–1291.

Prigatano, G. (1991). Rehabilitation of the frontal lobe damaged patient. In H. Eisenberg, A. Benton, & H. Levine (Eds.), *Frontal lobe function and dysfunction* (pp. 381–397). Oxford: Oxford University Press.

Rabbitt, P. (2004). *Methodology of frontal and executive function*. Psychology Press.

Ramsden, P. (2003). Review of the Delis-Kaplan Executive Function System. In B. Plake, J. Impara, & R. Spies (Eds.), *The fifteenth mental measurements yearbook with tests in print* (Vol. 15). Lincoln, NE: Buros Institute of Mental Measurement.

Robinson, A. L., Heaton, R. K., Lehman, R. A., & Stilson, D. W. (1980). The utility of the Wisconsin Card Sorting Test in detecting and localizing frontal lobe lesions. *Journal* *of Consulting and Clinical Psychology*, *48*(5), 605–614. doi:10.1037/0022-006X.48.5.605

Rolls, E. T., Hornak, J., Wade, D., & McGrath, J. (1994). Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *Journal of Neurology, Neurosurgery & Psychiatry*, *57*(12), 1518–1524. doi:10.1136/jnnp.57.12.1518

Ross, T. P. (2003). The reliability of cluster and switch scores for the Controlled Oral Word Association Test. *Archives of Clinical Neuropsychology*, *18*(2), 153–164. doi:10.1016/S0887-6177(01)00192-5

Ruff, R. M., Light, R. H., Parker, S. B., & Levin, H. S. (1996). Benton Controlled Oral Word Association Test: Reliability and updated norms. *Archives of Clinical Neuropsychology*, *11*(4), 329–338. doi:10.1016/0887-6177(95)00033-X

Sachs, B. C. (2011). The diagnostic utility of a multi-task verbal fluency paradigm in frontal and temporal lobe epilepsy: An analysis of fluency type and qualitative performance. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *71*(12-B), 7737.

Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology*, *19*(4), 532–545. doi:10.1037/0894-4105.19.4.532

Savitz, J. B., & Jansen, P. (2003). The Stroop Color-Word Interference Test as an indicator of ADHD in poor readers. *The Journal of Genetic Psychology: Research and Theory on Human Development*, *164*(3), 319–333. doi:10.1080/00221320309597986

Schneider, W., Schumann-Hengsteler, R., & Sodian, B. (2014). *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability, and theory of mind*. Psychology Press.

Shallice, T., & Burgess, P. (1991). Higher-order cognitive impairments and frontal lobe lesions in man. In *Frontal lobe function and dysfunction* (pp. 125–138). Oxford: Oxford University Press.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (Complete samples). *Biometrika*, *52*(3/4), 591–611. doi:10.2307/2333709

Sohlberg, M. M., & Mateer, C. A. (2001). *Cognitive rehabilitation: An integrative neuropsychological approach*. Guilford Press.

Steinmetz, J.P., Brunner, M., Loarer, E., & Houssemand, C. (2010). Incomplete psychometric equivalence of scores obtained on the manual and the computer version of the Wisconsin Card Sorting Test? *Psychological Assessment*, *22*(1), 199–202. doi:10.1037/a0017661

Strauss, E., Sherman, E., & Spreen, O. (2006). *A Compendium of neuropsychological tests: Administration, norms, and commentary* (Third). Oxford: Oxford University Press.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. doi:10.1037/h0054651

Stuss, D. T. (1991). Interference effects on memory function in postleukotomy patients: An attentional perspective. In *Frontal lobe function and dysfunction*. Oxford: Oxford University Press.

Stuss, D. T. (2011). Functions of the frontal lobes: Relation to executive functions. *Journal of the International Neuropsychological Society*, *17*(05), 759–765. doi:10.1017/S1355617711000695

Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: A conceptual view. *Psychological Research*, *63*(3-4), 289–298. doi:10.1007/s004269900007

Stuss, D. T., Gow, C. A., & Hetherington, C. R. (1992). "No longer Gage": Frontal lobe dysfunction and emotional changes. *Journal of Consulting and Clinical Psychology*, *60*(3), 349–359. doi:10.1037/0022-006X.60.3.349

Stuss, D. T., Levine, B., Alexander, M. P., Hong, J., Palumbo, C., Hamer, L., … Izukawa, D. (2000). Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: Effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia*, *38*(4), 388–402. doi:10.1016/S0028-3932(99)00093-7

Stuss, D. T., Picton, T. W., & Alexander, M. P. (2001). Consciousness, self-awareness, and the frontal lobes. In S. P. Salloway, P. F. Malloy, & J. D. Duffy (Eds.), *The frontal lobes and neuropsychiatric illness* (pp. 101–109). Arlington, VA: American Psychiatric Publishing, Inc.

Stuss, D. T., Pogue, J., Buckle, L., & Bondar, J. (1994). Characterization of stability of performance in patients with traumatic brain injury: Variability and consistency on reaction time tests. *Neuropsychology*, *8*(3), 316–324. doi:10.1037/0894-4105.8.3.316

Stuss, D. T., & Trites, R. L. (1977). Classification of neurological status using multiple discriminant function analysis of neuropsychological test scores. *Journal of Consulting and Clinical Psychology*, *45*(1), 145. doi:10.1037/0022-006X.45.1.145

Sumitani, S., Tanaka, T., Tayoshi, S., Ota, K., Kameoka, N., Ueno, S., & Ohmori, T. (2006). Activation of the prefrontal cortex during the Wisconsin Card Sorting Test as measured by multichannel near-infrared spectroscopy. *Neuropsychobiology*, *53*(2), 70–6.

Swanson, J. (2005). The Delis-Kaplan Executive Function System: A review. *Canadian Journal of School Psychology*, *20*(1-2), 117–128. doi:10.1177/0829573506295469

Swick, D., Honzel, N., Larsen, J., Ashley, V., & Justus, T. (2012). Impaired response inhibition in veterans with post-traumatic stress disorder and mild traumatic brain injury. *Journal of the International Neuropsychological Society*, *18*(05), 917–926. doi:10.1017/S1355617712000458

Tate, R. L. (1999). Executive dysfunction and characterological changes after traumatic brain injury: Two sides of the same coin? *Cortex*, *35*(1), 39–55. doi:10.1016/S0010-9452(08)70784-6

Taylor, J. (2007). Heritability of Wisconsin Card Sorting Test (WCST) and Stroop Color-Word Test performance in normal individuals: Implications for the search for endophenotypes. *Twin Research and Human Genetics*, *10*(06), 829–834. doi:10.1375/twin.10.6.829

Terada, S., Sato, S., Honda, H., Kishimoto, Y., Takeda, N., Oshima, E., … Uchitomi, Y. (2011). Perseverative errors on the Wisconsin Card Sorting Test and brain perfusion imaging in mild Alzheimer's disease. *International Psychogeriatrics*, *23*(10), 1552–9. doi:10.1017/S1041610211001463

Teuber, H.-L. (2009). The riddle of frontal lobe unction in man. *Neuropsychology Review*, *19*(1), 25–46. doi:10.1007/s11065-009-9088-z

Thorndike, R. L. (1949). *Personnel selection; test and measurement techniques* (Vol. viii). Oxford: Wiley.

Tomarken, A. J., & Serlin, R. C. (1986). Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. *Psychological Bulletin*, *99*(1), 90–99. doi:10.1037/0033-2909.99.1.90

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology*, *14*(2), 167–177. doi:10.1016/S0887-6177(97)00095-4

Tranel, D., Kemmerer, D., Adolphs, R., Damasio, H., & Damasio, A. R. (2003). Neural correlates of conceptual knowledge for actions. *Cognitive Neuropsychology*, *20*(3-6), 409–432. doi:10.1080/02643290244000248

Tucha, O., Smely, C., & Lange, K. W. (1999). Verbal and figural fluency in patients with mass lesions of the left or right frontal lobes. *Journal of Clinical and Experimental Neuropsychology*, *21*(2), 229–236. doi:10.1076/jcen.21.2.229.928

Vasilopoulos, T., Franz, C. E., Panizzon, M. S., Xian, H., Grant, M. D., Lyons, M. J., … Kremen, W. S. (2012). Genetic architecture of the Delis-Kaplan Executive Function System Trail Making Test: Evidence for distinct genetic influences on executive function. *Neuropsychology*, *26*(2), 238–250. doi:10.1037/a0026768

Vaughan, L., & Giovanello, K. (2010). Executive function in daily life: Age-related influences of executive processes on instrumental activities of daily living. *Psychology and Aging*, *25*(2), 343–355. doi:10.1037/a0017729

Walsh, K. W. (1978). *Neuropsychology: A clinical approach*. Oxford: Churchill Livingstone.

Waters-Wood, S. M., Xiao, L., Denburg, N. L., Hernandez, M., & Bechara, A. (2012). Failure to learn from repeated mistakes: Persistent decision-making impairment as measured by the Iowa Gambling Task in patients with ventromedial prefrontal cortex lesions. *Journal of the International Neuropsychological Society*, *18*(05), 927–930. doi:10.1017/S135561771200063X

Wecker, N. S., Kramer, J. H., Hallam, B. J., & Delis, D. C. (2005). Mental flexibility: Age effects on switching. *Neuropsychology*, *19*(3), 345–352. doi:10.1037/0894-4105.19.3.345

Weigl, E. (1941). On the psychology of so-called processes of abstraction. *The Journal of Abnormal and Social Psychology*, *36*(1), 3–33. doi:10.1037/h0055544

Weinberger, D. R., Berman, K. F., Suddath, R., & Torrey, E. F. (1992). Evidence of dysfunction of a prefrontal-limbic network in schizophrenia: A magnetic resonance imaging and regional cerebral blood flow study of discordant monozygotic twins. *The American Journal of Psychiatry*, *149*(7), 890–897.

Welch, B. L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika*, *38*(3/4), 330–336. doi:10.2307/2332579

Wiens, A. N., Bryan, J. E., & Crossen, J. R. (1993). Estimating WAIS-R FSIQ from the National Adult Reading Test-Revised in normal subjects. *Clinical Neuropsychologist*, *7*(1), 70–84. doi:10.1080/13854049308401889

Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and Plasticity. *Child Development Perspectives*, *6*(4), 354–360. doi:10.1111/j.1750-8606.2012.00246.x

Zelazo, P. D., Craik, F. I. M., & Booth, L. (2004). Executive function across the life span. *Acta Psychologica*, *115*(2–3), 167–183. doi:10.1016/j.actpsy.2003.12.005

Appendix A

Informed Consent Form

You are invited to participate in a research study conducted by Sunshine Collins, M.S., M.A. of Indiana University of Pennsylvania (IUP) and supervised by Dr. David LaPorte. The following information is provided to help you to make an informed decision whether or not to participate. If you have any questions, please do not hesitate to ask. You are eligible to participate because you enrolled with the subject pool at IUP.

The purpose of this study is to evaluate the relationship between three popular tests of executive functioning. You will be asked to actively participate in a testing session that involves the administration of these measures, which is estimated at a 1 hour total commitment. You may have been chosen to receive a non-standard administration for experimental purposes. The benefits of participating in this study will include contributing to research in the area of executive functions and accruing research credit as required for PSYCH 101. There are no anticipated risks associated with this study beyond possible lack of interest in the activities.

Your participation in this study is voluntary and you are free to rescind your consent to participate at any time. Your refusal to participate will not impact your enrollment in the subject pool or PSYC 101. If you choose to participate, you may change your mind and drop out of the study at any time by notifying the primary investigator, Sunshine Collins, and all information that you have provided for this study will be destroyed. If you choose to participate, all information that you provide for this study will be held in strict confidence and will be published only in aggregate form. Information obtained in the study may be published in scientific journals or presented at scientific meetings, but your identity will be kept strictly confidential.

If you are willing to participate in this study, please sign the statement below. Please keep the extra copy for your records.

Sunshine Collins, M.S., M.A.
Graduate Student
Principle Investigator
Indiana University of Pennsylvania
1020 Uhler Hall
Oakland Ave.
Indiana, PA 15701
(724) 357-6228

David LaPorte, Ph.D.
Director of Clinical Training
Faculty Advisor
Indiana University of Pennsylvania
1020 Uhler Hall
Oakland Ave.
Indiana, PA 15701
(724) 357-4524

Appendix B

Demographics Questionnaire

1. Please circle which age or age range most accurately represents your current age:
   a. 18
   b. 19
   c. 20
   d. 21-23
   e. 24-26
   f. 27-29
   g. 30 and over

2. Please circle the cultural group that best describes you:
   a. American Indian or Alaskan Native
   b. Asian or Pacific Islander
   c. Black
   d. White, Non-Hispanic
   e. Hispanic or Latino

3. Please indicate your gender:
   a. Male
   b. Female

4. Please indicate your current year in college:
   a. Freshman
   b. Sophomore
   c. Junior
   d. Senior
   e. Super senior (4+years)

5. Have you ever taken a test where you were asked to sort cards either here or somewhere else?
   a. Yes
   b. No

6. Have you ever had a head injury (concussion)? This could have happed due to sports, accidents, falls, etc.
   a. Yes
   b. No

Appendix C

Table 26

D-KEFS Sorting Test Comparison by Condition

D-KEFS Sorting Test	Control condition	Minimal Instruction Condition	Feedback Condition
Screening Pretest	I'd like you to read these words out loud and tell me if there are any words you would like me to explain. Go ahead. Do you know the meanings of all of these words?	Removed	Retained in standard format
Practice Card Set	I'm going to show you six cards that can be sorted in different ways. I want to see how many different ways you can sort cards. Let me show you what I mean with these cards. Look at these cards. Watch how I sort them into *two groups*, with *three cards* in each group. Next I'll explain how I sorted them by saying, this group has circles, and this group has squares. Notice how explain *both* groups, not just one of them. Now watch while I sort them another way, again with *two* groups and *three* cards in each group. I will explain how I sorted them by saying, this group has boy's names, and this group has girl's names. Do you have any questions about how I did this?	Removed	Retained in standard format
Free Sorting Condition, Card Set 1	I'm going to show you six new cards that can be sorted in many different ways. I'd like to see how many different ways you can sort these cards. Each time, make only *two groups* with *three cards* in each group. The three cards in each group should be the same in some way. After you sort the cards into two groups, tell me how you did it. Be sure to tell me how you sorted *both* groups, not just one of them. Once you sort the cards one way, do not sort them that way again. Work as quickly as you can.	I'm going to show you six ~~new~~ cards that can be sorted in many different ways. I'd like to see how many different ways you can sort these cards. Each time, make only *two groups* ~~with three cards in each group~~ [and make sure to use all of the cards*]*. The ~~*three*~~ cards in each group should be the same in some way. After you sort the cards into two groups, tell me how you did it. Be sure to tell me how you sorted *both* groups, not just one of them. Once you sort the cards one way, do not sort them that way again. Work as quickly as you can.	I'm going to show you six new cards that can be sorted in many different ways. I'd like to see how many different ways you can sort these cards. Each time, make only *two groups* with *three cards* in each group. The three cards in each group should be the same in some way. After you sort the cards into two groups, tell me how you did it. Be sure to tell me how you sorted *both* groups, not just one of them. [I cannot tell you how to sort the cards but I will tell you each time whether you are right or wrong.] Once you sort the cards one way, do not sort them that way again. Work as quickly as you can. Here is a page that will help you

	Here is a page that will help you remember these rules. Now try sorting these cards in as many different ways as you can. Ready? Begin. Now try to sort them in a different way.	~~Here is a page that will help you remember these rules.~~ Now try sorting these cards in as many different ways as you can. Ready? Begin. Now try to sort them in a different way.	remember these rules. Now try sorting these cards in as many different ways as you can. Ready? Begin. Now try to sort them in a different way.
Free Sorting Condition, Card Set 2	I'm going to show you six new cards that can be sorted in many different ways. I'd like to see how many different ways you can sort these cards. Each time, make only *two groups* with *three cards* in each group. The three cards in each group should be the same in some way. After you sort the cards into two groups, tell me how you did it. Be sure to tell me how you sorted *both* groups, not just one of them. Once you sort the cards one way, do not sort them that way again. Work as quickly as you can. Again, here is the page that will help you remember these rules. Now try sorting these cards in as many different ways as you can. Ready? Begin. Now try to sort them in a different way.	Instruction removed, stimuli switched without further guidance	I'm going to show you six new cards that can be sorted in many different ways. I'd like to see how many different ways you can sort these cards. Each time, make only *two groups* with *three cards* in each group. The three cards in each group should be the same in some way. After you sort the cards into two groups, tell me how you did it. Be sure to tell me how you sorted *both* groups, not just one of them. [I cannot tell you how to sort the cards but I will tell you each time whether you are right or wrong.] Once you sort the cards one way, do not sort them that way again. Work as quickly as you can. Again, here is the page that will help you remember these rules. Now try sorting these cards in as many different ways as you can. Ready? Begin. Now try to sort them in a different way.
Sort Recognition Condition, Card Set 1	Now I'm going to put these cards into two groups of three cards each. The three cards in each group will be the same in some way. I want you to tell me how the cards are the same in each group. Be sure to tell me how I sorted *both* groups, not just one of them. I will use a *different* way of sorting the cards each time I put them into groups.	Retained in standard format	Now I'm going to put these cards into two groups of three cards each. The three cards in each group will be the same in some way. I want you to tell me how the cards are the same in each group. Be sure to tell me how I sorted *both* groups, not just one of them. I will use a *different* way of sorting the cards each time I put them into groups. [I cannot tell you how I sorted the cards, but I will tell you each time whether you are right or wrong.]
Sort Recognition, Card Set 2	Like before, I'm going to put these cards into two groups of three cards each. The three cards in each group will be the same in some way. I want you to tell me how the cards are the same in each group. Be sure to tell me how I sorted *both* groups, not just one of them. I will use a *different* way of sorting the cards each time I	Instruction removed, stimuli switched without further guidance	Like before, I'm going to put these cards into two groups of three cards each. The three cards in each group will be the same in some way. I want you to tell me how the cards are the same in each group. Be sure to tell me how I sorted *both* groups, not just one of them. I will use a *different* way of sorting the cards each time I

	put them into groups.		put them into groups. [I cannot tell you how I sorted the cards, but I will tell you each time whether you are right or wrong.]
Miscellaneous	None	Rule summary pages removed, with the associated instruction lines	None

Appendix D

Debriefing Form

Thank you for participating in this study. The purpose of this research is to explore the merits of a popular test of executive function. In this study you were administered tests commonly used to assess executive functioning or intellectual ability. You were also asked to provide limited demographic information.

If you have any questions or would like more information about this study, please contact Sunshine Collins by phone ((724) 357-6228) or email (s.m.collins@iup.edu). If you would like more information about executive functions, please refer to the following articles:

Stuss, D. T., & Benson, D. F. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin*, *95*(1),3–28. doi: 10.1037/0033-2909.95.1.3

Stuss, D. T., & Levine, B. (2002). Adult clinical neuropsychology: Lessons from studies of the frontal lobes. *Annual Review of Psychology*, *53*, 401–33.

Thank you for your participation in this study!